สถิติ
สำหรับการวิจัยทางเกษตร
Statistics for Agricultural Research

โดย

ศ. ดร. ไพศาล เหล่าสุวรรณ

สาขาเทคโนโลยีการผลิตพืช
สำนักวิชาเทคโนโลยีการเกษตร
มหาวิทยาลัยเทคโนโลยีสุรนารี
2545

สถิติสำหรับการวิจัยทางเกษตร

ISBN 974 – 533 – 075 – 2

ฉบับแก้ไขครั้งที่ 3

จัดพิมพ์และใช้ประโยชน์โดย
มหาวิทยาลัยเทคโนโลยีสุรนารี
อ. เมือง จ. นครราชสีมา 30000

2545

คำนำ

(ฉบับแก้ไขปรับปรุง)

นับตั้งแต่ได้เขียนหนังสือเรื่องวิธีการวิจัยขึ้นในปี พ.ศ. 2515 ใช้สอนในระดับปริญญาตรีอยู่หลายปี ต่อมาได้ปรับปรุงใหม่ในปี 2531, 2535 และได้ปรับปรุงอีกครั้งในปี 2544 ดังที่เห็นอยู่นี้ ความหนาก็เพิ่มขึ้นเรื่อย ๆ ตามความประสงค์ของการใช้ประโยชน์

ตำราเล่มนี้เหมาะสำหรับใช้ประโยชน์ในการสอนระดับปริญญาตรี และบัณฑิตศึกษาในสาขาเกษตรศาสตร์ และวิทยาศาสตร์ชีวภาพ อย่างไรก็ดีตัวอย่างที่ยกมาแสดงส่วนใหญ่เป็นทางด้านพืช แต่ผู้สอนวิชานี้อาจหาตัวอย่างในสาขาของตนมาใช้ได้ การเขียนใช้สไตล์ของผู้เขียนเอง คือ ง่าย ลดความสลับซับซ้อน เน้นในด้านให้สามารถใช้ประโยชน์ได้ ตำราเล่มนี้อาจใช้ได้กับหลักสูตรที่มีการสอนวิชาการวางแผนการทดลองเพียงวิชาเดียวโดยไม่มีพื้นฐานวิชาสถิติมาก่อนเลย เพราะได้สรุปพื้นความรู้ไว้ครบถ้วนแล้ว และใช้ได้กับการสอนปริญญาตรีทั่วไป และถึงระดับบัณฑิตศึกษา โดยที่อาจใช้ตำราระดับสูงกว่านี้ร่วมด้วย ทั้งนี้เพราะไม่อาจจะรวบรวมรายละเอียดทุกอย่างไว้ในเล่มนี้ได้ครบ

ตำราเล่มนี้ยังไม่ได้จัดพิมพ์อย่างถาวร พร้อมที่จะปรับปรุงแก้ไขได้ทุกเวลา ดังนั้นหากผู้ใช้ท่านใดเห็นส่วนบกพร่องผิดพลาดควรแก้ไข โปรดแจ้งที่โทร. 044 224155, Fax 044 225150

ที่จะไม่ลืมขอบคุณ คือเจ้าหน้าที่และศิษย์ทุกคนที่ช่วยจัดพิมพ์ และจัดทำรูปเล่มดังที่เห็นนี้

ไพศาล เหล่าสุวรรณ
15 พฤษภาคม 2545

สารบัญ

บทที่ 1
พื้นฐานทางสถิติ*

1.1 คำนำ

วิชาสถิติ[1]นับว่าเป็นวิชาที่มีความสำคัญยิ่งวิชาหนึ่ง เราใช้ความรู้จากวิชานี้เพื่อช่วยในการตัดสินใจในปัญหาต่าง ๆ ได้มาก และถือว่าเป็นการตัดสินใจที่ถูกต้องด้วยหลักวิชาการ เพราะมีการตั้งสมมุติฐานว่าเหตุการณ์จะเป็นอย่างนั้นอย่างนี้ แล้วมีการรวบรวมข้อมูลมาทดสอบเพื่อหาคำตอบ ตัวอย่างเช่น ถ้านายแดงมีความต้องการที่จะสมัครรับเลือกตั้งเป็นผู้แทนราษฎรในจังหวัดหนึ่ง แต่ไม่แน่ใจว่าจะได้รับความนิยมมากน้อยเพียงใด จึงจัดให้มีการสำรวจความคิดเห็นจากประชาชนบางส่วนในเขตที่ตนจะลงสมัคร แล้วนำความคิดเห็นเหล่านี้มาวิเคราะห์ดูก่อนจะตัดสินใจว่าควรจะสมัครรับเลือกตั้งหรือไม่ จะเห็นได้ว่าวิธีการเช่นนี้เป็นการใช้วิชาการทางสถิติเข้ามาช่วยในการตัดสินใจ ในการทดลองทางเกษตรอาจต้องมีการตัดสินใจเลือกใช้วิธีการเพาะปลูกพืชวิธีใดวิธีหนึ่งในหลาย ๆ วิธีหรือเลือกใช้พืชพันธุ์ใดพันธุ์หนึ่งในจำนวนหลายพันธุ์ ฯลฯ เมื่อพบกับปัญหาเช่นนี้เราก็สามารถใช้วิธีการทางสถิติเข้าช่วยในการตัดสินใจได้เช่นกัน

1.2 ความหมายเบื้องต้น

สถิติคืออะไร สถิติหมายถึงวิธีการที่กระทำต่อข้อมูล[2] ซึ่งหมายถึงการรวบรวมข้อมูล การนำเสนอข้อมูล (ในรูปของตาราง กราฟ หรือแผนภูมิ) การวิเคราะห์ข้อมูล และการตีความหมายข้อมูล เราอาจแบ่งวิชาสถิติออกได้เป็น 2 ภาคใหญ่ ๆ คือ (1) **สถิติพรรณนา**[3] ซึ่งเกี่ยวกับการรวบรวมและนำเสนอข้อมูล หมายถึง (ก) **การหาแนวโน้มเข้าสู่ส่วนกลาง**[4] : เช่น **การหาค่าเฉลี่ย**[5] การหามัธยฐาน[6] ฯลฯ และ (ข) **การหาการกระจาย**[7] : ซึ่งได้แก่ การหาพิสัย[8] การหาค่าเบี่ยงเบนเฉลี่ย[9] การหาวาเรียนซ์[10] การหาส่วนเบี่ยงเบนมาตรฐาน[11] ฯลฯ และ (2) **สถิติอนุมาณ**[12] ซึ่งเกี่ยวข้องกับการให้ข้อสรุปเกี่ยวกับข้อมูลที่รวบรวมได้

เพื่อให้เข้าใจเกี่ยวกับสถิติ 2 ภาคนี้ได้ดีขึ้น ขอยกตัวอย่างดังนี้ ในการสอบวิชาต่าง ๆ 3 วิชา นักศึกษาคนที่หนึ่งได้คะแนนร้อยละ 63, 41 และ 55 ส่วนคนที่สองได้ร้อยละ 60, 58 และ 59 นั่นคือนักศึกษาคนแรกได้คะแนนเฉลี่ยร้อยละ 53 ส่วนคนที่สองได้คะแนนเฉลี่ยร้อยละ 59 การสำรวจและการเสนอข้อมูลแบบนี้เรียกว่าสถิติพรรณนา คือเราเสนอเฉพาะข้อมูลและค่าเฉลี่ย เพื่อบอกให้ทราบว่าข้อมูลมีลักษณะอย่างไรบ้าง แต่ถ้าเราให้ข้อสรุปต่อไปว่า นักศึกษาคนที่สองเรียนเก่งกว่านักศึกษาคนแรก ก็เรียกว่าสถิติอนุมาณ

***คำที่ตามด้วยตัวเลขในหนังสือเล่มนี้ได้รับการแปลหรือบัญญัติจากคำภาษาอังกฤษที่รวมอยู่ตอนท้ายแต่ละบท**

ประชากร[13] คำว่าประชากรมิได้หมายถึงกลุ่มของพืชหรือสัตว์ แต่ในทางสถิติหมายถึงกลุ่มของข้อมูลเกี่ยวกับลักษณะของพืช สัตว์ หรือสิ่งของต่าง ๆ กลุ่มนี้ต้องมีค่าของสมาชิกทั้งหมดตามขอบเขตที่กำหนด เช่น ความสูงของชายไทยอายุ 25 ปี ทั่วประเทศ ในที่นี้ข้อความว่า "อายุ 25 ปี ทั่วประเทศ" คือขอบเขตที่กำหนด ประชากรแบบนี้เรียกว่า เป็นประชากรที่มีขอบเขตจำกัด[14]ประชากรบางชนิดมีขอบเขตไม่จำกัด[15] เช่น จำนวนด้านหัวที่ได้จากการโยนเหรียญ ซึ่งอาจมีจำนวนครั้งของการโยนที่ไม่จำกัดแน่นอน

ตัวอย่าง[16] ตัวอย่างหมายถึงกลุ่มของข้อมูลที่เกี่ยวข้องกับลักษณะ หรือสิ่งของชนิดหนึ่ง ๆ ซึ่งตัวอย่างนี้ เป็นเพียงบางส่วนของประชากรใดประชากรหนึ่ง ตัวอย่างนี้ใช้เป็นตัวแทนของประชากร ทั้งนี้เมื่อไม่สามารถนำข้อมูลทั้งหมดของประชากรมาศึกษาได้ เช่น ในการวัดความสูงของชายไทยอายุ 25 ปี อาจสุ่มมาวัดจำนวน 100 คน ก็ใช้แทนความสูงของชายไทยทั่วประเทศ ซึ่งไม่อาจทำการวัดได้ทั่วถึง

ตัวแปร[17] ตัวแปรคือข้อมูลที่แปรผันขึ้นลง สูง ต่ำ หรือมากน้อยแตกต่างกันไปตามธรรมชาติของประชากร เช่น ในกระสอบใบหนึ่งมีข้าวโพด 100 ฝัก เมื่อนำมาวัดความยาวแต่ละฝักปรากฏว่ายาวไม่เท่ากัน ข้อมูลเหล่านี้เรียกว่า ตัวแปร ในทางสถิติอาจให้สัญลักษณ์ตัวแแปรว่าเป็น X เช่น ในกรณีของความยาวของฝักข้าวโพดก็ให้สัญลักษณ์เป็น $X_1, X_2, X_3, \ldots, X_{100}$ ซึ่งอาจมีค่าเป็น 18, 25, 23, ..., 20 ซม. ตามลำดับ ถ้าเราไม่ทราบว่ามีข้าวโพดกี่ฝักแน่นอนก็อาจเรียกว่ามี n ฝัก ดังนั้น X จะมีค่าจาก X_1 จนถึง X_n

1.3 การสุ่มตัวอย่าง

การสุ่มตัวอย่าง[18] คือกระบวนการหรือวิธีการปฏิบัติเพื่อให้ได้มาซึ่งข้อมูลของตัวอย่างแต่ละชุด เช่นในแปลงปลูกข้าวโพดแปลงหนึ่งมีข้าวโพด 1,000 ฝัก เราต้องการตัวอย่างเพียง 30 ฝัก นำมาวัดความยาว การได้มาซึ่งฝักเหล่านี้ต้องกระทำโดยสุ่ม ซึ่งอาจใช้วิธีการสุ่มแบบทั่วถึง[19] จุดประสงค์ในการสุ่มก็เพียงเพื่อให้ข้อมูลแต่ละตัวมีโอกาสเข้าสู่ชุดตัวอย่างได้เท่า ๆ กัน ในการสุ่มแต่ละครั้งนั้นมีการปฏิบัติต่อสมาชิกที่ได้รับการสุ่ม 2 แบบ คือใส่สมาชิกคืนที่[20] และไม่ใส่คืนที่[21] เช่นในกรณีของฝักข้าวโพดนี้ เมื่อเราสุ่มได้ฝักใดและบันทึกข้อมูลแล้วคือวัดความยาวแล้ว ก็โยนกลับลงไปในกองตามเดิม หรือไม่โยนใส่ในกองเดิม ก็เรียกว่า เป็นการสุ่มแบบใส่สมาชิกคืนที่เดิม และไม่ใส่คืนที่เดิมตามลำดับ

การสุ่มตัวอย่างมีเทคนิคและรายละเอียดมาก แต่เราจะทบทวนเฉพาะวิธีการสุ่มอย่างง่าย[22] วิธีการสุ่มแบบนี้อาจทำได้ 2 วิธี คือ

(1) วิธีจับสลาก เช่น ประชากรหนึ่งมีขนาด 100 ตัวแปร ซึ่งมีการจัดระเบียบเป็นลำดับเลขที่ไว้แล้ว ถ้าต้องการชุดตัวอย่างที่มีขนาด 5 ตัวแปร (n = 5) ก็เขียนหมายเลข 1 ถึง 100 ลงในแผ่นกระดาษ 4 เหลี่ยมแผ่นเล็ก ๆ ม้วนให้กลม ใส่ลงไปในภาชนะอะไรก็ได้ จากนั้นก็เขย่า แล้วหยิบขึ้น

มา 5 แผ่น เช่น หยิบได้หมายเลข 10, 91, 28, 43 และ 62 ตามลำดับ ก็แสดงว่าชุดตัวอย่างของเรามีหมายเลขดังกล่าวนี้ ต่อจากนั้นก็ตรวจดูข้อมูลของหมายเลขเหล่านั้นว่ามีค่าเท่าใด เช่น ถ้าเป็นฝักข้าวโพดก็อาจยาว 18, 20, 24, 22 และ 21 ซม. ตามลำดับ นี่คือชุดตัวอย่างที่เราได้รับในการสุ่มครั้งนี้

(2) วิธีสุ่มโดยใช้ตารางเลขสุ่ม ตารางเลขสุ่มคือตารางที่มีค่าต่าง ๆ ปรากฏขึ้นอย่างสุ่ม บางส่วนของตารางนี้ปรากฏอยู่ท้ายเล่มของหนังสือนี้ (ตาราง ผ. 1) ตารางเต็มรูปแบบมี 10,000 ค่า คือมีข้อมูลจำนวน 100 แถว และ 100 สดมภ์ การสร้างตารางนี้ทำเพียงง่าย ๆ คือ เขียนเลข 0 ถึง 9 ใส่ในฉลาก 10 ใบ แล้วสุ่มขึ้นมาครั้งละ 1 ใบ ได้เลขอะไรก็จดเอาไว้ แล้วใส่คืนลงไปทำ 10,000 ครั้ง ก็ได้ 10,000 ค่า การแยกเป็นแถวและสดมภ์ก็ไม่มีความประสงค์อื่นใด นอกจากจะจัดให้เป็นระเบียบเท่านั้น การใช้ตารางนี้มีวิธีการดังนี้ (1) สุ่มหาลำดับของสมาชิกในชุดตัวอย่าง และ (2) ตรวจสอบว่าสมาชิกนั้น ๆ มีค่า (เช่น น้ำหนัก) เท่าไร เช่น ในประชากรที่มี 100 ตัวแปร เช่น ข้าวโพด 100 ฝัก ใส่หมายเลขทุกฝักตั้งแต่หมายเลข 1 ถึง 100 ถ้าเราต้องการชุดตัวอย่างที่ $n = 5$ ก็สุ่มตัวเลขจากตารางมาเป็นคู่ ๆ จำนวน 5 ชุด เช่น สุ่มได้ค่า 80, 35, 08, 42, 99 ก็แสดงว่าชุดตัวอย่างของเราประกอบด้วยสมาชิกหมายเลข 80, 35, 8, 42 และ 99 ตามลำดับ (การสุ่ม) ต่อจากนั้นก็ไปสำรวจดูว่าสมาชิกเหล่านั้นมีค่าเท่าไร คือมีความยาวของฝักเท่าไร หรือน้ำหนักเท่าไร ในการสุ่มนั้นถ้าเลขชุดใดเกิดซ้ำก็ให้ตัดทิ้งไป สังเกตว่าถ้าประชากรมีสมาชิก 100 ตัวแปร ก็อ่านตัวเลข 2 หลัก คือตั้งแต่ 00 ถึง 99 ถ้าประชากรมีจำนวนเกิน 100 ตัวแปร เช่น 120 ตัวแปร ก็ใช้เลข 3 หลัก คือ 000 ถึง 119 (ถ้าสุ่มได้ค่า อื่น ๆ ให้ตัดทิ้งไป) ในทั้งสองกรณีนี้หมายเลข 00 และ 000 อาจใช้แทนสมาชิกหมายเลขต่ำสุดหรือสูงสุดก็ได้

1.4 การนำเสนอข้อมูล

การนำเสนอข้อมูล[23] คือการจัดระเบียบของข้อมูลในแบบต่าง ๆ เพื่อให้สามารถใช้ประโยชน์ต่อไป การนำเสนอมีได้หลายแบบ ดังนี้

(1) การนำเสนอแบบตาราง[24] ตารางคือข้อมูลที่มีการจัดระเบียบทั้งในแนวตั้งและแนวนอน ซึ่งเรียกว่าเป็นการจัดสดมภ์และแถวตามลำดับ ถ้าข้อมูลมีค่าซ้ำ ๆ กัน ก็มีการจัดเป็นชั้น ๆ และแต่ละชั้นเราสำรวจว่ามีสมาชิกจำนวนเท่าใด การนำเสนอข้อมูลแบบนี้เรียกว่า การทำตารางแจกแจงความถี่[25]

(2) การนำเสนอแบบอื่น ๆ การนำเสนอข้อมูลอาจกระทำโดยการใช้กราฟ ใช้แผนภูมิและรูปแบบต่าง ๆ ก็ได้ ในการเสนอแบบกราฟนั้น อาจใช้กราฟรูปสี่เหลี่ยมผืนผ้า ซึ่งเรียกว่า ฮิสโตแกรม[26] พื้นที่ของสี่เหลี่ยมผืนผ้าแต่ละรูปจะสอดคล้องกับความถี่ หรือจำนวนสมาชิกในแต่ละกลุ่ม อนึ่ง ถ้าเราลากเส้นผ่านยอดบนตรงจุดกลางของแต่ละชั้น ก็จะได้กราฟอีกชนิดหนึ่ง ซึ่งเราเรียกว่า รูปหลายเหลี่ยม[27] การนำเสนอข้อมูลอาจจะเป็นแบบอื่น ๆ อีกก็ได้ เช่น ใช้รูปหลายเหลี่ยมความถี่สะสม[28] คือนำความถี่จากแต่ละชั้นในตารางแจกแจงความถี่มารวมเข้าด้วยกันแล้ว แล้วเขียนเป็นกราฟ การนำเสนอข้อมูลอาจใช้รูปโค้งความถี่[29] คือทำให้รูปหลายเหลี่ยมมีขอบมนขึ้น ก็จะกลายเป็นเส้นโค้งเรียบ รูปโค้งความถี่

อาจจะเป็นแบบการกระจายปกติ[30] ซึ่งมีรูปแบบระฆัง หรืออาจมีความเบ้[31] หรืออาจมียอดสูงได้หลายยอด[32]

1.5 แนวโน้มเข้าสู่ส่วนกลาง

แนวโน้มเข้าสู่ส่วนกลางหรือแนวโน้มเข้าหาศูนย์[33] คือค่าที่สามารถใช้เป็นตัวแทนของกลุ่ม เป็นค่าที่อาจบอกได้ว่าค่าอื่น ๆ ที่เราไม่ทราบนั้น อาจมีขนาดหรือลักษณะอย่างไร เราใช้ค่านี้เพื่อบอกขนาด ปริมาณ หรือ ลักษณะของข้อมูลกลุ่มหนึ่ง ๆ และอาจใช้เป็นตัวแสดงความแตกต่างจากข้อมูลกลุ่มอื่น ๆ โดยปกติแนวโน้มเข้าสู่ส่วนกลางจัดเป็นค่ากลาง ๆ คือไม่แตกต่างจากค่าอื่น ๆ มากนัก และค่าดังกล่าวนี้มีอยู่หลายชนิดคือ

(1) ค่าเฉลี่ยหรือตัวกลางเลขคณิต[34] ค่าเฉลี่ย คือ ค่าที่เกิดจากผลรวมของค่าทั้งหมดที่หารด้วยจำนวนข้อมูล ค่าเฉลี่ยของประชากรเราใช้สัญลักษณ์ μ (มิว) ส่วนค่าเฉลี่ยของชุดตัวอย่างใช้สัญลักษณ์ ว่า $\overline{X}$ (เอกซ์บาร์)

ในทางคณิตศาสตร์ เรามักใช้อักษรเป็นสัญลักษณ์แทนข้อมูล อักษรที่ใช้อาจเป็น X หรือ Y เมื่อมีหลายค่า ก็ให้ i หรือ j บอกลำดับค่าเหล่านั้น เช่น ในการชั่งน้ำหนักเมล็ดของถั่วเหลืองพันธุ์ สจ. 4 จำนวน 5 ชุด ชุดละ 100 เมล็ด ก็อาจแสดงสัญลักษณ์และข้อมูลดังนี้

สัญลักษณ์	X_1	X_2	X_3	X_4	X_5
ข้อมูล (กรัม)	16.4	16.3	16.6	15.8	15.9

สมการสำหรับหาค่าเฉลี่ย คือ

$$\overline{X} = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{n}$$

หรือเขียนว่า $\overline{X} = (X_1 + X_2 + X_3 + X_4 + X_5)/n$

คือนำทุกค่ามารวม (บวก) กัน แล้วหารด้วยจำนวนข้อมูล ซึ่งถ้ามีข้อมูลมากก็อาจเขียนว่า

$$\overline{x} = \frac{(X_1 + X_2 + ... + X_n)}{n}$$

$$= \frac{\sum_{i=1}^{n} X}{n}$$

$$= \frac{\sum X_i}{n} \qquad ...(1\text{-}1)$$

เมื่อให้ $\sum$ (ซิกมา) แทนผลรวม, n เป็นจำนวนสมาชิกของชุดตัวอย่าง และ i = 1, 2, 3, ... , n (อ่านว่า i มีค่าค่าตั้งแต่ 1, 2, 3, จนถึง n) ใช้เป็นตัวบอกอันดับของสมาชิกของชุดตัวอย่าง

ค่าเฉลี่ยของประชากร (μ) ก็หาโดยวิธีเดียวกัน เพียงแต่บวกทุกค่าที่ประชากรนั้นพึงมีเท่านั้น

จากตัวอย่างข้างบนหาได้ว่า ค่าเฉลี่ยของน้ำหนัก 100 เมล็ด ของถั่วเหลืองพันธุ์ สจ. 4 คือ

$$\overline{X} = \frac{\sum X_i}{n}$$

$$\overline{X} = \frac{16.4 + 16.3 + 16.6 + 15.8 + 15.9}{5}$$

$$= \frac{81}{5} = 16.2 \text{ กรัม /100 เมล็ด}$$

ค่าดังกล่าวนี้ใช้เป็นตัวบอกขนาดเมล็ดของถั่วเหลืองพันธุ์ สจ. 4 ซึ่งแสดงว่า แม้จะสุ่มขึ้นมาอีกกี่ครั้งก็ใกล้เคียง 16.2 กรัม นั่นเอง

(2) ค่าเฉลี่ยเรขาคณิต[35] ค่าเฉลี่ยเรขาคณิต คือรากที่ n ของผลคูณของข้อมูลใช้สัญลักษณ์ว่า GM หาได้โดยสมการ

$$GM = \sqrt[n]{X_1 \times X_2 \times ... \times X_n}$$

การใช้วิธีนี้มีความยุ่งยากมาก อาจใช้ค่าล็อก[36] ดังนี้

$$\begin{aligned} \log GM &= \log \sqrt[n]{(X_1 \times X_2 \times ... \times X_n)} \\ &= (\log X_1 + \log X_2 + ... + \log X_n)/n \\ &= (\sum \log X_i)/n \\ GM &= \text{antilog} \qquad ...(1\text{-}2) \end{aligned}$$

เมื่อถอดค่าล็อก[37] ก็ได้ค่าเฉลี่ย

ตัวอย่าง 1.5.1

สมมุติว่าลักษณะความสูงของต้นข้าวฟ่างควบคุมโดยยีน 3 คู่ เมื่อมียีนหนึ่งคู่ ข้าวฟ่างสูง 40 ซม. มีสองคู่สูง 80 ซม. และเมื่อมีสามคู่สูง 160 ซม. จงหาผล เฉลี่ยของยีนทั้ง 3 คู่

วิธีทำ **จะเห็นได้ว่าผลของยีนแต่ละคู่เพิ่มเป็นแบบเท่าตัวหรือแบบคูณ ยีนแสดงผลเช่นนี้เรียกว่าแสดงผลแบบเรขาคณิต[38] ดังนั้นค่าเฉลี่ยหาได้จากสมการ (1-2) ดังนี้**

$$\begin{aligned} \log GM &= (\log 40 + \log 80 + \log 160)/3 \\ &= (1.60206 + 1.90309 + 2.20412)/3 \\ &= 5.70927/3 \\ &= 1.90309 \\ GM &= 80 \text{ ซม.} \end{aligned}$$

คือโดยเฉลี่ยแล้วยีนทั้ง 3 คู่ให้ความสูง 80 ซม.

(3) ค่าเฉลี่ยฮาร์โมนิก[39] ค่าเฉลี่ยฮาร์โมนิกคือส่วนกลับของค่าเฉลี่ยเลขคณิต ใช้ในบางโอกาสเท่านั้น หาได้จากสมการ

$$HM = \frac{n}{\sum \frac{1}{X}} \qquad \ldots(1\text{-}3)$$

ตัวอย่าง 1.5.2

เขามีเงินอยู่ 240 บาท ครึ่งหนึ่งนำไปซื้อสมุดเล่มละ 6 บาท อีกครึ่งหนึ่งนำไปซื้อสมุดเล่มละ 4 บาท จงหาราคาเฉลี่ยของสมุดที่ซื้อทั้งหมด

วิธีทำ ค่าเฉลี่ยดังกล่าวหาได้จากสมการดังนี้

$$HM = \frac{2}{\frac{1}{6} + \frac{1}{4}} = 4.80$$

คือสมุดเหล่านี้ราคาเฉลี่ยเล่มละ 4.80 บาท

นอกจากนี้แนวโน้มเข้าสู่ส่วนกลาง อาจวัดโดยใช้มัธยฐาน[40] ซึ่งหมายถึง ค่าที่แบ่งการกระจายออกเป็น 2 ส่วน แต่ละส่วนมีพื้นที่หรือจำนวนข้อมูลเท่ากัน เช่น ข้อมูลชุดหนึ่งมีค่าเป็น 2, 4, 7, 9, 12 มัธยฐานของเลขชุดนี้คือ 7 หรืออาจใช้ฐานนิยม[41] ฐานนิยมหมายถึงข้อมูลที่ปรากฏมากครั้งที่สุด เช่น ข้อมูลชุดหนึ่งมีค่าเป็น 2, 4, 4, 7, 6 ฐานนิยมของข้อมูลนี้ได้แก่ 4 นั่นเอง การวัดแนวโน้มเข้าสู่ส่วนกลางวิธีอื่น ๆ ที่จะไม่กล่าวถึงในที่นี้ก็มี คือ ควอไตล์[42] เดไซล์[43] และเปอร์เซ็นไตล์[44]

1.6 การวัดการกระจาย

การกระจายของข้อมูล[45] คือความแผ่กว้างของข้อมูลนั่นเอง เช่น ข้อมูลชุดหนึ่งมีค่า 5, 10 และ 15 ส่วนอีกชุดหนึ่งมีค่า 9, 10 และ 11 ข้อมูล 2 ชุดนี้มีค่าเฉลี่ยเท่ากัน แต่ข้อมูลชุดแรกมีค่าสูงต่ำแตกต่างกันมากกว่า คือมีการกระจายที่แผ่กว้างมากกว่า ซึ่งจะเห็นได้ว่าในบางครั้งการใช้เพียงค่าเฉลี่ยเพื่อบอกลักษณะของข้อมูลยังไม่เป็นการเพียงพอ อาจต้องใช้ค่าอื่น ๆ เช่น การกระจาย ควบคู่ไปด้วย จึงจะทราบลักษณะที่แท้จริงของข้อมูล การวัดการกระจายมีหลายวิธีคือ

(1) พิสัย[46] พิสัยคือ ความแตกต่างระหว่างค่าสูงสุดและค่าต่ำสุดในข้อมูลชุดนั้น เช่น ในตัวอย่างข้างบน ข้อมูลชุดแรกมีพิสัยเท่ากับ 10 (คือ 15 – 5) ส่วนข้อมูลชุดที่สองมีพิสัยเท่ากับ 2 (คือ 11 – 9) ประโยชน์ของพิสัยคือ สามารถตรวจสอบอัตราการกระจายได้อย่างรวดเร็ว ไม่ต้องใช้เทคนิคมากนัก เข้าใจง่าย แต่ก็มีข้อเสีย เช่น ถ้าข้อมูลที่มีค่าสูงหรือต่ำผิดปกติ ก็จะทำให้ได้พิสัยที่สูงผิดปกติไปด้วย และชุดตัวอย่างจากประชากรเดียวกันหลาย ๆ ชุด อาจให้พิสัยต่างกัน และข้อเสียประการสุดท้ายของพิสัยคือ เราไม่อาจใช้ประโยชน์ได้อย่างต่อเนื่อง

(2) ส่วนเบี่ยงเบนเฉลี่ย[47] ส่วนเบี่ยงเบนเฉลี่ย คือค่าเฉลี่ยของค่าเบี่ยงเบนจากตัวกลางเลขคณิต สมการสำหรับหาค่านี้ คือ

$$\text{ส่วนเบี่ยงเบนเฉลี่ย (M.D)} = \frac{\sum\left|X_i - \overline{X}\right|}{n} \qquad \ldots(1\text{-}4)$$

ทั้งนี้ $X_i - \overline{X}$ แต่ละค่าเป็นค่าสัมบูรณ์[48] คือไม่มีเครื่องหมายบวกหรือลบ ตัวอย่างเช่น ข้อมูลชุดหนึ่งมีค่าต่าง ๆ เป็น 72, 81, 86, 69 และ 57 ซึ่งได้ค่าเฉลี่ย 73 ค่าเบี่ยงเบนแต่ละค่าเท่ากับ – 1, 8, 13, – 4 และ – 16 ตามลำดับ ดังนั้น

$$\text{M.D} = \frac{1 + 8 + 13 + 4 + 16}{5} = 8.4$$

ซึ่งพูดได้ว่า โดยเฉลี่ยแล้วข้อมูลแต่ละตัวห่างจากค่าเฉลี่ย 8.4 นั่นเอง ที่น่าสังเกตประการหนึ่งในตอนนี้คือผลรวมของค่าเบี่ยงเบน (ที่ไม่เป็นค่าสัมบูรณ์) จะมีค่าเท่ากับศูนย์ คือผลรวมของ – 1, 8, 13, – 4 และ – 16 จะมีค่าเท่ากับศูนย์

(3) วาเรียนซ์และส่วนเบี่ยงเบนมาตรฐาน[49] การที่จะบอกว่าสมาชิกประชากรหนึ่ง ๆ มีความแตกต่างกันมากน้อยเท่าใด หรือมีความปรวนแปรเท่าใด โดยดูจากขนาดของผลรวมของค่าเบี่ยงเบนก็ทำไม่ได้ เพราะผลรวมของค่านี้เป็นศูนย์ แต่อาจนำมายกกำลังสองเสียก่อนแล้วรวมกัน เราเรียกว่าผลรวมของค่ายกกำลังสอง[50] ถ้าหารด้วยจำนวนค่าสังเกตของประชากร (N) เราเรียกว่าค่าเฉลี่ยของค่ายกกำลังสอง[51] ค่าดังกล่าวนี้เรียกทางสถิติว่า วาเรียนซ์ ซึ่งถ้าเป็นของประชากรใช้สัญญลักษณ์ว่า σ^2 (ซิกยกกำลังสอง) ถ้าเป็นของตัวอย่างใช้สัญญาลักษณ์ว่า s^2 โดยมีสมการตามลำดับดังนี้

$$\sigma^2 = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + \ldots + (X_n - \mu)^2}{N}$$

$$= \frac{\sum (X_i - \mu)^2}{N} \qquad \ldots(1\text{-}5)$$

X เป็นตัวแปร i=1,2,...,N, N คือจำนวนค่าสังเกตทั้งหมดของประชากรนั้น ถ้าเป็นวาเรียนซ์ของตัวอย่างก็ใช้สมการ

$$s^2 = \frac{(X_1-\overline{X})^2+(X_2-\overline{X})^2+\ldots+(X_n-\overline{X})^2}{n-1}$$

$$= \frac{\sum(X_i-\overline{X})^2}{n-1} \qquad \ldots(1\text{-}6)$$

โดยที่ $i = 1, 2, \ldots, n$ (อ่านว่า i มีค่าจาก 1, 2, จนถึง n) ทั้งนี้อาจพิสูจน์ได้ว่า

$$\sum(X_i-\overline{X})^2 = \sum X_i^2 - \frac{(\sum X_i)^2}{n}$$

ดังนั้นสมการหา s^2 ที่ใช้กันทั่วไปคือ

$$s^2 = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}{n-1} \qquad \ldots(1\text{-}7)$$

ตัวหารที่ใช้คำนวณวาเรียนซ์ของตัวอย่างใช้ $n-1$ ค่า $n-1$ นี้เรียกว่า อัตราความเป็นอิสระ[52] ย่อว่า df เมื่อใช้ df เป็นตัวหารจะได้ค่าประมาณของวาเรียนซ์ที่ไม่แตกต่างจากวาเรียนซ์ของประชากร จึงเรียกว่าเป็นค่าประมาณไม่ลำเอียง[53]

ค่าถอดรากสองของวาเรียนซ์เรียกว่าส่วนเบี่ยงเบนมาตรฐาน ส่วนเบี่ยงเบนมาตรฐานของตัวอย่างใช้สัญลักษณ์ว่า s ส่วนเบี่ยงเบนมาตรฐานของประชากรใช้ σ คือหาได้ง่าย ๆ โดยใช้สมการ (1-5), (1-6) หรือ (1-7) นั่นเอง เพียงแต่ให้มีการถอดรากสองเท่านั้น ดังนั้นส่วนเบี่ยงเบนมาตรฐานหาได้จากสมการ

$$s = \sqrt{\frac{\Sigma(X_i-\overline{X})^2}{n-1}} \qquad \ldots(1\text{-}8)$$

$$= \sqrt{s^2}$$

จากตัวอย่างในตอน 1.5 เกี่ยวกับน้ำหนัก 100 เมล็ดของถั่วเหลืองพันธุ์ สจ. 4 อาจหาวาเรียนซ์โดยใช้สมการ (1-5) ดังนี้

$$s^2 = \frac{(16.4-16.2)^2+(16.3-16.2)^2+\ldots+(15.9-16.2)^2}{5-1}$$

$$= 0.115$$

และหาได้ว่าส่วนเบี่ยงเบนมาตรฐาน (s) $= \sqrt{0.115} = 0.339$ กรัม/100 เมล็ด

วาเรียนซ์และส่วนเบี่ยงเบนมาตรฐาน เป็นค่าที่บอกความปรวนแปรของชุดข้อมูล โดยถือเอาค่าเฉลี่ยเป็นหลัก คือถ้าวาเรียนซ์หรือส่วนเบี่ยงเบนมาตรฐานมีค่าสูง ก็แสดงว่าค่าต่าง ๆ ของข้อมูล

ชุดนั้นห่างจากค่าเฉลี่ยมาก ส่วนเบี่ยงเบนมาตรฐานมีค่าได้ทั้งบวกและลบ ทั้งนี้เพราะใช้เป็นมาตรฐานบอกว่าตัวแปรแต่ละตัวห่างจากค่าเฉลี่ยมากน้อยเพียงใด

เพื่อเป็นการพิสูจน์ว่าถ้าใช้ $n-1$ เป็นตัวหารแล้วจะได้ว่าเรียนซ์ที่ไม่ลำเอียง คือ ค่าเฉลี่ยของวาเรียนซ์จากทุกชุดตัวอย่างของประชากรหนึ่ง จะมีค่าเท่ากับวาเรียนซ์ของประชากรนั้น ก็ขอแสดงตัวอย่างดังนี้ : สมมติว่ามีประชากรกลุ่มหนึ่ง ซึ่งมีค่า 3 ค่า คือ 1, 8 และ 9 ถ้าสุ่มตัวอย่างที่มีขนาด $n = 2$ โดยใส่สมาชิกคืนที่ทุกครั้งที่สุ่ม[54] แล้วคำนวณค่าเฉลี่ย วาเรียนซ์และส่วนเบี่ยงเบนมาตรฐานจากประชากรนี้หาได้ว่า

$$\mu = 1 + 8 + 9 = 6.0$$

$$\sigma^2 = \frac{(1-6)^2 + (8-6)^2 + (9-6)^2}{3} = 12.67$$

และ $\sigma = \sqrt{12.67} = 3.559$

การสุ่มตัวอย่างแบบใส่สมาชิกคืนที่เดิมจะได้จำนวนตัวอย่าง $= N^2 = 9$ และเมื่อหาค่าต่าง ๆ ก็ได้ผลดังตาราง 1.6.1 การใส่คืนที่เดิม หมายถึงว่า มีโอกาสที่จะถูกสุ่มมาใหม่ จนได้ตัวอย่างที่มีสมาชิกเหมือนกัน เช่น ตัวอย่างในตาราง 1.6.1

ตาราง 1.6.1 ผลจากการสุ่มแบบใส่คืนที่เดิมของประชากรที่มีสมาชิก 3 ค่า คือ 1, 8, 9

ตัวอย่างที่	ข้อมูล	$\bar{x}$	s^2	s
1	1 , 1	1.0	0	0
2	1 , 8	4.5	24.5	4.950
3	1 , 9	5.0	32.0	5.657
4	8 , 1	4.5	24.5	4.950
5	8 , 8	8.0	0	0
6	8 , 9	8.5	0.5	0.707
7	9 , 1	5.0	32.0	5.657
8	9 , 8	8.5	0.5	0.707
9	9 , 9	9.0	0	0
รวม		**54.0**	**114.0**	**22.628**
เฉลี่ย		**6.0**	**12.67**	**2.514**

จากผลรวมของค่าต่าง ๆ ในตาราง เมื่อหาค่าเฉลี่ยโดยหารด้วย 9 ก็จะได้ค่าเฉลี่ยของ $\overline{X}$, s^2 ที่เท่ากับ μ และ σ^2 พอดี จึงจัดได้ว่าทั้ง $\overline{X}$ และ s^2 ที่หาโดยใช้ df เป็นตัวหาร เป็นค่าประมาณที่ไม่ลำเอียงของ μ และ σ^2 ตามลำดับ แต่ค่าเฉลี่ยของ s เป็นค่าประมาณที่ลำเอียงของ σ อยู่นั่นเอง

ในตอนนี้อาจสรุปเพิ่มเติมเกี่ยวกับค่าต่าง ๆ ที่ได้ศึกษามาแล้วดังนี้ ค่าเฉลี่ย วาเรียนซ์ และส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง $(\overline{X}, s^2, s)$ เรียกว่าสถิติ[55] ส่วนค่าเหล่านี้ของประชากร (μ, σ^2, σ) เรียกว่า**พารามิเตอร์**[56]

ค่าวัดความคลาดเคลื่อนอีกค่าหนึ่งคือ **ความคลาดเคลื่อนมาตรฐาน**[57] ซึ่งอาจเขียนย่อว่า S.E. ใช้เป็นค่าแสดงขอบเขตของค่าเฉลี่ยหลาย ๆ ค่า เช่น ให้นักศึกษา 10 คน วัดความยาวของฝักข้าวโพดคนละ 12 ฝัก แล้วหาค่าเฉลี่ย ก็ได้ค่าเฉลี่ยมา 10 ค่า ค่าเฉลี่ยเหล่านี้มักไม่เท่ากัน ถ้านับว่าค่าเฉลี่ยเหล่านี้ เป็นตัวแปรชุดใหม่ ก็สามารถหาส่วนเบี่ยงเบนมาตรฐานได้โดยใช้สมการที่กล่าวมาแล้ว (สมการ 1-5) อย่างไรก็ดี เราอาจหาความคลาดเคลื่อนมาตรฐานได้โดยตรง จากการถอดรากสองของวาเรียนซ์ ที่หารด้วยขนาดของตัวอย่าง ดังนี้

$$\begin{aligned} \text{S.E.} &= \sqrt{\frac{\sum (X_i - \overline{X})^2}{n(n-1)}} \\ &= \sqrt{\frac{s^2}{n}} \end{aligned} \qquad \text{...(1-9)}$$

เช่นในกรณีของน้ำหนัก 100 เมล็ด ของถั่วเหลืองก็หาได้ว่า

$$\begin{aligned} \text{S.E.} &= \sqrt{\frac{0.46}{5(4)}} \\ &= 0.152 \text{ กรัม/100 เมล็ด} \end{aligned}$$

ถ้าเราใช้ค่านี้ไปบวกหรือหักลบจากค่าเฉลี่ย ก็ใช้บอกได้ว่าขอบเขตของค่าเฉลี่ยมีค่าเท่าใด และเราใช้ค่านี้สำหรับหาว่าค่าเฉลี่ยขนาดนั้นขนาดนี้ มีโอกาสเกิดขึ้นมากน้อยเท่าใด

1.7 สัมประสิทธิ์ของความปรวนแปร

สัมประสิทธิ์ของความปรวนแปร[58] คือค่าแปลง ซึ่งได้จากการนำส่วนเบี่ยงเบนมาตรฐานไปหารด้วยค่าเฉลี่ยและคูณด้วย 100 เขียนย่อว่า CV (ย่อมาจาก coefficient of variation) และหาได้จากสมการ

$$\text{CV}(\%) = \frac{s}{\overline{x}} \times 100 \qquad \text{...(1-10)}$$

เมื่อ s และ $\overline{X}$ คือส่วนเบี่ยงเบนมาตรฐานและค่าเฉลี่ยตามลำดับ เช่น กรณีของน้ำหนัก 100 เมล็ดของถั่วเหลืองหาได้ว่า CV = (0.339/16.2) x 100 = 2.09 เปอร์เซ็นต์

คุณสมบัติที่น่าสนใจของสัมประสิทธิ์ของความปรวนแปร

(1) ในสิ่งมีชีวิตหนึ่ง ๆ หรือข้อมูลชนิดหนึ่ง ๆ จะมี CV ที่คงที่หรือมีช่วงที่แน่นอน ไม่ว่าจะวัดข้อมูลนั้นจากที่ใดก็ตาม เช่น ผลผลิตของถั่วเหลืองมี CV = 12 ถึง 18 เปอร์เซ็นต์ ขนาดเมล็ดถั่วเขียวมี CV = 4 ถึง 6 เปอร์เซ็นต์ ขอให้สังเกตว่า ถึงแม้การทดลองแต่ละครั้งได้ค่าเฉลี่ยไม่เท่ากัน แต่ CV ค่อนข้างคงที่ เพราะถ้าหากการทดลองมีค่าเฉลี่ยสูง ส่วนเบี่ยงเบนมาตรฐานก็สูงตามไปด้วย จึงทำให้ผลหารของค่าทั้งสองออกมาค่อนข้างคงที่ การทดลองใดก็ตาม ถ้าให้ CV สูงกว่าที่ควรจะเป็นมาก ความเชื่อมั่นก็ยิ่งน้อยลง ดังนั้นเราจึงอาจใช้ CV เป็นตัววัดหรือตัวชี้ความถูกต้องเที่ยงตรงของแต่ละการทดลองได้ทางหนึ่ง

(2) ค่า CV จะสูงหรือต่ำ จะมีช่วงปรวนแปรมาก-น้อย ขึ้นอยู่กับชนิด ของพืช-สัตว์ทดลอง ชนิดของลักษณะที่ศึกษา ลักษณะที่มีความปรวนแปรตามสภาพแวดล้อมมาก เช่น ผลผลิตหรือความสูงของต้นพืช น้ำหนักตัวของสัตว์ ฯลฯ ย่อมจะให้ CV สูง แต่ลักษณะที่ปรวนแปรน้อย ๆ เช่น ขนาดเมล็ดหรือเปอร์เซ็นต์น้ำมันของถั่วเหลือง ย่อมให้ CV ต่ำ ดังนี้เป็นต้น

เราอาจใช้ CV หาขนาดแปลงทดลองของพืช ในการทดลองทางพืชหลายชนิดพบว่า เมื่อแปลงทดลองค่อย ๆ มีขนาดใหญ่ขึ้น CV ก็จะลดลงอย่างช้า ๆ เราอาจเลือกใช้ขนาดของแปลงที่เหมาะสม เพื่อได้ผลทั้งทางประหยัด และได้การทดลองที่มีประสิทธิภาพ

1.8 แบบฝึกหัด

1. ในเวลาสามชั่วโมงปรากฏว่า เซลล์ของแบคทีเรียเพิ่มจาก 1,000 เซลล์ เป็น 4,000 เซลล์ จงคำนวณหาเปอร์เซ็นต์เพิ่มเฉลี่ยต่อชั่วโมง

2. นายแดงมีเงินอยู่ 300 บาท เมื่อเข้าไปร้านค้าร้านแรก เขาใช้เงินไป 100 บาท เพื่อซื้อเสื้อตัวละ 20 บาท ในร้านที่สองจ่ายไป 100 บาท เพื่อซื้อเสื้อตัวละ 25 บาท และในร้านที่สามจ่ายไปอีก 100 บาท เพื่อซื้อเสื้อตัวละ 50 บาท จงคำนวณหาราคาเฉลี่ยของเสื้อจากทุกร้าน

3. จงหาวาเรียนซ์ของชุดตัวอย่างดังต่อไปนี้

   ก. 3, 5, 6, 4, 2, 7, 1

   ข. 13, 15, 16, 14, 12, 17, 11

   ค. 30, 50, 60, 40, 20, 70, 10

4. จงหาส่วนเบี่ยงเบนมาตรฐานของตัวอย่างที่มีค่าดังนี้ 1, 1, 1, 1, 1, 0, 0, 0, 0 โดยใช้วิธีปกติซึ่งแสดงในสมการ (1-8) และถ้าสมมุติว่าอัตราส่วนของ 1 มีค่าเป็น p และอัตราส่วน 0 มีค่าเป็น q จงแสดงให้เห็นว่า $s = \sqrt{pq}$

5. จงพิสูจน์ให้เห็นว่า

ก. $\sum(X_i - \overline{X}) = 0$

ข. $\sum(X_i - \overline{X})^2 = \sum X_i^2 - \frac{(\sum X)^2}{n}$

ค. $\text{Var } kX_i = k^2 \text{ Var } X_i$

ง. $\text{Var } (k + X_i) = k + \text{Var } X_i$

คำในบท

(1) statistics, (2) data, (3) descriptive statistics, (4) central tendency, (5) mean, (6) median, (7) dispersion, (8) range, (9) mean deviation, (10) variance, (11) standard deviation, (12) statistical inference (inductive statistics), (13) population, (14) finite population, (15) infinite population, (16) sample, (17) variable, (18) sampling, (19) random sampling, (20) sampling with replacement, (21) sampling without replacement, (22) simple random sampling, (23) presentation of data, (24) tabular presentation, (25) frequency distribution, (26) histogram, (27) frequency polygon, (28) ogive, (29) frequency curve, (30) normal distribution, (31) skewness, (32) multimodal, (33) central tendency, (34) mean (arithmetic mean), average, (35) geometric mean, (36) log, (37) antilogarithm, (38) geometric gene action (multiplicative), (39) harmonic mean, (40) median, (41) mode, (42) quartile, (43) decile, (44) percentile, (45) dispersion, (46) range, (47) mean deviation, (48) absolute value, (49) variance and standard deviation, (50) sum of squares, (51) mean square, (52) degree of freedom (df), (53) unbiased estimate, (54) sampling with replacement, (55) statistic, (56) parameter, (57) standard error of mean, (58) coefficient of variation (CV).

บทที่ 2
การกระจาย

2.1 คำนำ

การกระจาย[1] จัดเป็นธรรมชาติของการเกิดข้อมูลชนิดต่าง ๆ ทั้งนี้เมื่อข้อมูลแต่ละชนิดประกอบด้วยตัวแปรมากมาย และมีขนาดหรือค่าต่าง ๆ กัน การปรากฏของตัวแปรเหล่านั้นมีความถี่มากน้อยตามแต่โอกาสหรือความน่าจะเป็น คือในช่วงหนึ่ง ๆ ข้อมูลที่มีค่าหนึ่งอาจจะปรากฏมากกว่าในช่วงอื่น ทั้งนี้เพราะในช่วงดังกล่าวนั้นมีข้อมูลค่านั้นมากกว่าข้อมูลค่าอื่น ๆ นั่นเอง เมื่อพิจารณาจากข้อมูลทั้งหมดแล้ว โอกาสของข้อมูลจะเคลื่อนไหวแบบต่าง ๆ กันไป ทั้งนี้แล้วแต่ลักษณะหรือชนิดของการทดลอง ลักษณะของการเคลื่อนไหวของข้อมูลเราเรียกว่า การกระจาย

2.2 ความน่าจะเป็น

ความน่าจะเป็นหรือบางครั้งเรียกว่าโอกาส[2] หมายถึงอัตราส่วนของเหตุการณ์ชนิดหนึ่งต่อเหตุการณ์ทั้งหมด เช่น จากไพ่สำรับหนึ่ง ซึ่งมี 52 ใบ มีไพ่หน้าต่าง ๆ สี่ชนิด ชนิดละ 13 ใบ ดังนั้นไพ่หน้าใดก็ตามมีความน่าจะเป็น 13/52 หรือ 1/4 ดังนี้ เป็นต้น ความน่าจะเป็นมีบทบาทสำคัญอย่างยิ่งต่อการกระจายชนิดต่าง ๆ ข้อมูลใดมีความน่าจะเป็นสูง ก็จะเป็นข้อมูลที่ปรากฏเป็นจำนวนมากครั้งกว่าข้อมูลที่มีความน่าจะเป็นต่ำ

ความน่าจะเป็นเราแทนด้วย P เช่นเหตุการณ์ [3]ทั้งหมดมี N ครั้ง มีการเกิดเหตุการณ์ย่อย A จำนวน a ครั้ง ดังนั้นความน่าจะเป็นของเหตุการณ์ A คือ $P(A) = a/N$ และความน่าจะเป็นส่วนที่ไม่เป็นของเหตุการณ์ A (เรียกว่าเหตุการณ์ B) คือ $P(B) = (N-a)/N$ ตัวอย่างเช่น เราโยนเหรียญบาท 100 ครั้ง สมมุติว่าได้ด้านหัว (H) จำนวน 48 ครั้ง ดังนั้น $P(H) = 48/100$ และความน่าจะเป็นของด้านก้อย (T) คือ $P(T) = (100-48)/100$ หรือหาได้ว่า $P(T) = 1 - P(H)$

ตัวอย่างข้างบนนี้ให้ความจริง 2 อย่างคือ (1) ความน่าจะเป็นของแต่ละเหตุการณ์มีค่าระหว่าง 0 และ 1 เช่น ในการโยนเหรียญ 2 ครั้ง ถ้าไม่ได้ด้านหัวเลย ก็หาได้ว่า $P(H) = 0$ ถ้าได้ด้านหัว 1 ครั้ง ก็ให้ $P(H) = 1/2$ ถ้าได้ด้านหัวทั้ง 2 ครั้ง ก็ได้ $P(H) = 2/2 = 1$ และ (2) ผลบวกของความน่าจะเป็นของทุก ๆ เหตุการณ์ เท่ากับ 1 เสมอ เช่น ในกรณีของการโยนเหรียญเราก็ได้ $P(H) + P(T) = 1$ ในเรื่องของความน่าจะเป็นนั้น จำนวนเหตุการณ์ทั้งหมดของการทดลองครั้งหนึ่ง ๆ เรียกว่า ขอบเขตของข้อมูล[4] เช่น ในการโยนเหรียญ 2 ครั้ง อาจมีขอบเขตเป็น HH, HT, TH, TT ซึ่งหมายถึง การที่

เหรียญขึ้นด้านหัวทั้ง 2 ครั้ง, ด้านหัวขึ้นก่อนแล้วด้านก้อยขึ้นตาม ฯลฯ ตามลำดับ จากตัวอย่างของขอบเขตข้อมูลนี้เราอาจให้หลัก 2 อย่าง แก่ความน่าจะเป็น ดังนี้คือ

(1) กฎในการบวก[5] ในการสุ่มตัวอย่าง เราอาจจะวางข้อกำหนดแบบมีทางเลือกคือ ถ้าไม่เกิดเหตุการณ์หนึ่งก็อาจเกิดอีกเหตุการณ์หนึ่ง เช่น จากเมล็ดถั่วเหลือง 100 เมล็ด มีเมล็ดแตก 10 เมล็ด เมล็ดเป็นโรค 5 เมล็ด เพราะฉะนั้นถ้าเราหยิบมา 1 เมล็ด โอกาสที่จะได้เมล็ดแตกเท่ากับ 1/10 และโอกาสที่จะได้เมล็ดเป็นโรคเท่ากับ 1/20 แต่ถ้าเราให้มีทางเลือกคือ โอกาสที่เมล็ดแตกหรือเป็นโรค ก็หาได้โดยการบวกความน่าจะเป็นทั้งสองชนิดนี้เข้าด้วยกัน

$$P(\text{เมล็ดแตกหรือเป็นโรค}) = 1/10 + 1/20 = 3/20$$

ซึ่งเห็นว่าโอกาสที่มีทางเลือกมีขนาดสูงกว่าโอกาสใดโอกาสหนึ่งเดี่ยว ๆ ถ้าให้สมการนี้ใช้ได้ทั่วไปก็อาจปรับปรุงดังนี้

$$P(A \text{ หรือ } B \text{ หรือ } \dots N) = P(A) + P(B) + \dots + P(N) \qquad \dots(2\text{-}1)$$

จงสังเกตว่าการเกิดเหตุการณ์ทั้ง 2 ชนิด คือเมล็ดแตกและเมล็ดเป็นโรคดังตัวอย่าง เป็นเหตุการณ์ที่เกิดขึ้นคนละครั้งของการสุ่ม[6] เพราะอาการเมล็ดแตกและเมล็ดเป็นโรค ไม่ปรากฏอยู่ด้วยกันหรือไม่เกิดในเมล็ดเดียวกันก็ได้ ดังนั้นในกรณีเช่นนี้เราอาจสร้างเหตุการณ์ขึ้นมาใหม่ โดยการบวกความน่าจะเป็นของแต่ละเหตุการณ์ย่อย ๆ เข้าด้วยกัน

(2) กฎในการคูณ[7] เมื่อเรามีเหตุการณ์มาก ๆ เราอาจวางข้อกำหนดว่า ให้เหตุการณ์หนึ่งเกิดขึ้นแล้วอีกเหตุการณ์หนึ่งเกิดตาม หรือเหตุการณ์หนึ่งเกิดพร้อมกับอีกเหตุการณ์หนึ่ง ข้อกำหนดนี้เป็นการสร้างเหตุการณ์ใหม่ ซึ่งมีความน่าจะเป็น หรือโอกาส เท่ากับการนำเอาความน่าจะเป็นของเหตุการณ์ที่เป็นอิสระกันมาคูณกัน เช่นการการสุ่มเมล็ดถั่วเหลืองมา 1 เมล็ด โอกาสที่จะได้เมล็ดแตก และเป็นโรคด้วยเท่ากับ (1/10) x (1/20) = 1/200 หรือในการโยนเหรียญ 2 ครั้ง โอกาสที่จะได้ด้านหัวทั้งสองครั้งเท่ากับ (1/2) x (1/2) = 1/4 ถ้าจะให้ใช้ได้ทั่วไปก็ใช้สมการจากการคูณความน่าจะเป็นย่อย ๆ ดังนี้

$$P(A \text{ และ } B \text{ และ } \dots \text{ และ } N) = P(A)\,P(B) \dots P(N) \qquad \dots(2\text{-}2)$$

2.3 การกระจายของตัวแปรเต็มหน่วย

ตัวแปรเต็มหน่วย[8] คือตัวแปรที่มีค่าเป็นตัวเต็ม เช่น 1, 2, 10, 15, 18, 21 ซึ่งไม่มีค่าใด ๆ หลังจุดทศนิยม ตัวแปรเช่นนี้ได้แก่ จำนวนด้านหัวด้านก้อยจากการโยนเหรียญ จำนวนหน้าคิงจากการดึงไพ่จากสำรับ จำนวนครั้งที่เกิดอุบัติเหตุบนท้องถนน ฯลฯ การกระจายของตัวแปรพวกนี้เป็นการกระจายที่ไม่ต่อเนื่อง[9] การกระจายแบบไม่ต่อเนื่องมีอยู่หลายชนิด แต่ที่เรารู้จักกันดีคือการกระจายทวินาม[10] ซึ่งเป็นการกระจายของเหตุการณ์ที่มีวิธีการเกิดเพียง 2 ชนิด (ทวิ = สอง) เช่น ในการโยนเหรียญก็มี

การเกิดเพียง 2 ทาง คือด้านหัวและด้านก้อย อย่างไรก็ดีเหตุการณ์ชนิดอื่นอาจทำให้เป็น 2 ทาง ก็ได้ เช่น แม้ลูกเต๋าจะมี 6 หน้า แต่ถ้าบอกว่าเป็นการกระจายของหน้าคู่และหน้าคี่ ก็จะกลายเป็นการกระจายทวินาม

ในตอนนี้จะอธิบายเพิ่มเติมถึงการกระจายทวินาม ตัวอย่างเช่น ในการโยนเหรียญบาท 3 เหรียญ 1 ครั้ง พบว่า อาจเกิดได้อย่างมาก 8 วิธี เป็นลำดับ คือ (1) ขึ้นด้านหัวทุกเหรียญ (2) ขึ้นด้านหัว 2 เหรียญ ด้านก้อย 1 เหรียญ (3) ขึ้นด้านหัว 1 เหรียญ ด้านก้อย 2 เหรียญ และ (4) ขึ้นด้านก้อยทุกเหรียญ ดังแสดงในตาราง 2.3.1 แต่เมื่อวิเคราะห์อย่างละเอียดพบว่า ถ้าขึ้นทั้งด้านหัวและด้านก้อย เช่น ด้านหัว 2 เหรียญ ด้านก้อย 1 เหรียญ มีการเกิด 3 วิธี โดยสรุปแล้วมีจำนวนวิธี 1 : 3 : 3 : 1 แต่ความน่าจะเป็นของแต่ละวิธีย่อย ๆ เท่ากับ a/N = 1/8 ดังนั้นความน่าจะเป็นในภาพรวมของการโยนเหรียญจำนวน 3 เหรียญ คือ $\frac{1}{8}:\frac{3}{8}:\frac{3}{8}:\frac{1}{8}$

ตาราง 2.3.1 ผลจากการโยนเหรียญบาท 3 อัน 1 ครั้ง, H=head (ด้านหัว), T=tail (ด้านก้อย)

วิธีการเกิด	เหรียญ 1	เหรียญ 2	เหรียญ 3	รวม	เหตุการณ์[1]	จำนวน
1	H	H	H	3 H : 0 T	3 H	1
2	H	H	T			
3	H	T	H	2 H : 1 T	2 H	3
4	T	H	H			
5	H	T	T			
6	T	H	T	1 H : 2 T	1 H	3
7	T	T	H			
8	T	T	T	0 H : 3 T	0 H	1

ถ้าให้ P(H) = p, P(T) = q และ P(H) = P(T) = 1/2 อาจแสดงการกระจายและตาราง 2.3.1 เป็นความน่าจะเป็นและการกระจายของความน่าจะเป็นดังตาราง 2.3.2 การกระจายดังกล่าวนี้ เป็นการกระจายไบโนเมียล การกระจายนี้เท่ากับการกระจายของการโยนเหรียญ 3 อัน 1 ครั้ง

การหาการกระจายโดยใช้วิธีในตาราง 2.3.1 และ 2.3.2 ต้องใช้เวลามาก และกระทำไม่ได้เมื่อเหตุการณ์มากขึ้น ดังนั้นอาจใช้สมการ (2-3) สมการนี้เรียกว่า สมการกระจายไบโนเมียล[11] โดยแทนด้วย X เช่น การโยนเหรียญ X = 0 ไม่ขึ้นด้านหัวเลย, X = 3 ขึ้นด้านหัว 3 เหรียญ ดังนั้น P(X = 0) คือ โอกาสที่ไม่ขึ้นด้านหัวเลย

$$P(X) = \underbrace{\frac{N!}{X!(N-X)!}}_{1} \underbrace{p^{X} q^{N-X}}_{2} \qquad \ldots(2\text{-}3)$$

ตาราง 2.3.2 การกระจายของความน่าจะเป็นของการโยนเหรียญ 3 อัน 1 ครั้ง

วิธีการเกิด	เหรียญ 1	เหรียญ 2	เหรียญ 3	ความน่าจะเป็น	ความน่าจะเป็น
1	p	p	p	p^3	$1\left(\frac{1}{2}\right)^3 = \frac{1}{8}$
2	p	p	q		
3	p	q	p	$3p^2q$	$3\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right) = \frac{3}{8}$
4	q	p	p		
5	p	q	q		
6	q	p	q	$3pq^2$	$3\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^2 = \frac{3}{8}$
7	q	q	p		
8	q	q	q	q^3	$1\left(\frac{1}{2}\right)^3 = \frac{1}{8}$

ในสมการนี้ ส่วนที่ 1 เป็นสูตรจัดหมู่ ใช้หาสัมประสิทธิ์ของไบโนเมียล[12] สูตรนี้ได้เรียนกันมาแล้วในวิชาคำนวณ ทั้งนี้ N คือจำนวนเหตุการณ์ X = 0, 1, 2, ..., N และ " ! " คือ แฟกทอเรียล[13] คือมีการคูณตั้งแต่ 1 ถึง N หรือถึง X

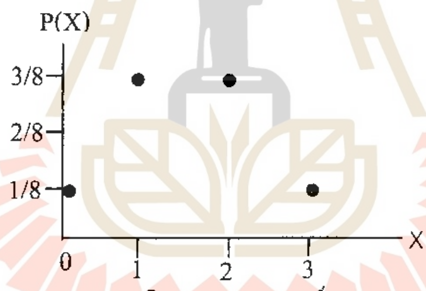
ส่วนที่สองของสมการใช้หาความน่าจะเป็น โดย p เป็นความน่าจะเป็นของเหตุการณ์หนึ่ง และ q เป็นความน่าจะเป็นของอีกเหตุการณ์หนึ่ง เช่น ในกรณีของการโยนเหรียญ ก็หาได้ว่า p = P (H) = 1/2 ; q = P(T) = 1/2 หรือ $q = 1 - p$

เมื่อใช้สมการ (2-3) เพื่อแก้โจทย์ข้างบนก็ให้การกระจายของเหตุการณ์ต่าง ๆ ดังนี้

$$P(X=0) = \frac{3!}{0!(3-0)!}(1/2)^0(1/2)^3$$
$$= \frac{3\times 2\times 1}{(1)(3\times 2\times 1)}(1)(1/8)$$
$$= 1/8$$

$$P(X=1) = \frac{3!}{1!(3-1)!}(1/2)^1(1/2)^2$$
$$= \frac{3\times 2\times 1}{(1)(2\times 1)}(1/2)(1/4)$$
$$= 3/8$$

ซึ่งอาจหาโดยวิธีคล้ายคลึงกันได้ว่า $P(X = 2) = 3/8$ และ $P(X = 3) = 1/8$ นั่นก็คือความน่าจะเป็นของการได้ด้านหัว 0, 1, 2 และ 3 ครั้ง ของตัวอย่างนี้เท่ากับ 1/8, 3/8, 3/8 และ 1/8 ตามลำดับ จึงแสดงการกระจายข้างบนดังรูป 2.3.2



รูป 2.3.2 การกระจายของเหตุการณ์ในการโยนเหรียญ 3 ครั้ง

ค่าเฉลี่ยของการกระจายไบโนเมียลหาได้จากสมการ

$$\mu = \sum X/N$$

ทั้งนี้ในตัวอย่างข้างบน $N = 1/8 + 3/8 + 3/8 + 1/8 = 1$, $\Sigma X = \Sigma [X][P(X)] = (0)(1/8) + (1)(3/8) + (2)(3/8) + (3)(1/8) = 12/8$ ดังนั้น

$$\mu = (12/8)/1$$
$$= 1.5$$

ส่วนวาเรียนซ์หาได้จากสมการ

$$\sigma^2 = \sum (X-\mu)^2 P(X)$$

ซึ่งอาจพิสูจน์ได้ว่า

$$\mu \;=\; Np \text{ และ } \sigma^2 = Npq$$

เช่นในกรณีของตัวอย่างของเราซึ่งมี $N = 3$, $p = q = 1/2$ ก็หาได้ว่า $\mu = 1.5$ และ $\sigma^2 = 0.75$

นอกจากการกระจายทวินามแล้ว การกระจายของตัวแปรเต็มหน่วยยังมีได้อีกหลายแบบ ได้แก่ การกระจาย Bernoulli, การกระจายไฮเปอร์จีโอเมตริก[14], การกระจายของอัตราส่วน[15], การกระจายพหุนาม[16] และการกระจายปัวซอง[17] แต่จะไม่อธิบายโดยละเอียดในที่นี้

2.4 การกระจายของตัวแปรต่อเนื่อง

ตัวแปรต่อเนื่อง[18] ได้แก่ตัวแปรที่มีค่าไม่ลงตัว คือมีค่าติดต่อถึงหลังจุดทศนิยม ตัวแปรเหล่านี้ได้แก่ความสูง น้ำหนัก ปริมาตร ความกว้าง ความหนา ความลึก ฯลฯ ซึ่งเป็นข้อมูลที่ได้จากการชั่ง ตวง และวัด เช่น ถ้าเราสังเกตความยาวของฝักข้าวโพด จำนวน 100 ฝัก จะไม่มีฝักใดเลยที่มีความยาวที่ลงตัวเป็นเซ็นติเมตร นอกจากเราวัดอย่างหยาบ ๆ เท่านั้น ถ้าสังเกตข้อมูลที่ปรวนแปรแบบนี้มาก ๆ จะเห็นได้ว่ามีขนาดลดหลั่นติดต่อกันไป ถ้าวาดเป็นรูปกราฟก็จะให้เส้นติดต่อกันไป การกระจายแบบนี้เรียกว่าการกระจายแบบต่อเนื่อง[19] ความน่าจะเป็นของแต่ละเหตุการณ์ในการกระจายแบบนี้ต้องบอกกันเป็นช่วงหรือระยะ เช่น อาจให้หาความน่าจะเป็นของความยาวฝักข้าวโพดที่ยาวระหว่าง 14.50 – 15.50 ซม. ซึ่งอาจเขียนว่า

$$P(14.50 < X < 15.50) \;=\; a$$

เมื่อร่นระยะเข้าไปชิด X ครึ่งหนึ่งของระยะเดิม ความน่าจะเป็นก็จะลดลงครึ่งหนึ่งเช่นกัน คือได้

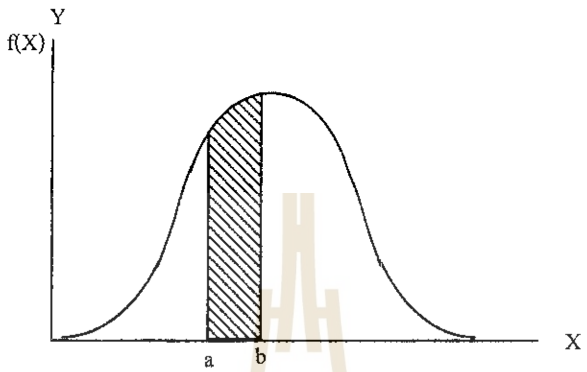
$$P(14.75 < X < 15.25\,) \;=\; a/2$$

ถ้าลดลงไปอีกจนเป็นค่าเดี่ยว ก็จะทำให้ได้ความน่าจะเป็นที่มีค่าเป็นศูนย์

$$P(X \;=\; 15.00) = 0$$

คือแสดงว่าการได้ข้าวโพดยาว 15.00 ซม. นั้น มีความน่าจะเป็นเท่ากับศูนย์นั่นเอง

จากที่อธิบายมานี้เห็นได้ว่า การบอกความน่าจะเป็นของตัวแปรต่อเนื่องนั้น ต้องบอกกันเป็นช่วงหรือระหว่าง จะบอกเป็นค่าโดด ๆ ไม่ได้ ตัวอย่างการกระจายแบบนี้แสดงไว้ในรูป 2.4.1 เนื้อที่ใต้ส่วนโค้งนี้มีค่ารวมกันเท่ากับ 1 ความน่าจะเป็นของแต่ละเหตุการณ์ เช่น ระหว่าง a และ b จะเป็นสัดส่วนหรือฟังก์ชันของเนื้อที่[20] หรือเขียนว่า $f(X) = P(a < X < b)$ เมื่อ a และ b เป็นจุด 2 จุด บนแกน X



รูป 2.4.1 การกระจายต่อเนื่อง

การกระจายต่อเนื่องชนิดที่ควรทราบมีดังนี้

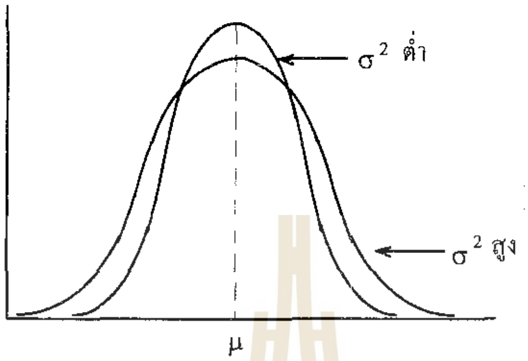
(1) การกระจายปกติ การกระจายแบบปกติ[21] นับว่าเป็นการกระจายที่มีความสำคัญอย่างยิ่ง ค่าสังเกตหรือข้อมูลในชีววิทยามักมีการกระจายแบบนี้ เช่น ผลผลิตของพืช น้ำหนักของคน-สัตว์ ความสูงของต้นพืช ความเฉลียวฉลาดของคน หรือแม้แต่ข้อมูลจากการทดลองหรือการทดสอบบางอย่าง เช่น คะแนนในการสอบวิชาสถิติของนักศึกษา การต้านทานต่อแรงดึงของเส้นลวด อายุการใช้งานของหลอดไฟฟ้า ฯลฯ การกระจายปกติจะมีลักษณะดังต่อไปนี้คือ **มีรูปร่างคล้ายระฆังคว่ำ ด้านซ้ายและด้านขวามีพื้นที่เท่ากัน เรียกว่ามีรูปร่างแบบสมมาตร[22] ค่าเฉลี่ยอยู่ตรงกลางโค้งกระจาย และหางทั้งสองข้างไม่มีจุดจบหรือถึงอนันต์**

รูปร่างของการกระจายแบบปกติขึ้นอยู่กับวาเรียนซ์ ถ้าวาเรียนซ์มีค่ามาก การกระจายของข้อมูลก็มีความแผ่กว้างมาก แต่ถ้าวาเรียนซ์มีค่าน้อยการกระจายก็แคบ และโค้งกระจายก็จะสูง ดังรูป 2.4.2 และโค้งกระจายก็จะเคลื่อนที่ตามค่าเฉลี่ยดังรูป 2.4.3 คือเมื่อค่าเฉลี่ยเคลื่อนที่ไปทางด้านขวา โค้งกระจายก็จะเคลื่อนตามไป ดังนั้นการกระจายปกติจึงปรวนแปรตาม μ และ σ^2

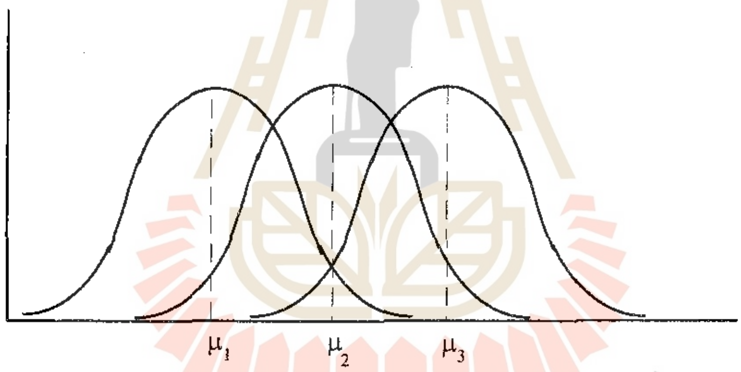
ตามทฤษฎีแล้ว ปลายโค้งของการกระจายแบบปกติไม่มีจุดสิ้นสุด มันจะวิ่งไปใกล้ ๆ กับแกน X แต่ไม่เคยสัมผัสกับแกน X ในทางทฤษฎีเราถือว่าเนื้อที่ส่วนปลาย ๆ มีค่าน้อยมากจนอาจตัดทิ้งก็ได้ เมื่อให้ μ เป็นจุดศูนย์กลาง ก็อาจประมาณพื้นที่ทั้งสองด้านดังนี้

$$P(\mu - 3\sigma < X < \mu + 3\sigma) \;=\; 99.73 \text{ เปอร์เซ็นต์}$$
$$P(\mu - 2\sigma < X < +2\sigma) \;=\; 95.45 \text{ เปอร์เซ็นต์}$$
$$P(\mu - 1\sigma < X < \mu + 1\sigma) \;=\; 68.27 \text{ เปอร์เซ็นต์}$$

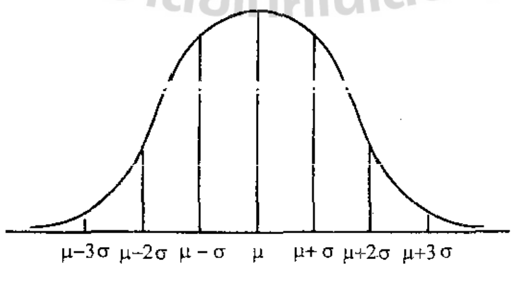
ซึ่งหมายถึงว่า ค่าสังเกต 99.73, 95.45 และ 68.27 เปอร์เซ็นต์ อยู่ในส่วนที่ห่างจาก μ ไปทั้งสองด้านจำนวน $3\sigma, 2\sigma$ และ 1σ ตามลำดับดังแสดงในรูป 2.4.4



รูป 2.4.2 การกระจายปกติที่มี μ เหมือนกัน แต่ σ^2 ต่างกัน



รูป 2.4.3 การกระจายแบบปกติที่มี σ^2 เหมือนกัน แต่ μ ต่างกัน



รูป 2.4.4 การกระจายแบบปกติ

(2) การกระจายคะแนนมาตรฐาน คะแนนมาตรฐาน[23] คือค่าแปลงของคะแนนดิบโดยใช้สมการ

$$Z = \frac{X - \mu}{\sigma} \qquad \ldots(2\text{-}4)$$

เมื่อมีตัวแปรมาก ๆ ก็ได้การกระจายของคะแนนมาตรฐานที่มี $\mu = 0$ และ $\sigma^2 = 1$ ซึ่งมีลักษณะการกระจายแบบปกติ รูปร่างสมมาตร เนื้อที่ภายใต้โค้งเท่ากับ 1 เนื้อที่ 68.27, 95.45 และ 99.73 เปอร์เซ็นต์ อยู่ห่างจากค่าเฉลี่ย ± 1, 2 และ ± 3 คะแนนมาตรฐาน ค่าบางค่าที่ควรทราบคือ เนื้อที่ 99 และ 95 เปอร์เซ็นต์ อยู่ระหว่างคะแนนมาตรฐาน (Z) $\pm$ 2.58 และ $\pm$1.96 ตามลำดับ

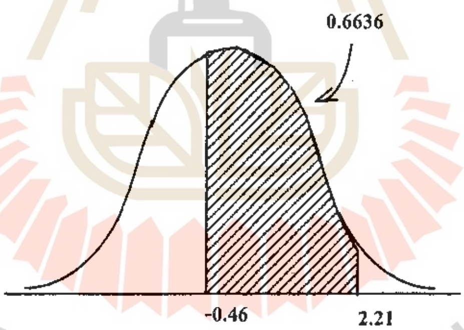
ตารางที่ 2 ของภาคผนวกแสดงพื้นที่ของคะแนนมาตรฐานต่าง ๆ จากค่า Z ค่าเดียว เราสามารถหาพื้นที่ทางด้านขวาหรือซ้ายของรูปกระจายแบบปกติก็ได้ ถ้ามี Z จำนวน 2 ค่า ก็หาพื้นที่ระหว่างกลางได้ ซึ่งแสดงวิธีหาพอเป็นตัวอย่างดังนี้ :

ตัวอย่าง **2.4.1**

จงหาพื้นที่ระหว่าง Z = $-$0.46 และ Z = 2.21

วิธีทำ

พื้นที่ดังกล่าวนี้แสดงในรูปดังนี้



ซึ่งเกิดจากการรวมพื้นที่ระหว่าง Z = $-$0.46 กับ Z = 0 และระหว่าง Z = 0 กับ Z = 2.21 ซึ่งเท่ากับ 0.1772 + 0.4864 = 0.6636

ตัวอย่าง **2.4.2**

ตัวแปรชุดหนึ่งมี μ = 24 และ σ = 12 จงหาความน่าจะเป็นของค่าระหว่าง 17.4 และ 58.8

วิธีทำ **ในกรณีเช่นนี้เราเขียนสมการว่า P(17.4 < X < 58.8) = ? ซึ่งก็หาตามขั้นตอนได้ดังนี้**

$$P(17.4 < X < 58.8) = P(Z_1 < Z < Z_2) = ?$$

$$Z_1 = \frac{X - \mu}{\sigma} = \frac{17.4 - 24}{12} = -0.55$$

$$Z_2 = \frac{58.8 - 24}{12} = 2.90$$

ดังนั้น $P(17.4 < X < 58.8) = P(-0.55 < Z < 2.90)$ **หรือเท่ากับเนื้อที่ระหว่าง** $Z = -0.55$ **และ** $Z = 2.90$ **ซึ่งเท่ากับ** $0.2088 + 0.4981 = 0.7069$

จงสังเกตว่าพื้นที่หรือความน่าจะเป็นไม่มีค่าเป็นลบ เครื่องหมายบวก-ลบ ของค่า Z เป็นเพียงการระบุว่าค่านั้นอยู่ด้านซ้ายหรือด้านขวาเท่านั้น

ตารางมาตรฐานใช้ได้กับข้อมูลที่มีการกระจายแบบต่อเนื่องเท่านั้น แต่ถ้าข้อมูลมีค่าแบบไม่ต่อเนื่อง เราก็อาจแก้ไขโดยการบวกหรือลบด้วย 0.5 เช่น

ตัวอย่าง 2.4.3

ในการสำรวจขนาดของฟาร์มสุกรต่างๆ ปรากฏว่าได้ $\mu = 100$ ตัว และ $\sigma = 10$ ตัว จงหาโอกาสที่มีฟาร์มซึ่งมีสุกร 120 ตัว หรือมากกว่า

วิธีทำ **ในตัวอย่างนี้ข้อมูลได้จากการนับ จึงจัดเป็นตัวแปรเต็มหน่วย ความน่าจะเป็นของการมีฟาร์มขนาด 120 ตัว นั้น ควรนับจากฟาร์มที่มีขนาด 119.5 ตัว คือ**

$Z = (119.5 - 100)/10 = 1.95$

ดังนั้น $P(X > 120) = P(Z > 1.95)$

$= 0.5000 - 0.4744$

$= 0.0256$

ดังนั้นโอกาสที่มีฟาร์มซึ่งมีสุกร 120 ตัว หรือมากกว่าเท่ากับ 0.0256

ประโยชน์อันหนึ่งของการกระจายปกติ คือใช้ประมาณการกระจายทวินาม โดยประมาณได้ใกล้เคียงเมื่อ N มีค่าสูง และ p หรือ q ใกล้ 0.5 หรือเมื่อ Np หรือ Nq มากกว่า 5

ตัวอย่าง 2.4.4

ในการโยนเหรียญ 12 ครั้ง จงคำนวณความน่าจะเป็นของการได้ด้านหัว 4 ครั้ง โดยใช้วิธีการหาการกระจายทวินาม และประมาณจากการกระจายปกติ

วิธีทำ **เมื่อใช้วิธีการกระจายทวินามก็จะได้**

$P(X = 4) = {}_{12}C_4(1/2)^4 (1/2)^8 = 495(1/2)^4 (1/2)^8 = 0.12$

เมื่อประมาณโดยใช้วิธีการกระจายแบบปกติ ก็หาได้ว่า $\mu = np = 12(1/2) = 6,$

$\sigma = \sqrt{Npq} = \sqrt{(12)(1/2)(1/2)} = \sqrt{3}$,

และ $P(X = 4) = P(3.5 < X < 4.5)$ ดังนั้น

$Z_1 = (3.5-6)/\sqrt{3} = -1.44$

$Z_2 = (4.5-6)/\sqrt{3} = -0.87$

ซึ่งเมื่อประมาณโดยวิธีการกระจายปกติก็พบว่า

$P(3.5 < X < 4.5) = P(-1.44 < Z < -0.87)$

$= 0.4251 - 0.3078$

$= 0.1173$

ซึ่งใกล้เคียงกับการหาโดยวิธีการกระจายทวินามมาก

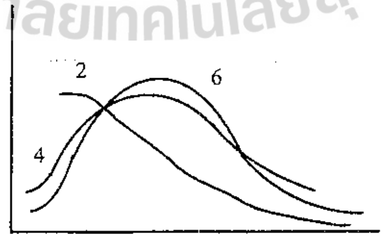
(3) การกระจายของไคสแควร์[24] ไคสแควร์ (χ^2) คือค่าที่หาได้โดยใช้สมการ

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \qquad \ldots(2\text{-}5)$$

ทั้งนี้ $(n-1)s^2 = \sum(X-\bar{X})^2$ เท่ากับผลบวกค่ากำลังสอง[25] ดังนั้นค่าไคสแควร์ คือผลบวกค่ากำลังสองหารด้วยวาเรียนซ์ของประชากรนั่นเอง ถ้าเราสุ่มตัวอย่างที่มีขนาดเท่ากับ n มา 1 ชุด ก็ได้ค่าไคสแควร์ 1 ค่า ดังนั้นถ้าสุ่มมามากชุดโดยมีขนาดเท่ากัน และจากประชากรเดียวกัน ก็ได้ค่าไคสแควร์มาก ๆ ค่า เมื่อนำค่าไคสแควร์เหล่านี้มาลงจุดกราฟ โดยให้แกน X เป็นขนาดของไคสแควร์ และให้แกน Y เป็นความถี่[26] ก็ได้การกระจายไคสแควร์ที่มี df = n – 1 รูปร่างของการกระจายจะเบ้ทางขวา จะเบ้มากหรือน้อยขึ้นอยู่กับขนาดของ df ดังรูป 2.4.6 ทั้งนี้อาจพิสูจน์ทางคณิตศาสตร์ได้ว่าค่าไคสแควร์

$$\chi^2 = \sum\frac{(o-e)^2}{e} \qquad \ldots(2\text{-}6)$$

ในสมการนี้ o = ค่าสังเกต[27] และ e = ค่าคาดหมาย[28] สมการนี้ใช้ทดสอบความสอดคล้องระหว่างค่าสังเกตและค่าคาดหมาย[29]



รูปที่ 2.4.6 การกระจายของค่าไคสแควร์ที่มี df 2, 4, และ 6

(4) การกระจายของค่าเฉลี่ยของตัวอย่าง ถ้ามีประชากรอยู่ชุดหนึ่ง เราอาจสุ่มตัวอย่างได้มากมาย ตัวอย่างแต่ละชุดอาจทำการคำนวณหาค่าเฉลี่ยโดยใช้สมการ $X = \sum X / n$ เมื่อนำค่าเฉลี่ยเหล่านั้นมาหาการกระจาย โดยให้แกน X เป็นขนาดของค่าเฉลี่ย และแกน Y เป็นความถี่ก็จะได้การกระจายที่เรียกว่าการกระจายของค่าเฉลี่ยของตัวอย่าง[30] ซึ่งมีคุณสมบัติที่น่าสนใจคือ

(1) ถ้าคิดว่าค่าเฉลี่ยของแต่ละตัวอย่างคือตัวแปรชนิดหนึ่ง คือ $\overline{X}_1 = X_1, \overline{X}_2 = X_2, \ldots$ เมื่อคำนวณค่าเฉลี่ยของการกระจาย $(\mu_{\overline{X}})$ ก็จะได้ค่าใกล้เคียงค่าเฉลี่ยของประชากร (μ) อย่างมาก คือ ถ้าสุ่มได้ครบทุกตัวอย่างเท่าที่ประชากรนั้นพึงจะมี ก็ทำให้ $\mu_{\overline{X}} = \mu$ พอดี ดังนั้น $\overline{X}$ จึงเป็นค่าประมาณที่ไม่ลำเอียงของ μ

(2) เมื่อหาวาเรียนซ์ของค่าเฉลี่ยเหล่านี้โดยใช้สมการ μ

$$\sigma^2_{\overline{x}} = \sum (\overline{X} - \mu)^2 / m$$

ทั้งนี้ให้ m = จำนวนตัวอย่าง ซึ่งเมื่อคำนวณดูจริง ๆ ก็พบว่า $\sigma^2_{\overline{x}}$ มีค่าใกล้เคียงวาเรียนซ์ของประชากรหารด้วยขนาดของตัวอย่าง คือ $\sigma^2_{\overline{X}} = \sigma^2 / n$ วาเรียนซ์นี้เรียกว่าวาเรียนซ์ของค่าเฉลี่ย เมื่อถอดรากที่สองก็ได้ส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ย $(\sigma_{\overline{X}})$ ซึ่งเราเรียกว่าความคลาดเคลื่อนมาตรฐาน[31] ก็ได้ แต่ถ้าเราไม่ทราบ σ^2 ก็อาจประมาณจากตัวอย่างเดียวโดยใช้สมการ $s^2_{\overline{x}} = s^2 / n$ และ $s_{\overline{X}} = s / \sqrt{n}$ ดังที่กล่าวมาแล้วในบทที่ 1 นั่นเอง

ยกตัวอย่าง เช่น เรามีประชากรอยู่กลุ่มหนึ่ง แล้วสุ่มตัวอย่างขนาด n = 10 มา 1 ชุด ซึ่งมีข้อมูลเป็น 32, 31, 11, 30, 19, 24, 53, 44, 19 และ 30 ซึ่งหาได้ว่า $s^2 = 151.56$ ดังนั้น

$$s^2_{\overline{X}} = s^2 / n = 151.56 / 10 = 15.156$$

$$s_{\overline{X}} = s / \sqrt{n} = 12.3 / \sqrt{10} = 3.89$$

ดังนั้นจึงเห็นว่า ถึงแม้เราไม่ทราบความคลาดเคลื่อนมาตรฐาน ก็สามารถคำนวณได้จากส่วนเบี่ยงเบนมาตรฐานนั่นเอง

เราอาจให้ข้อมูลและความรู้เพิ่มเติม จากการสุ่มตัวอย่างจากประชากรดังนี้

(1) ในประชากรที่มีขนาดจำกัด ไม่ว่าจะมีการกระจายปกติหรือไม่ก็ตามที แต่ถ้าสุ่มตัวอย่างโดยใส่คืนที่เดิม ก็จะหาได้ว่า $\mu_{\overline{X}} = \mu$ และ $\sigma^2_{\overline{X}} = \sigma^2 / n$ แต่ถ้าไม่ใส่คืนที่เดิมก็หาได้ว่า $\mu_{\overline{X}} = \mu$ และ $\sigma^2_{\overline{X}} = (\sigma^2/n)(N - n / N - 1)$ เมื่อ N = ขนาดของประชากร และ n = ขนาดของตัวอย่าง

(2) การกระจายที่มีรูปร่างไม่เป็นแบบปกติ ถ้าสุ่มโดยคืนที่เดิม ค่าเฉลี่ยก็กระจายแบบปกติหรือใกล้เคียงปกติ ถ้าสุ่มโดยไม่คืนที่เดิมไม่แน่ใจว่าจะเป็นแบบปกติหรือไม่

(3) การกระจายที่มีรูปร่างแบบปกติ การสุ่มแบบคืนที่หรือไม่คืนที่เดิมก็ตาม แต่ค่าเฉลี่ยจะกระจายแบบปกติ

จากข้อสรุปข้างบนเห็นว่า ค่าเฉลี่ยของชุดตัวอย่างที่มีการกระจายแบบปกติ เมื่อแปลงค่าเฉลี่ยเหล่านี้เป็นคะแนนมาตรฐานคล้ายสมการ (2-4) ดังนี้

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} \qquad \text{...(2-7)}$$

ก็ได้คะแนนมาตรฐานที่กระจายแบบปกติ โดยมีค่าเฉลี่ย = 0 และวาเรียนซ์ = 1

สมการ (2-7) มีวิธีการใช้ประโยชน์คล้ายสมการ (2-4) เช่น ใช้หา $P(X > a)$ ถ้า a เป็นค่าอะไรก็ได้

ตัวอย่าง 2.4.5

ประชากรกลุ่มหนึ่งมีสมาชิกเป็น 1, 1, 1, 3, 4, 5, 6, 6, 6 และ 7 ถ้า สุ่มตัวอย่างขนาด n = 36 มาชุดหนึ่ง มีความน่าจะเป็นหรือโอกาสเท่าไรที่จะได้ค่าเฉลี่ยระหว่าง 3.9 ถึง 4.4 โดยให้ใกล้เคียงค่านี้ที่สุด

วิธีทำ จากประชากรนี้หาได้ว่า $\mu = 4$, และ $\sigma_{\overline{X}} = 0.37$ ความน่าจะเป็นใกล้เคียง 3.9 และ 4.4 คือ $P(3.85 < X < 4.45)$ ดังนั้น

$$Z_1 = (3.85 - 4) / 0.37 = -0.405$$

$$Z_2 = (4.45 - 4) / 0.37 = 1.216$$

ดังนั้น

$P(-0.405 < Z < 1.216)$

= เนื้อที่ระหว่าง Z = 0 และ Z = 1.216 บวกเนื้อที่ระหว่าง Z = 0 และ Z = − 0.405

= 0.3879 + 0.1578

= 0.5457

คือโอกาสที่จะได้ค่าเฉลี่ยระหว่าง 3.9 ถึง 4.4 เท่ากับ 0.5457

(5) การกระจายแบบที เมื่อสุ่มตัวอย่างที่มีขนาดเล็ก ($n < 30$) จากประชากรที่ไม่ทราบวาเรียนซ์ เราอาจคำนวณค่าคล้ายคะแนนมาตรฐานดังสมการ (2-7) โดยใช้ความคลาดเคลื่อนมาตรฐาน ซึ่งประมาณจากชุดตัวอย่างที่สุ่ม ก็ได้ค่าที่เรียกว่า ค่า t หรือสตูเดนต์ที[32] ดังสมการ

$$t = \frac{\overline{X} - \mu}{s_{\overline{X}}} \qquad \text{...(2-8)}$$

ซึ่ง $s_{\bar{x}} = s/\sqrt{n}$

เมื่อมีตัวอย่างขนาดเดียวกันมาก ๆ ถ้านำมาเขียนกราฟก็ได้การกระจายคล้าย ๆ การกระจายคะแนนมาตรฐาน มีค่าเฉลี่ย $=0$ และวาเรียนซ์ $=n/(n-2)$ คือวาเรียนซ์มากกว่า 1 เล็กน้อย ทั้งนี้ $n>2$ การกระจายนี้เรียกว่าการกระจายแบบที[33] มีรูปแผ่กว้างกว่าการกระจายของคะแนนมาตรฐานเล็กน้อย โดยเฉพาะอย่างยิ่งเมื่อชุดตัวอย่างมีขนาดเล็กการแผ่กว้างก็มีมาก เมื่อชุดตัวอย่างขนาดใหญ่ขึ้น การกระจายก็ใกล้เคียงคะแนนมาตรฐานขึ้น

ตารางที่ 4 ในภาคผนวกแสดงค่า t เนื่องจากการกระจายเป็นแบบสมมาตร ดังนั้นค่าในตารางอาจเป็นบวกหรือลบก็ได้ แถวบนสุดของตารางแสดงความน่าจะเป็นของค่า t ซึ่งเป็นความน่าจะเป็นของทั้งสองหาง เมื่อจะใช้สำหรับหางเดียวก็ให้ลดความน่าจะเป็นลงมาครึ่งหนึ่งดังตัวอย่างเช่นนี้

ที่ df = 10 ค่า t สูงกว่า 1.812 หรือต่ำกว่า – 1.812 มีความน่าจะเป็น 10 เปอร์เซ็นต์

ที่ df = 10 ค่า t สูงกว่า 1.812 มีความน่าจะเป็น 5 เปอร์เซ็นต์

ซึ่งจะกล่าวเพิ่มเติมถึงวิธีการใช้ตาราง t เมื่อพูดถึงการทดสอบสมมุติฐานต่อไป

(6) การกระจายค่าของเอฟ ค่า F เกิดจากอัตราส่วนของวาเรียนซ์ของสองประชากรหรือสองตัวอย่าง เช่น จากประชากรหนึ่งถ้าเราสุ่มตัวอย่างมา 1 ชุด ก็สามารถคำนวณค่าไคสแควร์โดยใช้สมการ

$$\chi^2 = (n_1 - 1)s_1^2/\sigma_1^2$$

ซึ่งคำนวณต่อไปได้ว่า

$$\chi^2/(n_1 - 1) = s_1^2/\sigma_1^2$$

ถ้าสุ่มมาอีกตัวอย่างหนึ่งก็คำนวณได้ว่า

$$\chi^2/(n_2 - 1) = s_2^2/\sigma_2^2$$

เมื่อนำอัตราส่วนทั้งสองกลุ่มมาหารกันก็ได้ค่า F คือ

$$F = \frac{\chi^2/(n_1-1)}{\chi^2/(n_2-1)}$$

$$= \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$$

ถ้า $\sigma_1^2 = \sigma_2^2$ ก็หาได้ว่า

$$F = \frac{s_1^2}{s_2^2} \qquad \ldots(2\text{-}9)$$

ทั้งนี้เมื่อ $s_1^2 > s_2^2$ ดังนั้นค่า F คืออัตราส่วนของวาเรียนซ์นั่นเอง ทั้งนี้ให้ชื่อว่าเป็นการกระจายแบบ F เพื่อเป็นเกียรติแก่ R. A. Fisher มักใช้ค่า F ในการทดสอบสมภาพ[34] ของวาเรียนซ์และทดสอบความแตกต่างของค่าเฉลี่ย

ตารางที่ 9 ในภาคผนวกเป็นตาราง F แสดงเฉพาะความแตกต่างระหว่าง 5 และ 1 เปอร์เซ็นต์เท่านั้น ในตารางมี df 2 ค่า คือแถวบนสุดและสดมภ์แรกของตาราง ใช้สำหรับตัวตั้งและตัวหารตามลำดับ ในตารางนั้นมีค่า F 2 ค่า คือ 5 และ 1 เปอร์เซ็นต์ เช่น df ตัวตั้ง = 10 และ df ตัวหาร = 7 พบว่าที่ระดับความแตกต่าง 0.05 นั้น F มีค่าเท่ากับ 3.63 ซึ่งหมายถึงว่าเนื้อที่ทางด้านขวาของการกระจายที่สูงกว่า F = 3.63 นั้น มีค่าเป็น 5 เปอร์เซ็นต์ของเนื้อที่ทั้งหมด หรือกล่าวอีกนัยหนึ่งว่าเนื้อที่ระหว่าง F = 0 ถึง F = 3.63 เป็น 95 เปอร์เซ็นต์ของเนื้อที่ทั้งหมด

2.8 แบบฝึกหัด

1. ถ้าเราทอดลูกเต๋า 4 ครั้ง และให้ k เป็นจำนวนครั้งที่ลูกเต๋าขึ้นหน้า 3 จงหาการกระจายของความน่าจะเป็น k = 0, 1, 2, 3 และ 4 (แนะนำ : P (หน้า 3) = 1/6)

2. จากข้อมูลที่มีการกระจายแบบต่อเนื่อง ซึ่งมี $\mu = 40$ และ $\sigma = 6$ จงคำนวณหาความน่าจะเป็นของค่า ก. ต่ำกว่า 32 ข. สูงกว่า 27 และ ค. ระหว่าง 42 และ 51

3. ประชากรกลุ่มหนึ่งมีการกระจายแบบปกติ มีค่าเฉลี่ย 80 และมีส่วนเบี่ยงเบนมาตรฐาน 10 ถ้าสุ่มตัวอย่างขนาด n = 50 จงหาความน่าจะเป็นของค่าเฉลี่ย ก. ต่ำกว่า 71 ข. มากกว่า 76 ค. ระหว่าง 77 และ 82 และ ง. ไม่สูงกว่า 83

4. ถ้าสุ่มตัวอย่างที่มีขนาด (n) = 17 มาจากประชากรที่มีการกระจายแบบปกติจำนวนหลายตัวอย่าง จากแต่ละตัวอย่างคำนวณหาค่า t

   ก. มีความน่าจะเป็นเท่าไรที่จะได้ t ระหว่าง – 2.12 และ 2.12

   ข. มีความน่าจะเป็นเท่าไรที่จะได้ t สูงกว่า 2.12

   ค. มีความน่าจะเป็นเท่าไรที่จะได้ t สูงกว่า 2.12 หรือต่ำกว่า – 2.12

5. ชุดตัวอย่างมีขนาด (n) = 25 จากประชากรที่มี $\sigma^2 = 6$ จงหา

   ก. ความน่าจะเป็นที่ s สูงกว่า 9.1

   ข. ความน่าจะเป็นที่ s อยู่ระหว่าง 3.462 และ 10.745

คำในบท

(1) distribution, (2) probability, (3) event, (4) sample space, (5) addition rule, (6) mutually exclusive event, (7) multiplicative rule, (8) discrete variable, (9) discrete distribution, (10) binomial distribution, (11) binomial expansion, (12) coefficient of binomial, (13) factorial, (14) hypergeometric distribution, (15) distribution of the proportion, (16) multinomial distribution, (17) Possion distribution, (18) continuous variable, (19) continuous distribution, (20) density function, (21) normal distribution, (22) symmetrical, (23) standard core, (24) Chi-square, (25) sum of square, (26) frequency, (27) observed value, (28) expected value, (29) goodness-of-fit test, (30) distribution of sample mean, (31) standard error (S.E.), (32) student's t, (33) t-distribution, (34) equality, (35) Poisson distribution.

บทที่ 3
การประมาณและการทดสอบสมมุติฐาน

3.1 คำนำ

ในบทก่อนเราได้กล่าวถึง μ และ σ^2 อย่างไรก็ดี ในประชากรทั่วไปนั้นเรามักไม่ทราบค่านี้ ทางที่จะทราบได้คือการประมาณจากตัวอย่าง เช่น เมื่อเราทราบ $\overline{X}$ ก็นำไปใช้ประมาณ μ เมื่อทราบ s^2 ก็นำไปประมาณ σ^2 ค่าประมาณมีอยู่ 2 ชนิด คือ (1) ค่าประมาณเดี่ยว[1] และ (2) ค่าประมาณเป็นช่วง[2] วิธีการอีกอันหนึ่งคือ เมื่อเราไม่ทราบค่าต่าง ๆ ดังกล่าวแล้ว ก็อาจค้นหาโดยวิธีการทดสอบสมมุติฐาน[3]ซึ่งได้แก่ การคาดคะเนค่าของประชากร หรือพารามิเตอร์ แล้วทดสอบว่าถูกหรือผิด การทดสอบสมมุติฐานจัดเป็นการอนุมาณเชิงสถิติ[4] เป็นวิธีการตัดสินจากข้อมูลที่ได้มาโดยการสำรวจ หรือเก็บรวบรวม โดยการทดลอง

3.2 การประมาณ

ดังที่กล่าวมาแล้วว่า การประมาณคือการให้ค่าพารามิเตอร์แก่ประชากร โดยดูจากค่าของตัวอย่าง วิธีการนี้ประหยัด ไม่ต้องสำรวจประชากรทั้งหมด ยิ่งกว่านั้นประชากรบางชนิดก็ไม่มีอยู่จริง เช่น เราไม่มีทางทราบเลยว่าปริมาณฝนตกแต่ละวันมีกี่มิลลิเมตร จะทราบได้ก็โดยวัดจากบางท้องที่ ในบางแห่งอาจตกมากหรือน้อยกว่านั้น แต่ถือว่าการประมาณเป็นการให้ค่าใกล้เคียง ทั้งนี้เพราะเชื่อว่าค่าที่เกิดบ่อย มักเป็นค่าที่ใกล้เคียงพารามิเตอร์นั่นเอง การประมาณมี 2 ชนิด ดังนี้

(1) การประมาณแบบเดี่ยว การประมาณเดี่ยวคือ การใช้ค่าตัวเดียวโดด ๆ เพื่อประมาณอีกค่าหนึ่ง เช่น ใช้ $\overline{X}$ ประมาณ μ การประมาณนี้อาจได้ค่าใกล้เคียงเท่านั้น ทั้งนี้เราทราบจากบทก่อนแล้วว่า ในการเก็บข้อมูลนั้น มีโอกาสสูงมากที่จะได้ค่าเฉลี่ยใกล้ μ คือโอกาสที่ $\overline{X}$ ห่างจาก μ ไปด้านซ้ายและขวาเพียง 1 S.E.[5] เท่ากับ 0.6827 เช่น $\mu = 10$ และ 1 S.E. $\pm$ 1.5 ดังนั้นมีโอกาสถึง 0.6827 ที่จะได้ $\overline{X}$ ระหว่าง 8.5 ถึง 11.5 ซึ่งไม่ห่างมากนัก ยิ่งกว่านั้นถ้าตัวอย่างที่สุ่มมีขนาดใหญ่ ค่าประมาณก็จะถูกต้องยิ่งขึ้น

การประมาณเดี่ยวอาจเป็นการประมาณวาเรียนซ์ คือเมื่อสุ่มตัวอย่างมาแล้วเราอาจใช้ s^2 เพื่อประมาณ σ^2

ค่าประมาณทั้งสองชนิดนี้ จัดเป็นค่าประมาณที่ไม่ลำเอียง ค่าประมาณที่ไม่ลำเอียง คือ ค่าประมาณที่ให้ค่าเฉลี่ยเท่ากับพารามิเตอร์ เช่น จากประชากรหนึ่งเราสุ่มตัวอย่างมามาก ๆ หรือ สุ่มมาจนครบ

ทุกตัวอย่างจากประชากรนั้น แล้วนำค่าเฉลี่ยและวาเรียนซ์แต่ละชนิดมาบวกกัน แล้วหารด้วยจำนวนตัวอย่าง ค่าเฉลี่ยนี้จะเท่ากับ μ และ σ^2 ซึ่งถือว่าเป็นค่าประมาณที่ไม่ลำเอียง อาจเขียนเป็นสมการว่า ค่าประมาณจะไม่ลำเอียงเมื่อ $E(\overline{X}) = \mu$ และ $E(s^2) = \sigma^2$

(2) การประมาณแบบช่วง การประมาณแบบช่วง คือการประมาณที่ให้ค่าที่ต้องการอยู่ระหว่างกลาง และใส่โอกาสหรือความน่าจะเป็นพ่วงเป็นความเชื่อมั่นเข้าไปด้วย เช่น โอกาสที่นักศึกษาสอบตกในวิชาสถิติมีจำนวน 1 ถึง 5 คน เท่ากับ 0.90 ซึ่งเขียนเป็นสมการว่า $P(1 < X < 5) = 0.90$ อย่างนี้เป็นต้น โอกาส 0.90 เรียกว่าความเชื่อมั่น คือเชื่อแน่ได้ว่าในแต่ละครั้งที่สอบนั้นมีโอกาสอยู่ 0.90 ที่จะพบว่ามีนักศึกษาตกอย่างน้อย 1 คน และอย่างมาก 5 คน

อย่างไรก็ดี เรานิยมใช้โอกาสระดับมาตรฐาน คือ 0.95 หรือ 0.99 เมื่อให้ผลรวมของโอกาสเท่ากับ 1 และให้ α คือโอกาสที่เป็นเหตุการณ์อย่างอื่น ก็อาจเขียนสมการการประมาณแบบช่วงว่า

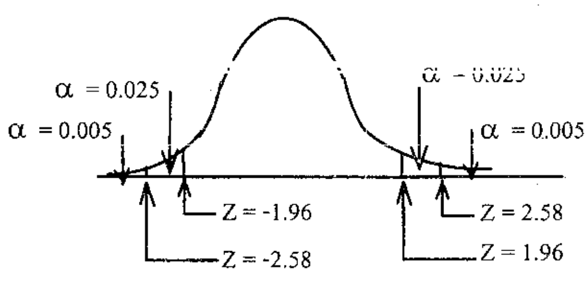
$$P(X < a < Y) = 1 - \alpha \qquad \ldots(3\text{-}1)$$

เมื่อ X และ Y เป็นตัวจำกัดล่างและบนตามลำดับ และ $1-\alpha$ คือสัมประสิทธิ์ของความเชื่อมั่น[6] ช่วงระหว่าง X และ Y นี้เรียกว่า ช่วงเชื่อมั่น[7]

การประมาณแบบช่วง อาจใช้ประมาณค่าเฉลี่ย วาเรียนซ์ ความน่าจะเป็น และอัตราส่วน ซึ่งอธิบายเฉพาะบางค่าดังนี้

(1) ช่วงเชื่อมั่นของค่าเฉลี่ยเมื่อทราบ σ^2 เราทราบแล้วว่า เมื่อประชากรมีการกระจายแบบปกตินั้น จะมีโอกาสอยู่ 0.6827, 0.9545 และ 0.9973 ที่จะพบว่าค่าเฉลี่ยของตัวอย่างอยู่ห่างจาก μ เท่ากับ 1, 2 และ 3 S.E. ตามลำดับ ข้อความเหล่านี้อาจจะนำมาดัดแปลงหาช่วงเชื่อมั่นที่เราต้องการคือที่ 95 และ 99 เปอร์เซ็นต์ ดังนี้

เมื่อเราทราบ σ^2 ก็อาจใช้คะแนนมาตรฐาน (Z) ในการหาช่วงเชื่อมั่นคือ โดยใช้สมการ $Z = (\overline{X} - \mu)/\sigma_{\overline{x}}$ นั่นเอง การกระจายของคะแนนมาตรฐานเป็นแบบปกติดังรูป 3.2.1 เนื้อที่ทั้งหมดมีค่าเท่ากับ 1 เนื้อที่ทางด้านหางที่ตัดออกไปทั้งหมดแทนว่า α ถ้าตัดไปด้านละครึ่งก็แทนว่า $\alpha/2$



รูป 3.2.1 การกระจายของคะแนนมาตรฐานและความสัมพันธ์ระหว่าง α และค่า Z

จากสมการ (3-1) อาจพูดว่าเนื้อที่ทางด้านขวาของ Y ถูกตัดไป $\alpha/2$ และด้านซ้ายของ X ตัดไป $\alpha/2$ ดังนั้นที่เหลือระหว่าง X และ $Y = 1 - \alpha$ ค่า α และค่า Z มีความสัมพันธ์กันดังนี้

เมื่อ $\alpha = 0.05$, ค่า $Z_{\alpha/2} = Z_{.025} = 1.96$

เมื่อ $\alpha = 0.01$, ค่า $Z_{\alpha/2} = Z_{.005} = 2.58$

ดังนั้นสมการสำหรับหาช่วงเชื่อมั่นของค่าเฉลี่ย คือ

$$P(Z_1 < Z < Z_2) = \alpha$$

คือพื้นที่ระหว่าง Z_1 และ Z_2 (ดูรูป 3.2.1) มีโอกาสเท่ากับ $1-\alpha$

จากสมการนี้สามารถคำนวณ โดยการแทนค่า Z_1, Z_2 และ μ จนได้สมการ (3-2) ดังนี้

$$P(\overline{X} - Z\sigma_{\overline{x}} < \mu < \overline{X} + Z\sigma_{\overline{x}}) = 1 - \sigma \qquad \ldots(3\text{-}2)$$

หรือที่ความเชื่อมั่น $1 - \alpha$, μ อยู่ระหว่าง

$$\overline{X} \pm Z\sigma_{\overline{x}} \qquad \ldots(3\text{-}3)$$

ที่ความเชื่อมั่น 95 เปอร์เซ็นต์ ก็เขียนว่า μ อยู่ระหว่าง

$$\overline{X} \pm 1.96\sigma_{\overline{x}} \qquad \ldots(3\text{-}4)$$

และที่ความเชื่อมั่น 99 เปอร์เซ็นต์ ก็เขียนว่า μ อยู่ระหว่าง

$$\overline{X} \pm 2.58\sigma_{\overline{x}} \qquad \ldots(3\text{-}5)$$

ตัวอย่าง 3.2.1

จากการวัดความยาวของปลาชนิดหนึ่ง 50 ตัว พบว่ามี $\overline{X} = 77.4$ มม. สมมุติว่ามี $\sigma = 10$ มม. จงหาช่วงเชื่อมั่นที่ 95 เปอร์เซ็นต์ ของค่าเฉลี่ยของประชากร (μ) ของปลาชนิดนี้

วิธีทำ จากสมการ (3-2) หรือ (3-4) ส่วนเบี่ยงเบนมาตรฐานมีค่า 10 มม. ดังนั้น $\sigma_{\overline{x}} = 10/\sqrt{50}$ จึงหาได้ว่าความเชื่อมั่น 95 เปอร์เซ็นต์ของ μ อยู่ระหว่าง

$$77.4 - (1.96)(10)/\sqrt{50}) < \mu < 77.4 + (1.96)(10/\sqrt{50}) = 0.95$$

คือหาได้ว่าที่ความเชื่อมั่น 95 เปอร์เซ็นต์ μ อยู่ระหว่าง 74.6 และ 80.2 มม.

(2) ช่วงเชื่อมั่นของค่าเฉลี่ยเมื่อไม่ทราบ σ^2 เมื่อไม่ทราบวาเรียนซ์ของประชากรก็ไม่สามารถคำนวณหา $\sigma^2_{\overline{x}}$ ถ้า n มีขนาดใหญ่ (n > 30) ก็ให้ใช้ $s_{\overline{x}}$ แทน $\sigma_{\overline{x}}$ ทั้งนี้ $s_{\overline{x}} = s/\sqrt{n}$ แล้วใช้สมการ (3-4) หรือ (3-5) เพื่อหาช่วงเชื่อมั่นต่อไป แต่ถ้าชุดตัวอย่างมีขนาดเล็ก (n < 30) ก็อาจหาช่วงเชื่อมั่นโดยใช้ค่า t ซึ่งคล้ายกับสมการ (3-2) ดังนี้

$$P(\overline{X} - ts_{\overline{x}} < \mu < \overline{X} + ts_{\overline{x}}) = 1 - \alpha \qquad \ldots(3\text{-}6)$$

หรือที่ความเชื่อมั่น $1-\sigma$, μ จะอยู่ระหว่าง

$$\overline{X} \pm ts_{\overline{x}} \qquad \ldots(3\text{-}7)$$

ในกรณีนี้ค่า t เปิดจากตาราง $t_{\alpha/2}$ ที่ df $= n-1$ ถ้าเป็นช่วงเชื่อมั่น 95 เปอร์เซ็นต์ $\alpha = 0.05$, $\alpha/2 = 0.025$ ถ้าเป็นช่วงเชื่อมั่น 99 เปอร์เซ็นต์ $\alpha = 0.01, \alpha/2 = 0.005$ ตาราง t มี 2 ชนิด คือ อาจเป็นตารางหางเดียว[8] หรือตารางสองหาง[9] ซึ่งแสดงความแตกต่างพอเป็นตัวอย่างสำหรับ df 24 ได้ดังนี้

ช่วงเชื่อมั่น	α จากตาราง		ค่า t
	ตารางหางเดียว[8]	ตารางสองหาง[9]	
95%	0.025	0.05	2.064
99%	0.005	0.01	2.797

ส่วนวิธีการหาช่วงเชื่อมั่นเหมือนกับตัวอย่างในหน้าก่อนทุกประการ เพียงแต่เปลี่ยนค่าคะแนนมาตรฐานเป็นค่า t เท่านั้น

3.3 ขนาดของตัวอย่าง

ในการทดลองแต่ละครั้ง เป็นการยากที่จะบอกว่าขนาดของตัวอย่างควรเป็นเท่าไร อย่างไรก็ดีเราสามารถประมาณขนาดของตัวอย่างได้โดยใช้สมการ $Z = (\overline{X} - \mu)/\sigma_{\overline{x}}$ ทั้งนี้ขึ้นอยู่กับความต้องการที่เราจะให้ $\overline{X} - \mu$ มีค่าเท่าไร ถ้าต้องการไม่มากกว่า $\pm$ E ก็สามารถเขียนว่า

$$Z = \frac{E}{\frac{\sigma}{\sqrt{n}}}$$

ซึ่งเมื่อจัดสมการนี้ใหม่ก็ได้สมการเป็น

$$n = \frac{Z^2_{\alpha/2}\,\sigma^2}{E^2}$$

n คือ ขนาดของตัวอย่างที่ทำให้ $\overline{X} - \mu$ มีค่าไม่เกิน E มีโอกาสเท่ากับ $1-\alpha$ ทั้งนี้ E คือ $\overline{X} - \mu$

ตัวอย่างเช่น ถ้าให้ E = 3, $\sigma^2 = 100$, $Z_{\alpha/2} = 0.005 = 2.58$ ก็หาได้ว่า

$$n = \frac{(2.58)^2(100)}{3^2} = 74$$

ซึ่งมีความหมายว่า ถ้า n = 74 มีโอกาส 0.99 ที่ค่าเฉลี่ยที่เราสุ่มได้ จะห่างจาก μ ไม่เกิน 3

ในการใช้วิธีการนี้เราต้องทราบ σ^2 แต่โดยทั่วไปแล้วเรามักประมาณค่านี้จากชุดตัวอย่าง

3.4 พื้นฐานของการทดสอบสมมุติฐาน

การทดสอบสมมุติฐาน คือการตัดสินค่าที่เราคาดคะเนเอาไว้ว่าถูกหรือผิด การทดสอบอาจนำไปสู่การยอมรับ[10] หรือการปฏิเสธ[11] สมมุติฐาน ค่าที่เรานำมาทดสอบ อาจเป็นค่าเฉลี่ย อาจเป็นวาเรียนซ์ หรือความน่าจะเป็นบางอย่าง รายละเอียดและตัวอย่างเกี่ยวกับการทดสอบสมมุติฐานอาจอธิบายโดยย่อ ๆ ได้ดังต่อไปนี้

(1) ชนิดของสมมุติฐาน สมมุติฐานในการทดสอบมีอยู่ 2 ชนิด คือสมมุติฐานที่น่าจะจริงหรือสมมุติฐานนัล[12] ซึ่งเขียนย่อ ๆ ว่า H_0 และสมมุติฐานแย้งหรือสมมุติฐานเลือก[13] ซึ่งเขียนย่อ ๆ ว่า H_1 ตัวอย่างเช่น เราตั้งสมมุติฐานนัลไว้ว่าน้ำหนักต่อ 100 เมล็ด ของถั่วเขียวพันธุ์อู่ทอง 1 เท่ากับ 6.5 กรัม แต่เราอาจตั้งข้อสงสัย หรือเรียกอีกอย่างหนึ่งว่าเป็นสมมุติฐานแย้งว่าน้ำหนักต่อ 100 เมล็ด สูงกว่า 6.5 กรัม ซึ่งอาจเขียนสมมุติฐานทั้งสองตามวิธีที่นิยมกันดังนี้

$$H_0 : \mu = 6.5 \text{ กรัม/100 เมล็ด}$$
$$H_1 : \mu > 6.5 \text{ กรัม/100 เมล็ด}$$

เมื่อทดสอบแล้ว ถ้าไม่ยอมรับ H_0 ก็ต้องรับ H_1 แต่ถ้าเราตั้งข้อสงสัยว่าน้ำหนักเมล็ดอาจน้อยกว่านี้ ก็เขียนว่า

$$H_0 : \mu = 6.5 \text{ กรัม/100 เมล็ด}$$
$$H_1 : \mu < 6.5 \text{ กรัม/100 เมล็ด}$$

ซึ่งทั้งสองตัวอย่างข้างบนนี้มีสมมุติฐานแย้งหรือ H_1 เป็นแบบด้านเดียวหรือหางเดียว อย่างไรก็ดี H_1 อาจเป็นแบบสองหางก็ได้ คือค่าอะไรก็ได้ที่สูงหรือต่ำกว่า 6.5 ถึงระดับสถิติแล้วไม่รับทั้งนั้น ซึ่งเขียนว่า

$$H_0 : \mu = 6.5 \text{ กรัม/100 เมล็ด}$$
$$H_1 : \mu \neq 6.5 \text{ กรัม/100 เมล็ด}$$

(2) ระดับนัยสำคัญ ระดับนัยสำคัญหรือระดับความแตกต่าง[14] คือโอกาสที่เราปฏิเสธ H_0 ซึ่งแทนโอกาสอันนี้ด้วย α ในการทดสอบสมมุติฐานทุกชนิดเรามักใช้ $\alpha = 0.05$ หรือ 0.01 เรียกว่าเป็นระดับ**นัยสำคัญ** และ **นัยสำคัญยิ่ง** ตามลำดับ[15] ซึ่งเรียกกันโดยทั่วไปว่าเป็นความแตกต่างในระดับ 0.05 และ 0.01

เมื่อเลือกระดับนัยสำคัญแล้ว หมายถึงว่าเราได้แบ่งการกระจายของสถิติทดสอบออกเป็น 2 ส่วน ส่วนหนึ่งคือ ส่วนยอมรับ H_0[16] อีกส่วนหนึ่งคือ ส่วนปฏิเสธ H_0[17] สมมุติว่าเราใช้คะแนนมาตรฐานเป็นสถิติทดสอบ และเลือกใช้ $\alpha = 0.05$ ก็หมายถึงเราแบ่งการกระจายของคะแนนมาตรฐานออกดังรูป 3.2.1 กรณีนี้พบว่าคะแนนมาตรฐานที่คำนวณได้สูงกว่า 1.645 ค่าก็จะตกในตอนปลายของหางทาง

ขวามือ ซึ่งทำให้เราปฏิเสธสมมุติฐานนัลที่ระดับความแตกต่าง 0.05 แต่ถ้าเท่ากับหรือน้อยกว่า 1.645 ก็ยอมรับสมมุติฐานนัล ค่าที่คอยชี้ว่าควรยอมรับหรือปฏิเสธสมมุติฐานนัลนี้เราเรียกว่า ค่าตัดสิน[18] เช่นในกรณีนี้ค่าตัดสินคือ 1.645

(3) การทดสอบแบบหางเดียวและสองหาง ในการทดสอบสมมุติฐานนั้น ผู้ทดสอบอาจต้องการทราบเพียงว่า ค่าที่กำหนดขึ้นน้อยกว่าหรือมากกว่าพารามิเตอร์ เช่น ในกรณีของขนาดเมล็ดถั่วเขียวนั้น ถ้าขนาดเมล็ดไม่เท่ากับ 6.5 กรัม/100 เมล็ด ก็หมายความว่าเมล็ดอาจมีขนาดโตกว่านั้น ในกรณีเช่นนี้ก็เขียนสมมุติฐานว่า

$$H_0 : \mu = 6.5 \text{ กรัม/100 เมล็ด}$$
$$H_1 : \mu > 6.5 \text{ กรัม/100 เมล็ด}$$

การทดสอบเช่นนี้ เรียกว่าเป็นการทดสอบแบบหางเดียว[19] ในกรณีที่ขนาดเมล็ดอาจน้อยกว่า 6.5 กรัม/100 เมล็ด ก็เขียนว่า $H_1 : \mu < 6.5$ กรัม/100 เมล็ด ซึ่งเป็นการทดสอบหางเดียวเช่นกัน ในกรณีแรกนั้น เป็นการทดสอบหางทางขวาของการกระจาย เช่น ที่ $\alpha = 0.05$ ค่า Z ที่คำนวณได้สูงกว่า 1.645 จึงปฏิเสธสมมุติฐานนัล ส่วนในกรณีหลังนั้นเป็นการทดสอบใช้หางทางซ้ายที่ $\alpha = 0.05$ ค่า Z ที่คำนวณได้ต่ำกว่า - 1.645 (หรือมากกว่าค่าสัมบูรณ์) จึงปฏิเสธสมมุติฐานนัล

บางครั้งผู้ทดสอบสนใจเพียงว่าค่าที่หาได้เท่าหรือไม่เท่ากับค่าที่กำหนด ถ้าไม่เท่าจะน้อยกว่าหรือมากกว่า ก็ปฏิเสธสมมุติฐานทั้งนั้น ดังนั้นก็อาจเขียนสมมุติฐานว่า

$$H_0 : \mu = 6.5 \text{ กรัม/100 เมล็ด}$$
$$H_1 : \mu \neq 6.5 \text{ กรัม/100 เมล็ด}$$

การทดสอบแบบนี้เรียกว่า การทดสอบสองหาง[20] เช่น ถ้า $\alpha = 0.05$ เราต้องเปิดค่าตัดสิน 2 ค่า คือที่ $-Z_{\alpha/2}$ และ $Z_{\alpha/2}$ ซึ่งได้ค่าที่เหมือนกันแต่อยู่ในหางต่างกัน คือ $-Z_{.025} = -1.96$ และ $Z_{.025} = 1.96$ ถ้าเราทดสอบค่าเฉลี่ยโดยใช้คะแนนมาตรฐาน ก็อาจแสดงค่าตัดสินพอเป็นตัวอย่างได้ดังนี้

	ระดับนัยสำคัญ (α)		
	0.10	0.05	0.01
ค่าตัดสินสำหรับการทดสอบหางเดียว	− 1.28 , 1.28	− 1.645, 1.645	− 2.33, 2.33
ค่าตัดสินสำหรับการทดสอบสองหาง	− 1.645, 1.645	− 1.96 , 1.96	− 2.58, 2.58

(4) ขั้นตอนของวิธีการทดสอบ เราได้ศึกษาถึงรายละเอียดที่จำเป็น สำหรับการทดสอบสมมุติฐานมาแล้ว ต่อไปนี้เป็นการสรุปถึงขั้นตอนการทดสอบ ในการทดสอบสมมุติฐานมักดำเนินเป็นขั้น ๆ ดังนี้

(1) ตั้งสมมุติฐาน H_0 และ H_1

(2) บอกสถิติทดสอบ

(3) บอกกฎตัดสิน[21] คือบอกระดับความแตกต่างและค่าตัดสิน ทั้งนี้บอกว่าเมื่อไร เราจึงยอมรับหรือปฏิเสธ H_0 ตัวอย่างของกฎตัดสินเช่น "ที่ระดับความแตกต่าง 0.05 เราปฏิเสธ H_0 ถ้า Z ที่คำนวณได้สูงกว่า $Z_{0.05} = 1.645$ ไม่เช่นนั้นก็ยอมรับ H_0"

(4) คำนวณ

(5) สรุปผลการทดสอบ

วิธีการทดสอบอาจไม่ดำเนินตามขั้นเหล่านี้ก็ได้ แต่ควรมีรายละเอียดที่ใกล้เคียงกัน

3.5 การทดสอบค่าเฉลี่ย

การทดสอบค่าเฉลี่ยของประชากรชนิดใดชนิดหนึ่ง อาจใช้วิธี 2 วิธี คือ

(1) ใช้คะแนนมาตรฐาน เมื่อทราบ σ^2 การทดสอบ μ ก็กระทำโดยใช้คะแนนมาตรฐาน คือ ใช้สมการ $Z = (\overline{X} - \mu)/\sigma_{\overline{X}}$ โดยที่ค่า μ ในสมการนี้คือค่า μ ใน H_0 นั่นเอง ซึ่งอาจแสดงโดยใช้ตัวอย่าง ดังนี้

ตัวอย่าง 3.5.1

คะแนนจากผลการสอบวิชาสถิติหลายปี พบว่ามีค่าเฉลี่ย (μ) = 65 คะแนน และมี σ = 16 คะแนน ต่อมามีการเปลี่ยนแปลงรายละเอียดและเทคนิคการสอนวิชานี้ เมื่อสุ่มตัวอย่างมา 64 คน พบว่ามีค่าเฉลี่ย 69 คะแนน จงทดสอบว่าค่าเฉลี่ยดังกล่าวนี้ สูงกว่า μ (65 คะแนน)

วิธีทำ **(1) สมมุติฐาน : $H_0 : \mu = 65$ คะแนน**

$H_1 : \mu > 65$ คะแนน

(จงสังเกตวิธีตั้ง H_0 และ H_1 ในที่นี้เราไม่กล่าวถึงค่าเฉลี่ย 69 โดยตรง แต่เมื่อปฏิเสธ 65 ก็เท่ากับยอมรับว่า 69 ได้จากประชากรที่มีค่าเฉลี่ยสูงกว่า 65 อนึ่งคำถามในโจทย์แสดงนัยว่า เป็นการทดสอบหางเดียว คือให้หาว่าสูงกว่าหรือไม่ คำถามที่ให้หาว่า "สูงกว่า หรือต่ำกว่า" เป็นการทดสอบแบบหางเดียว แต่ถ้าคำถามบอกให้แสดงว่า "ไม่เท่ากับ" ก็แสดงนัยว่าเป็นการทดสอบแบบสองหาง)

(2) กฎตัดสิน : ที่ระดับความแตกต่าง 0.05 เราปฏิเสธ H_0 ถ้า Z ที่คำนวณได้สูงกว่า $Z_{0.05} = 1.645$ (ซึ่งเปิดจากตาราง) ไม่เช่นนั้นก็ยอมรับ H_0 (ในระดับเดียวกัน) ทั้งนี้สถิติทดสอบ คือ

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

(3) คำนวณ : $\mu = 65$ คะแนน, $\sigma = 16$ คะแนน, $\bar{X} = 69$ คะแนน และ $n = 64$ ดังนั้น

$$Z = \frac{69-65}{16/\sqrt{64}} = 2.00$$

(4) $Z = 2.00 > Z_{0.05}$ ซึ่งเท่ากับ 1.645 จึงปฏิเสธ H_0 ที่ระดับความแตกต่าง 0.05 คือ สรุปได้ว่าค่าเฉลี่ยของประชากรนี้สูงกว่า 65 คะแนน

ตัวอย่าง 3.5.2

บริษัทผลิตยาน้ำบริษัทหนึ่งแสดงไว้ที่สลากยาว่า ยาน้ำที่ผลิตมีปริมาตร 100 ซีซี. มีคะแนนมาตรฐาน 4 ซีซี. เมื่อสุ่มยา 144 ขวด มาวัดปริมาตร พบว่ามีค่าเฉลี่ย 99 ซีซี. จงทดสอบที่ระดับ 0.01 ว่าปริมาตรยาน้อยกว่า 100 ซีซี

วิธีทำ (1) สมมุติฐาน : $H_0: \mu = 100$ ซีซี.

$H_1: \mu < 100$ ซีซี.

(2) กฎตัดสิน : ที่ระดับความแตกต่าง 0.01 ปฏิเสธ H_0 ถ้า Z ที่คำนวณได้น้อยกว่า $-Z_{0.01} = -2.33$ ทั้งนี้

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

(3) คำนวณ : $\mu = 100$ ซีซี., $\sigma = 4$ ซีซี., $\bar{X} = 99$ ซีซี., และ $n = 144$

$$Z = \frac{99-100}{4/\sqrt{144}} = -3.00$$

(4) $Z = -3.00 < -Z = -2.33$ (หรืออาจเขียนว่า $|Z| = 3.00 > Z_{0.01} = 2.33$)

จึงปฏิเสธ H_0 ที่ระดับความแตกต่าง 0.01 จึงสรุปได้ว่าปริมาตรยาในขวดน้อยกว่าที่บริษัทแสดงไว้

(2) ใช้สถิติ t เมื่อ $n > 30$ ถึงแม้เราไม่ทราบ σ^2 ก็สามารถประมาณได้จาก s^2 ของตัวอย่าง ดังนั้นก็สามารถทดสอบค่าเฉลี่ยโดยใช้คะแนนมาตรฐาน แต่เมื่อชุดตัวอย่างมีขนาดเล็กคือ $n < 30$ ค่า $(\bar{X}-\mu)/s/\sqrt{n}$ ไม่กระจายแบบคะแนนมาตรฐาน แต่มีการกระจายแบบ t ดังนั้นจึงใช้สถิตินี้เป็นสถิติทดสอบ การทดสอบโดยใช้สถิติชนิดนี้คล้ายคลึงกับการใช้คะแนนมาตรฐานนั่นเอง ใช้ค่า t_α และ $t_{\alpha/2}$ เป็นค่าตัดสินของการทดสอบแบบหางเดียวและสองหางตามลำดับ

ตัวอย่าง 3.5.3

บริษัทผลิตรถยนต์แห่งหนึ่งโฆษณาว่า รถเก๋งขนาดเล็กของบริษัทกินน้ำมันโดยเฉลี่ย 14 กม./ลิตร เมื่อทดสอบรถเก๋ง 16 คัน กินน้ำมันเฉลี่ย 13.4 กม./ลิตร และมีส่วนเบี่ยงเบนมาตรฐาน (s) 2.2 กม./ลิตร จงทดสอบ ที่ระดับความแตกต่าง 0.05 ว่าค่าเฉลี่ยที่แท้จริงต่ำกว่า 14 กม./ลิตร

วิธีทำ **(1) สมมุติฐาน : $H_0 : \mu = 14$ กม./ลิตร**

$H_1 : \mu < 14$ กม./ลิตร

(2) กฎตัดสิน : ที่ระดับความแตกต่าง 0.05 ปฏิเสธ H_0 ถ้า t ที่คำนวณได้น้อยกว่า $-t_{0.05}$ df 15 = -1.753 (จงสังเกตว่าในตาราง t ไม่มีค่าเป็นลบ แต่เราอาจทำให้เป็นลบโดยใส่เครื่องหมายเข้าไป ทั้งนี้ยึดถือ H_1 เป็นหลัก) ดังนี้

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

(3) คำนวณ : μ, $\overline{X}$ และ s มีค่า 14, 13.4 และ 2.2 กม./ลิตร ตามลำดับ และ n = 16 ดังนั้น

$$t = \frac{13.4 - 14}{2.2/\sqrt{16}} = -1.09$$

(4) $t = -1.09 > -t_{0.05} = -1.753$ จึงยอมรับ H_0 ที่ระดับความแตกต่าง 0.05 คือ สรุปว่าค่าเฉลี่ยที่แท้จริงไม่ต่ำกว่า 14 กม./ลิตร

3.6 แบบฝึกหัด

1. จงหาช่วงเชื่อมั่นของค่าเฉลี่ย ดังต่อไปนี้

ก. ช่วงเชื่อมั่นที่ 96 เปอร์เซ็นต์ เมื่อ $\overline{X} = 780$, $n = 30$, $\sigma^2 = 1600$

ข. ช่วงเชื่อมั่นที่ 95 เปอร์เซ็นต์ เมื่อ $\overline{X} = 7.4$, $n = 36$, $\sigma^2 = 0.25$

ค. ช่วงเชื่อมั่นที่ 99 เปอร์เซ็นต์ เมื่อ $\overline{X} = 67.45$, $n = 100$, $s = 2.93$

2. จากข้อ 1 ก. เราควรใช้ขนาดตัวอย่างเท่าใด จึงเชื่อมั่นได้ 96 เปอร์เซ็นต์ว่า ค่าเฉลี่ยของตัวอย่างห่างจาก μ ไม่เกิน 10

3. โรงงานผลิตหลอดไฟฟ้าแห่งหนึ่งกล่าวว่า อายุของหลอดไฟฟ้ามีการกระจายแบบปกติ มี $\mu = 800$ ชั่วโมง และมี $\sigma = 40$ ชั่วโมง ถ้าสุ่มตัวอย่างขนาด $n = 30$ มาชุดหนึ่งพบว่ามี $\overline{X} = 788$ ชั่วโมง จงทดสอบ $H_0 : \mu = 800$ ชั่วโมง และมี $H_1 : \mu < 800$ ชั่วโมง ที่ระดับความแตกต่าง 0.04

4. นักโบว์ลิ่งคนหนึ่งกล่าวว่า เขาสามารถทำแต้มเฉลี่ยได้ 196 แต้ม เมื่อให้เขาทดลองโยนลูกโบว์ลิ่งที่ซื้อมาใหม่ 50 เกมส์ ปรากฏว่าได้เฉลี่ย 188 แต้ม และมีส่วนเบี่ยงเบนมาตรฐาน 24.9 แต้ม ในระดับความแตกต่าง 0.01 เรายอมรับหรือไม่ว่าลูกโบว์ลิ่งใหม่นี้ ทำให้แต้มของเขาต่ำลง

5. ผลผลิตของข้าวพันธุ์หนึ่งปลูกในท้องที่ใกล้เคียงกัน 9 แห่ง ปรากฏว่าได้ 330, 300, 295, 337, 325, 282, 340, 309 และ 308 กก./ไร่ แต่ทางเจ้าหน้าที่เกษตรกล่าวว่าข้าวดังกล่าวให้ผลผลิตเฉลี่ย 330 กก./ไร่ จงทดสอบความจริงอันนี้ในระดับ 0.01

คำในบท

(1) point estimate, (2) interval estimate, (3) test of hypothesis, (4) statistical inference, (5) standard error, (6) confidence coefficient, (7) confidence interval, (8) one-tailed table, (9) two-tailed table, (10) accept, (11) reject, (12) null hypothesis, (13) alternative hypothesis, (14) level of significance (15) singnificant and highly significant level, (16) acceptance region, (17) rejection or critical region, (18) critical value, (19) one-tailed test, (20) two-tailed test, (21) decision rule.

บทที่ 4
การเปรียบเทียบสองตัวอย่าง

4.1 คำนำ

ในบางครั้งผู้ทดลองต้องการที่จะทดสอบเพื่อให้ทราบว่า ค่าเฉลี่ย 2 ค่า เช่น ผลผลิตของพืช 2 พันธุ์ วิธีการปลูกพืช 2 วิธี ผลของอาหารสัตว์ 2 สูตร ฯลฯ มีความแตกต่างกันหรือไม่ พันธุ์พืช วิธีการเพาะปลูก หรืออาหารสัตว์ เหล่านี้เรียกรวม ๆ ว่า วิธีการปฏิบัติ หรือปัญหา หรือตำหรับทดลอง ซึ่งในภาษาอังกฤษเรียกว่าทรีตเมนต์[1] การทดลองเพื่อเปรียบเทียบ 2 ปัญหานี้มีวิธีการและรายละเอียดอีกมาก ซึ่งจะกล่าวโดยสรุปไว้ในบทนี้

4.2 วาเรียนซ์ของความแตกต่างและวาเรียนซ์เฉลี่ย

เมื่อมีประชากร 2 กลุ่มที่เป็นอิสระต่อกัน คือการปรากฏหรือขนาดของสมาชิกในประชากรหนึ่งไม่เกี่ยวข้องกับอีกประชากรหนึ่ง เช่น ถั่วเหลืองพันธุ์ สจ.4 และ สจ.5 ถ้าเราหาน้ำหนัก 100 เมล็ดของถั่วเหลืองทั้ง 2 พันธุ์นี้ โดยการนับเมล็ดมาครั้งละ 100 เม็ด จำนวนหลายครั้ง แต่ละครั้งให้นำน้ำหนักเมล็ดมาหักลบกัน เมื่อหักลบกันแล้วก็จะได้ข้อมูลใหม่อีกชุดหนึ่งซึ่งเรียกว่าเป็นความแตกต่าง เมื่อนำข้อมูลนี้ไปหาวาเรียนซ์เราเรียกว่า เป็นวาเรียนซ์ของความแตกต่าง[2] ซึ่งอาจพิสูจน์ได้ว่าวาเรียนซ์ของความแตกต่างเท่ากับ

$$\mathrm{Var}(X_1 - X_2) = \sigma^2_{x_1} + \sigma^2_{x_2} \qquad \ldots(4\text{-}1)$$

คือวาเรียนซ์ของความแตกต่างเท่ากับผลบวกของวาเรียนซ์ของประชากร 2 กลุ่มนี้ ในทำนองเดียวกันอาจพิสูจน์ได้ว่า เมื่อนำข้อมูลของพืชทั้ง 2 กลุ่ม มารวมกัน ก็จะได้วาเรียนซ์ของผลบวก

$$\mathrm{Var}(X_1 + X_2) = \sigma^2_{x_1} + \sigma^2_{x_2} \qquad \ldots(4\text{-}2)$$

คือวาเรียนซ์ของผลรวมของข้อมูล 2 กลุ่ม จะเท่ากับผลบวกของวาเรียนซ์ของประชากร 2 กลุ่มนี้เช่นกัน

ความจริงดังกล่าวนี้อาจประยุกต์ใช้กับค่าเฉลี่ยด้วย ถ้าเราสุ่มชุดตัวอย่างมาจากประชากร 2 กลุ่ม ซึ่งมีการกระจายแบบปกติ กลุ่มละ 1 ตัวอย่าง ที่มี σ_1^2 และ σ_2^2 โดยให้ตัวอย่างมีขนาด n_1 และ n_2 เมื่อหาค่าเฉลี่ยได้ $\bar{X}_1$ และ $\bar{X}_2$ ตามลำดับ ค่าเฉลี่ยจากประชากรเหล่านี้มีวาเรียนซ์เท่ากับ σ_1^2/n_1 และ σ_2^2/n_2 ตามลำดับ ดังนั้นวาเรียนซ์ของความแตกต่างระหว่างค่าเฉลี่ย คือ

$$\sigma^2_{\bar{x}_1-\bar{x}_2} = \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2} = \sigma^2_{\bar{x}_1}+\sigma^2_{\bar{x}_2} \qquad \text{...(4-3)}$$

คือสรุปได้ว่า $\overline{X}_1 - \overline{X}_2$ มีการกระจายแบบปกติมีค่าเฉลี่ย $\mu_1 - \mu_2$ และมีวาเรียนซ์ $\sigma^2_{\bar{x}_1}+\sigma^2_{\bar{x}_2}$ ถ้าประชากรมีวาเรียนซ์เท่ากัน $(\sigma_1^2=\sigma_2^2=\sigma^2)$ และ $n_1=n_2=n$ ดังนั้น

$$\sigma^2_{\bar{x}_1}+\sigma^2_{\bar{x}_2} = \frac{2\sigma^2}{n} \qquad \text{...(4-4)}$$

แต่ในการศึกษาเกี่ยวกับข้อมูลต่าง ๆ เราไม่ทราบ σ^2 ดังนั้นเราอาจประมาณได้จาก s^2 ของตัวอย่าง ถ้า n_1 และ n_2 มีขนาดใหญ่ (n_1 , n_2 ต่างก็มากกว่า 30) ก็ใช้ s_1^2 และ s_2^2 แทน σ_1^2 และ σ_2^2 ตามลำดับ แต่ถ้า n_1,n_2 มีขนาดเล็กเราให้ค่าเฉลี่ยของ $s_1^2+s_2^2$ เป็นตัวประมาณ σ^2 ของประชากรทั้งสองค่าเฉลี่ย $s_1^2+s_2^2$ เรียกว่าวาเรียนซ์เฉลี่ย[3] ซึ่งเขียนย่อว่า s_p^2 ถ้า $n_1=n_2$ ก็หาได้ว่า

$$s_p^2 = (s_1^2+s_2^2)/2 \qquad \text{...(4-5)}$$

s_p^2 จะหาได้เมื่อ $\sigma_1^2=\sigma_2^2$ หรือตัวอย่างมาจากประชากรกลุ่มเดียวกันเท่านั้น ถ้าสงสัยว่าไม่เท่ากันก็ต้องทดสอบเสียก่อนโดยใช้สถิติ F คือใช้สมการ

$$F = s_1^2/s_2^2 \qquad \text{...(4-6)}$$

เมื่อให้ s_1^2 คือวาเรียนซ์ที่มีค่าสูงกว่า และ s_2^2 มีค่าน้อยกว่า

โดยปกติแล้ว s_p^2 หาได้จากสมการ

$$s_p^2 = \frac{\sum(X_1-\overline{X}_1)^2+\sum(X_2-\overline{X}_2)^2}{n_1+n_2-2} = \frac{SS_1+SS_2}{n_1+n_2-2} = \frac{\text{pooled sum of squares (SS)}}{\text{pooled df}} \qquad \text{...(4-7)}$$

ถ้าเราทราบ s_1^2 และ s_2^2 ก็หา SS_1 และ SS_2 ได้โดยใช้สมการ

$$SS_1 = (n_1-1)s_1^2$$
$$SS_2 = (n_2-1)s_2^2$$

ดังนั้นสมการก็จะเป็น

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \qquad \ldots(4\text{-}8)$$

และถ้า $n_1 = n_2 = n$ สมการ (4-8) ก็จะเหมือนกับสมการ (4-5) ทุกประการ

ดังนั้นในขั้นต่อไปเราประมาณวาเรียนซ์ของความแตกต่างได้โดยใช้สมการ

$$\begin{aligned} s^2_{\bar{x}_1-\bar{x}_2} &= \frac{s_p^2}{n_1} + \frac{s_p^2}{n_2} \\ &= s_p^2\left[\frac{1}{n_1} + \frac{1}{n_2}\right] \end{aligned} \qquad \ldots(4\text{-}9)$$

ถ้า $n_1 = n_2 = n$ ก็อาจพิสูจน์ได้ว่า

$$s^2_{\bar{x}_1-\bar{x}_2} = \frac{2s_p^2}{n} \qquad \ldots(4\text{-}10)$$

ในตอนหรือบทต่อ ๆ ไป เราอาจเขียนว่า $\sigma^2_{\bar{x}_1-\bar{x}_2} = \sigma^2_d$ และ $s^2_{\bar{x}_1-\bar{x}_2} = s^2_d$

4.3 การทดสอบความแตกต่างระหว่างค่าเฉลี่ยสองค่า

เมื่อเราต้องการจะเปรียบเทียบดูว่าประชากร 2 กลุ่มมีความแตกต่างกันหรือไม่ ก็ใช้วิธีการทดสอบให้เหมาะสมกับข้อมูลดังนี้

(1) เมื่อ X_1 และ X_2 กระจายแบบปกติ และทราบค่า σ_1^2 และ σ_2^2

เมื่อมีข้อกำหนดเช่นนี้ ความแตกต่างระหว่าง $\bar{X}_1$ และ $\bar{X}_2$ ก็กระจายแบบปกติ โดยมีค่าเฉลี่ย $\mu_1-\mu_2$ และมีวาเรียนซ์ $\sigma_1^2/n_1+\sigma_2^2/n_2$ ดังนั้นการทดสอบสมมุติฐานเกี่ยวกับความแตกต่างใช้สถิติ

$$Z = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \qquad \ldots(4\text{-}11)$$

โดย Z มีการกระจายแบบปกติ ค่าเฉลี่ยเท่ากับ 0 และวาเรียนซ์เท่ากับ 1

ตัวอย่าง 4.3.1

สมมุติว่าเราสุ่มตัวอย่างมาจากประชากร 2 กลุ่ม ซึ่งมีรายละเอียดดังนี้
$\sigma_1^2 = 9, \sigma_2^2 = 16$, $n_1 = n_2 = 16, \bar{X}_1 = 17.50$ **และ** $\bar{X}_2 = 15.00$ **จงทดสอบที่ระดับความแตกต่าง 0.05 ว่าประชากร 2 กลุ่ม นี้มีค่าเฉลี่ยต่างกันหรือไม่**

วิธีทำ **(1) สมมติฐาน :** $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

(2) กฎตัดสิน : ปฏิเสธ H_0 ที่ระดับความแตกต่าง 0.05 ถ้า Z ที่คำนวณได้อยู่นอกช่วง $Z_{0.025}$ = -1.96 และ $Z_{0.025}$ = 1.96 ดังสมการ (4-11)

(3) คำนวณ : แทนค่า $\overline{X}_1 = 17.50$, $\overline{X}_2 = 15.00$ $\sigma_1^2 = 9$, $\sigma_2^2 = 16$ **และ** $n_1 = n_2 = 16$ **ลงในสมการ (จงสังเกตว่า $\mu_1 - \mu_2$ จะมีค่าเป็นศูนย์ตาม H_0 ดังนั้น เราไม่จำเป็นต้องคำนึงถึงส่วนนี้ของสมการ)**

$$Z = \frac{17.50 - 15.00}{\sqrt{\frac{9}{16} + \frac{16}{16}}} = 2.00$$

(4) Z = 2.00 ไม่อยู่ระหว่าง $Z_{0.025} = -1.96$ **และ 1.96 จึงปฏิเสธ H_0 ที่ระดับความแตกต่าง 0.05 สรุปได้ว่าประชากรทั้งสองนี้มีความแตกต่างกัน**

(2) เมื่อประชากร X_1 และ X_2 ไม่กระจายแบบปกติ แต่ทราบ σ_1^2 และ σ_2^2 และ n_1 , n_2 มีขนาดใหญ่

เมื่อประชากรมีคุณสมบัติดังนี้ แต่ n_1 , $n_2 > 30$ ค่าเฉลี่ย $\overline{X}_1$ และ $\overline{X}_2$ ก็จะมีการกระจายใกล้เคียงแบบปกติ ดังนั้นการกระจายของความแตกต่าง $\overline{X}_1 - \overline{X}_2$ ก็ใกล้เคียงการกระจายแบบปกติด้วย จึงสามารถใช้สมการ (4-11) ในการทดสอบสมมุติฐาน

(3) เมื่อประชากร X_1 และ X_2 กระจายแบบปกติ ไม่ทราบ σ_1^2 และ σ_2^2

ถ้าประชากรมีคุณสมบัติดังนี้ ถึงแม้เราไม่ทราบวาเรียนซ์ของประชากร แต่ถ้า n_1 และ n_2 มีขนาดใหญ่ (n_1 , $n_2 > 30$) ก็สามารถใช้วาเรียนซ์ของตัวอย่างเป็นตัวประมาณ คือ อาจทดสอบสมมุติฐานโดยใช้สมการ (4-11) แต่ใช้ s_1^2 และ s_2^2 ประมาณ σ_1^2 และ σ_2^2 ตามลำดับ

สำหรับประชากรที่มีการกระจายแบบปกติ แต่ n_1 , n_2 มีขนาดเล็ก และเราไม่ทราบ σ_1^2 , σ_2^2 แต่ถ้าประชากรนี้มีวาเรียนซ์เท่ากัน คืออาจเป็นประชากรที่มีต้นกำเนิดจากแหล่งเดียวกันเช่น เป็นพืชชนิดเดียวกัน พันธุ์เดียวกัน พืชชนิดเดียวกันแต่ได้รับปุ๋ยคนละสูตร ฯลฯ ในกรณีเช่นนี้เราใช้ s_p^2 เพื่อประมาณ σ^2 ดังสมการ (4-8) , (4-9) หรือ (4-10) และทดสอบสมมุติฐานโดยใช้สมการ

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} \qquad \ldots(4\text{-}12)$$

การสร้างสมการ (4-12) ต้องการสมมุติฐาน $\sigma_1^2 = \sigma_2^2 = \sigma^2$

ตัวอย่าง 4.3.2

สมมุติว่ามีนักเรียน 2 กลุ่ม แต่ละกลุ่มใช้ตำราเรียนคนละชนิด หลังการสอบไล่ เราสุ่ม คะแนนมากลุ่มละ 2 ชุด โดยมีคะแนนดังนี้

กลุ่มที่ 1	60	65	70	75	80
กลุ่มที่ 2	50	53	60	67	70

จงทดสอบที่ระดับความแตกต่าง 0.01 ว่านักเรียนกลุ่มที่หนึ่งได้คะแนนเฉลี่ยสูงกว่ากลุ่มที่สอง

วิธีทำ

(1) สมมุติฐาน : $H_0 : \mu_1 = \mu_2$

: $H_1 : \mu_1 > \mu_2$

(2) กฎตัดสิน : ปฏิเสธ H_0 ที่ระดับความแตกต่าง 0.01 ถ้าค่า t ที่คำนวณ ได้สูงกว่า $t_{0.01}$, df 8 = 2.896 ทั้งนี้ใช้สถิติ

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

(3) คำนวณ :จากข้อมูลที่กำหนดให้พบว่า $\overline{X}_1 = 70$, $\overline{X}_2 = 60$ ดังนั้น

$$SS_1 = (60-70)^2 + \dots + (80-70)^2 = 250$$

$$SS_2 = (30-60)^2 + \dots + (70-60)^2 = 298$$

ดังนั้น $s_1^2 = 250/4 = 62.50$ และ $s_2^2 = 298/4 = 74.50$

ทดสอบสมมุติฐาน $\sigma_1^2 = \sigma_2^2$ โดยใช้สถิติ

$$F = \frac{s_1^2}{s_2^2} = \frac{\text{วาเรียนซ์ที่มากกว่า}}{\text{วาเรียนซ์ที่น้อยกว่า}} = \frac{74.50}{62.50} = 1.192$$

เมื่อเปิดตาราง F ที่ df 4 , 4 ก็ได้ค่า $F_{0.05} = 6.39$ ซึ่งมากกว่า F ที่คำนวณได้ จึงสรุปได้ว่า ตัวอย่างทั้งสองกลุ่มนี้มาจากประชากรที่มีวาเรียนซ์เท่ากัน ดังนั้น ก็สามารถใช้สมการ (4-12) เพื่อการทดสอบได้ เมื่อแทนค่าต่างๆ ก็จะได้

$$s_p^2 = \frac{SS_1 + SS_2}{n_1 + n_2 - 2} = \frac{250 + 298}{5 + 5 - 2} = 68.5$$

ดังนั้น

$$t = \frac{70 - 60}{\sqrt{\frac{2 \times (68.5)}{5}}} = 1.91$$

(4) $t = 1.91 < t_{0.01}$ df 8 = 2.896 จึงยอมรับ H_0 ที่ระดับความแตกต่าง 0.01 คือ สรุปได้ว่าผลการสอบของนักเรียนไม่แตกต่างกัน

(4) เมื่อ X_1 และ X_2 มีการกระจายแบบปกติ ไม่ทราบ σ_1^2 และ σ_2^2 แต่ $\sigma_1^2 \neq \sigma_2^2$ และ n_1, n_2 มีขนาดเล็ก

ในกรณีทดสอบทางสถิติแล้วพบว่าวาเรียนซ์ของประชากรไม่เท่ากัน เราก็ไม่อาจใช้สมการ (4-12) ในการทดสอบ แต่เราอาจใช้สมการ

$$t' = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad \text{...(4-13)}$$

ถ้า $n_1 = n_2 = n$ ก็เปิดตาราง t ที่ df = n − 1 แต่ถ้า n_1 ไม่เท่ากับ n_2 ก็คำนวณค่าตัดสินจาก

$$t'' = (w_1 t_1 + w_2 t_2)/(w_1 + w_2)$$

เมื่อ $w_1 = s_1^2 / n_1$, $w_2 = s_2^2 / n_2$, t_1 และ t_2 คือค่าจากตาราง t ที่ df = $n_1 - 1$ และ $n_2 - 1$ ตามลำดับ

4.4 การเปรียบเทียบแบบจับคู่

ในการทดลองเราอาจจัดหน่วยรองรับการทดลอง[4] ไว้เป็นคู่ ๆ ตามความคล้ายคลึงกัน เช่น เปรียบเทียบอาหารสุกร 2 สูตร โดยใช้สุกร 10 ครอก โดยนำสุกรมา ครอกละ 1 คู่ ให้สุกรตัวหนึ่งในแต่ละคู่กินอาหารสูตรแรก สุกรตัวที่สองในครอกเดียวกันกินอาหารสุกรสูตรที่สอง เมื่อสิ้นสุดการทดลอง ความแตกต่างระหว่างน้ำหนักของสุกรแต่ละคู่ ย่อมเกิดจากผลของสูตรอาหารนั่นเอง การทดลองเช่นนี้เป็นการทดลองแบบจับคู่

การทดลองแบบจับคู่ อาจดัดแปลงใช้สำหรับการทดลองทางเกษตร ได้หลายแบบ เช่น การทดลองเปรียบเทียบผลผลิตของพืช เมื่อมีการใช้ปุ๋ยและไม่ใช้ การปราบวัชพืชและไม่ปราบ การฉีดยาฆ่าแมลงและไม่ฉีด ฯลฯ โดยปลูกในแปลงใกล้ชิดกันในหลาย ๆ ท้องที่ คือสรุปได้ว่าใช้ได้เสมอสำหรับการทดลองที่มี 2 ปัญหา[5] จุดประสงค์ในการจับคู่ ก็เพื่อที่จะเพิ่มประสิทธิภาพของการทดลอง ทั้งนี้เพราะเราสามารถขจัดความปรวนแปร อันเกิดจากสาเหตุภายนอกที่ไม่เกี่ยวกับการทดลองออกไป เช่น ความปรวนแปรของ

หน่วยรองรับการทดลอง อันเนื่องมาจากพันธุกรรม ท้องที่ ความสามารถเฉพาะตัว ฯลฯ การทดลองแบบจับคู่จึงทำให้สามารถจับความแตกต่างเพียงเล็กน้อยได้

สมมุติว่าเราทำการทดลองเปรียบเทียบอาหารลูกสุกร 2 สูตร โดยใช้แบบจับคู่ 5 คู่ ก็อาจใช้สุกรเพศเดียวกันจาก 5 ครอก แต่ละครอกนำมาทดลอง 1 คู่ ให้ตัวหนึ่งในแต่ละคู่กินอาหารสูตรหนึ่ง ส่วนอีกตัวหนึ่งกินอาหารอีกสูตรหนึ่ง แล้วหาความแตกต่างระหว่างน้ำหนักเพิ่มของสุกร

ครอกที่	อาหาร 1	อาหาร 2	ความแตกต่าง
1	X_{11}	X_{21}	$D_1 = X_{11} - X_{21}$
2	X_{12}	X_{22}	$D_2 = X_{12} - X_{22}$
3	X_{13}	X_{23}	$D_3 = X_{13} - X_{23}$
4	X_{14}	X_{24}	$D_4 = X_{14} - X_{24}$
5	X_{15}	X_{25}	$D_5 = X_{15} - X_{25}$

ค่า D_1 ถึง D_5 กลายเป็นข้อมูลชุดใหม่ซึ่งเราอาจคำนวณหาค่าเฉลี่ยและวาเรียนซ์ดังนี้

$$\overline{D} = \sum D_i / n$$

เมื่อ $i = 1, 2, \ldots, n$; n = จำนวนคู่ทดลอง และอาจคำนวณหาวาเรียนซ์ได้ดังนี้

$$s_D^2 = \frac{\sum (D_i - \overline{D})^2}{n-1} \qquad \ldots(4\text{-}14)$$

$$s_{\overline{D}} = \sqrt{s_D^2 / n} \qquad \ldots(4\text{-}15)$$

ซึ่งอาจพิสูจน์ได้ว่า

$$s_D^2 = s_1^2 + s_2^2 - \frac{2\sum (X_1 - \overline{X}_1)(X_2 - \overline{X}_2)}{n-1}$$

ส่วนสุดท้ายในสมการนี้คือโควาเรียนซ์ ค่านี้จะสูงถ้าสมาชิกในแต่ละคู่สอดคล้องกันมาก ดังนั้นจึงยิ่งทำให้ $s_{\overline{D}}$ มีค่าน้อยลง และสามารถจับความแตกต่างได้มากยิ่งขึ้น ในการทดสอบสมมุติฐานนั้นเราใช้สถิติ

$$t = \frac{\overline{D} - \mu_D}{s_{\overline{D}}} \qquad \ldots(4\text{-}16)$$

df ที่เปิดค่า t คือ n - 1 เมื่อ n เป็นจำนวนคู่

ตัวอย่าง 4.4.1

ในการทดสอบผลของไวรัส 2 พวก บนใบยาสูบ โดยทาไวรัสพวกแรก (ก) ลงบนใบยาสูบซีกหนึ่งอีกพวกหนึ่ง (ข) ลงอีกซีกหนึ่ง ทาทั้งสิ้น 8 ใบ แล้วนับจำนวนจุดแผลได้ดังนี้

ใบที่	1	2	3	4	5	6	7	8
ไวรัส ก	31	20	18	17	9	8	10	7
ไวรัส ข	18	17	14	11	10	7	5	6

จงทดสอบความแตกต่างของผลไวรัสสองชนิดนี้ที่ระดับความแตกต่าง 0.05

วิธีทำ

(1) สมมุติฐาน : $H_0 : \mu_D = 0$

$H_1 : \mu_D \neq 0$

(2) กฎตัดสิน : ปฏิเสธ H_0 ที่ระดับความแตกต่าง 0.05 ถ้า t ที่คำนวณได้ไม่อยู่ระหว่าง $-t_{0.025}$ df 7 = −2.365 และ $t_{0.025}$ df 7 = 2.365 ทั้งนี้ใช้สถิติ

$$t = \frac{\bar{D} - \mu_D}{s_{\bar{D}}}$$

(3) คำนวณ : จากข้อมูลที่กำหนดให้พบว่า

ใบที่	1	2	3	4	5	6	7	8
$D(X_1 - X_2)$	13	3	4	6	−1	1	5	1

$$\bar{D} = \sum D_i / n = 32/8 = 4$$

$$s_D^2 = (13-4)^2 + ... + (1-4)^2 / 7 = 18.57$$

ดังนั้น

$$t = \frac{4}{\sqrt{18.57/8}} = 2.63$$

(4) เนื่องจาก t = 2.63 ไม่อยู่ระหว่าง ± 2.365 จึงปฏิเสธ H_0 คือ สรุปได้ว่าไวรัสนี้ให้ผลแตกต่างกัน

4.5 แบบฝึกหัด

1. ประชากรกลุ่มหนึ่งมี $\sigma^2 = 36$ เมื่อสุ่มตัวอย่างที่มี $n_1 = 16$ มาชุดหนึ่ง พบว่ามี $\overline{X}_1 = 33$ ในระยะต่อมามีการสุ่มชุดตัวอย่างมาอีกชุดหนึ่งมีขนาด $n_2 = 32$ ปรากฏว่า ได้ $\overline{X}_2 = 29$ จงทดสอบที่ระดับความแตกต่าง 0.05 ว่าค่าเฉลี่ยของประชากรนี้แตกต่างกัน

2. ในการเรียนวิชาสถิตินั้นได้แบ่งนักศึกษาออกเป็น 2 ห้อง ห้องละ 50 คน ทำการสอนโดยใช้ตำราประกอบต่างกัน เมื่อทำการสอบปรากฏว่าทั้งสองห้องให้ค่าเฉลี่ย $\overline{X}_1 = 77, \overline{X}_2 = 74$ เปอร์เซ็นต์ และวาเรียนซ์ 64 และ 49 ตามลำดับ ท่านคิดว่านักศึกษาห้องแรกได้คะแนนสูงกว่าห้องที่สองหรือไม่ จงทดสอบที่ระดับความแตกต่าง 0.01

3. ถ้าทำการสุ่มตัวอย่างขนาด $n_1 = 11$ และ $n_2 = 14$ มาจากประชากร 2 กลุ่ม ได้ $\overline{X}_1 = 75, \overline{X}_2 = 60, s_1 = 6.1$ และ $s_2 = 5.3$ จงทดสอบ $H_0 : \mu_1 = \mu_2$ โดยมี $H_1 : \mu_1 \neq \mu_2$ ที่ระดับความแตกต่าง 0.05

4. มีการสงสัยกันว่าถั่วลิสงดิบและถั่วลิสงคั่วมีคุณค่าทางอาหารแตกต่างกัน ดังนั้นจึงทดลองเลี้ยงหนูอย่างละ 9 ตัว เมื่อสิ้นสุดการทดลองได้น้ำหนักเพิ่มเป็นกรัมดังนี้

ถั่วลิสงดิบ	61	60	56	63	56	63	59	44	61
ถั่วลิสงคั่ว	55	54	47	59	51	61	57	62	58

จงทดสอบที่ระดับความแตกต่าง 0.01 ว่าถั่วลิสงคั่วมีคุณภาพต่ำกว่าถั่วลิสงดิบ

5. การวัดความดันโลหิตโดยใช้เครื่องวัด 2 ชนิด แต่ละชนิดทดลองวัดคนไข้คนเดียวกัน จำนวน 10 คนได้ผลดังนี้

คนที่ 1	1	2	3	4	5	6	7	8	9	10
เครื่องแรก	134	147	165	152	122	138	147	153	178	139
เครื่องที่สอง	131	140	164	153	122	135	148	147	165	133

จงทดสอบที่ระดับความแตกต่าง 0.05 ว่าผลจากการวัดจากเครื่องมือทั้งสองชนิดนี้ไม่แตกต่างกันทั้งนี้ให้ $H_0 : \mu_D = 0$

คำในบท

(1) treatment, (2) variance of difference, (3) pooled average variance, s_p^2 (4) experimental unit, (5) treatment.

บทที่ 5
รีเกรซชันเส้นตรง

5.1 คำนำ

ข้อมูลสองชนิดอาจเกิดในการทดลองหรือในการสำรวจครั้งเดียวกันก็ได้ ทั้งนี้เกิดขึ้นเพราะข้อมูลชนิดหนึ่งเป็น "เหตุ" ส่วนข้อมูลอีกชนิดหนึ่งเป็น "ผล"[1] เช่น เราพบว่า น้ำหนักของคน มีความเกี่ยวข้องกับความสูง ผลผลิตข้าวโพดขึ้นอยู่กับปริมาณปุ๋ยที่ใส่ลงไป คะแนนการสอบของนักศึกษาขึ้นอยู่กับไอคิว การกินน้ำมันของรถยนต์ขึ้นอยู่กับความเร็วที่วิ่ง ฯลฯ ข้อมูลที่เกี่ยวข้องแบบข้อมูลหนึ่งเป็นเหตุ และอีกข้อมูลหนึ่งเป็นผลเช่นนี้เป็นข้อมูลที่มีความสัมพันธ์กัน ซึ่งอาจเกิดจากสาเหตุเดียวกันหรือคนละสาเหตุก็ตาม แต่เราพอจะแยกได้ว่าข้อมูลใดเป็นเหตุและข้อมูลใดเป็นผล ข้อมูลที่มีความสัมพันธ์กันแบบที่ค่าหนึ่งขึ้นอยู่กับอีกค่าหนึ่งนี้ เรียกว่า เป็นข้อมูลที่มีรีเกรซชันต่อกัน ความสัมพันธ์ของข้อมูลในลักษณะนี้เรียกว่า ความสัมพันธ์แบบถดถอย

5.2 รีเกรซชันของ Y ต่อ X

ในการศึกษาเกี่ยวกับรีเกรซชันนั้น เราแยกข้อมูลออกได้เป็น 2 ชนิด ชนิดที่เป็นเหตุให้เป็น X และชนิดที่เป็นผลให้เป็น Y จึงพูดได้ว่า Y ขึ้นต่อ X หรือเรียกว่า Y มีรีเกรซชันต่อ X ทั้งนี้ X จัดเป็นตัวแปรอิสระ[2] คือแต่ละค่าเกิดขึ้นโดยธรรมชาติของตัวเอง ส่วน Y เป็นค่าที่ขึ้นโดยตรงต่อ X เรียกว่าเป็นตัวแปรไม่เป็นอิสระ[3] ในกรณีของตัวอย่างข้างบนเราอาจให้ความสูงของคนเป็น X และให้ น้ำหนักเป็น Y ดังนี้เป็นต้น

ในทางทฤษฎีนั้นแต่ละค่า X มี Y ได้หลายค่า เช่น คนที่สูง 165 ซม. นั้นอาจมีน้ำหนักตั้งแต่ 50 ถึง 80 กก. ก็ได้ คนที่สูง 170 ซม. อาจมีน้ำหนักตั้งแต่ 55 ถึง 85 กก. ก็ได้ ฯลฯ ซึ่งจะเห็นได้ว่าเมื่อมี X หลายค่า ก็มี Y ได้หลายกลุ่ม แต่ละกลุ่มนี้เราเรียกว่าประชากร อย่างไรก็ดีเราต้องตั้งเป็นข้อกำหนดไว้ว่า ประชากรเหล่านี้มีวาเรียนซ์เท่ากัน

ก่อนที่จะศึกษาต่อไปในเรื่องรีเกรซชัน จะขอยกตัวอย่างข้อมูลง่าย ๆ เพื่ออธิบายรายละเอียดเกี่ยวกับเรื่องนี้เสียก่อน เช่น ให้ X เป็นน้ำหนักปุ๋ยที่ใส่แปลงผักคะน้า และ Y เป็นน้ำหนักของผักที่เก็บเกี่ยวได้ (เป็น กก./แปลง) ดังนี้

ปุ๋ย	X	0	1	2	3	4	5
ผลผลิต	Y	2	4	6	8	10	12

จากตัวอย่างนี้พอจะเห็นได้ว่าผลผลิต (Y) ขึ้นอยู่กับปริมาณปุ๋ย (X) ที่ใส่ลงไป ซึ่งอาจแสดงความสัมพันธ์เป็นรูปได้ดังรูป 5.2.1 จากรูปดังกล่าวสรุปได้ดังนี้

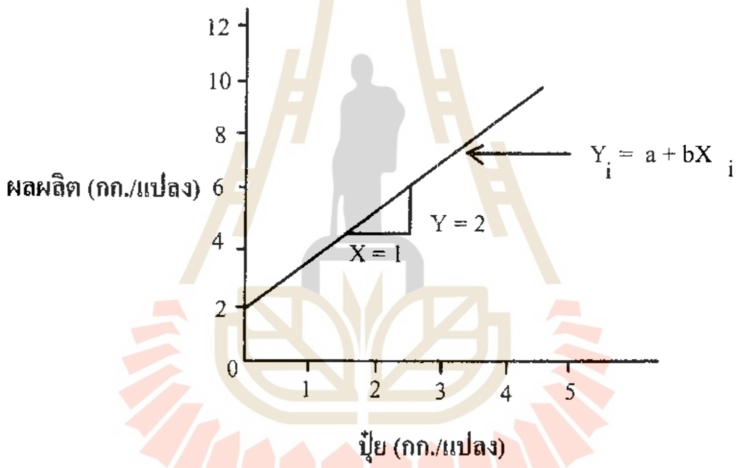
(1) เมื่อ X = 0, Y = 2 ค่าของ Y ที่ X = 0 นี้เราแทนด้วย a ค่า a เราเรียกว่า Y-intercept เป็นค่าที่บอกว่าเส้นรีเกรซชันตัดแกน Y ที่ใด

(2) เมื่อ X เพิ่ม 1 หน่วย, Y เพิ่ม 2 หน่วย จำนวนค่า Y ที่เพิ่ม (หรือลด) ต่อ การเพิ่ม X จำนวน 1 หน่วย เราเรียกว่าอัตราการลาดเท หรือสัมประสิทธิ์ของรีเกรซชัน[4] ซึ่งแทนด้วย b ในตัวอย่างนี้ b มีค่าเท่ากับ 2 ค่า b อาจมีค่าเป็นลบก็ได้ ถ้าตัวแปรหนึ่งเพิ่มขึ้นอีกตัวแปรลดลง

(3) ค่า Y ในเส้นตรงนี้มีค่าเท่ากับ a + bX_i สมการนี้เขียนว่า

$$Y = a + bX_i \qquad \ldots(5\text{-}1)$$

ถ้า a = 2, b = 2 ก็อาจเขียนว่า $Y = 2 + 2X_i$



รูป 5.2.1 แสดงความสัมพันธ์ระหว่างปริมาณปุ๋ย (X) และผลผลิตของคะน้า (Y)

ที่แสดงมานี้เป็นตัวอย่างง่าย ๆ แต่โดยปกติสำหรับแต่ละค่า X จะได้ Y มาโดยวิธีสุ่ม ดังนั้นเป็นการยากที่จะให้ X และ Y ทุกคู่วางอยู่บนเส้นตรงเดียวกัน ที่พออนุโลมได้ คือให้เส้นตรงผ่านใกล้จุด ต่าง ๆ มากที่สุด เส้นตรงที่ดีที่สุดสำหรับแต่ละชุดข้อมูล คือเส้นตรงที่ผ่านตำแหน่ง ซึ่งมีระยะระหว่าง Y ที่อยู่บนเส้นตรงและ Y ตามจุดต่าง ๆ ที่มีค่าต่ำที่สุด หรืออีกนัยหนึ่งคือให้ได้ผลบวกของกำลังสองของค่าเหล่านี้ต่ำที่สุด[5] (วิธีการนี้เราเรียกว่าวิธี least square)

สมการรีเกรซชัน (5-1) นั้น เรียกว่าเป็นสมการของตัวอย่าง คือเป็นของข้อมูลชุดหนึ่งที่ได้จากการสุ่มในประชากรนั้น อาจสมมุติให้ข้อมูลมีความสัมพันธ์แบบเส้นตรงดังสมการ

$$Y_i = \alpha + \beta X_i \qquad \ldots(5\text{-}2)$$

และประชากรของ Y_i สำหรับแต่ละค่า X ให้สมการว่า

$$Y_{pi} = \alpha + \beta X_i + \varepsilon_i \qquad \ldots(5\text{-}3)$$

ทั้งนี้ให้ α และ β เป็นพารามิเตอร์บอก Y-intercept และสัมประสิทธิ์ของรีเกรซชันตามลำดับ และให้ ε_i (epsilon) แทนค่าเบี่ยงเบนของ Y_{pi} จากค่าคาดหมาย และ Y_{pi} คือค่า Y ของกลุ่มประชากรของแต่ละค่า X

5.3 รีเกรซชันของตัวอย่าง

ประชากรเป็นเพียงสิ่งสมมุติ เป็นการยากที่เราจะทราบ α, β, X_i และ Y_i ที่แท้จริง แต่ก็อาจได้ค่านี้จากตัวอย่าง เช่น ในการศึกษาถึงความสัมพันธ์ระหว่างความสูงและน้ำหนักของนักศึกษาชาย 10 คน ดังแสดงในตาราง 5.3.1 นั้น เราให้ความสูงเป็น X และน้ำหนักเป็น Y เพราะน้ำหนักน่าจะขึ้นอยู่กับความสูงมากกว่าที่ความสูงจะขึ้นอยู่กับน้ำหนัก ถ้าเราเขียนกราฟแล้วนำค่า X และ Y มาจุดลงในจุดที่ค่านี้ตัดกัน ก็จะได้จุดกระจัดกระจายจุดเหล่านั้นแต่ละจุดมีค่าดังสมการ

$$Y_i = a + bX_i + e_i \qquad \ldots(5\text{-}4)$$

ตาราง 5.3.1 แสดงความสูงและน้ำหนักของนักศึกษาชายจำนวน 10 คน

คนที่	ความสูง (ซม.)	น้ำหนัก (กก.)	คนที่	ความสูง (ซม.)	น้ำหนัก (กก.)
1	156	46	6	165	54
2	157	47	7	167	55
3	160	51	8	170	61
4	162	52	9	172	60
5	163	54	10	174	62

แต่สมการดังกล่าวนี้ไม่สามารถใช้ในการทำนายได้ เราจำเป็นต้องหาสมการเส้นตรงคล้ายกับสมการ (5-1) แต่เป็นค่าที่ได้จากการคำนวณ จึงให้สมการเป็น

$$Y_c = a + bX_i \qquad \ldots(5\text{-}5)$$

Y_c ทุกค่าจะอยู่บนเส้นตรง จะเห็นได้ว่า e_i คือความแตกต่างระหว่างสมการ (5-4) และ (5-5) ซึ่งจัดเป็นค่าเบี่ยงเบนของค่าสำรวจจากค่าทำนาย คือ

$$e_i = Y_i - Y_c = Y_i - (a + bX_i) \qquad \ldots(5\text{-}6)$$

ดังนั้นเส้นตรงที่หาโดยใช้สมการ (5-5) จะให้ค่า $\sum e_i^2$ ต่ำที่สุด[6]

เส้นตรงหรือเส้นรีเกรซชันที่กล่าวถึงข้างบน คือเส้นที่ลากผ่าน Y_c สำหรับทุกค่า X_i ในการลากเส้นรีเกรซชันนั้นเรามีวิธีการดังนี้

1. หาจุด $\overline{X}$ และ $\overline{Y}$ เส้นตรงจะต้องผ่านจุดนี้ ซึ่งหาได้ว่า $\overline{X} = 164.6$, $\overline{Y} = 54.2$ (ตาราง 5.3.2)
2. หาจุดที่เส้นตรงต้องลากผ่านอีก 1 จุด แล้วลากเส้นเชื่อมระหว่าง 2 จุดนี้ ก็จะได้ส่วนหนึ่งของเส้นตรง

ในการหาจุดอีกจุดหนึ่งเราต้องหาค่า a และ b ของสมการ (5-5) เสียก่อน โดยวิธีดังต่อไปนี้ :

เมื่อคูณ X_i เข้ากับสมการ (5-4) แล้วบวกกันก็จะได้

$$\sum X_i Y_i = a\sum X_i + b\sum X_i^2 \qquad \ldots(5\text{-}7)$$

สังเกตุว่า e_i หายไปจากสมการ (5-7) เพราะ $\sum e_i = 0$

เมื่อบวกค่าในสมการ (5-4) เข้าด้วยกันก็จะได้

$$\sum Y_i = na + b\sum X_i \qquad \ldots(5\text{-}8)$$

เมื่อคูณ n เข้ากับสมการ (5-7) ก็จะได้

$$n\sum X_i Y_i = na\sum X_i + nb\sum X_i^2 \qquad \ldots(5\text{-}9)$$

เมื่อคูณ $\sum X_i$ เข้ากับสมการ (5-8) ก็จะได้

$$\sum X_i \sum Y_i = na\sum X_i + b(\sum X_i)^2 \qquad \ldots(5\text{-}10)$$

เมื่อนำสมการ (5-9) ลบด้วยสมการ (5-10) ก็จะได้

$$n\sum X_i Y_i - \sum X_i \sum Y_i = nb\sum X_i^2 - b(\sum X_i)^2$$

ดังนั้นจึงหาได้ว่า

$$b = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2} \qquad \ldots(5\text{-}11)$$

$$= \frac{\sum X_i Y_i - \dfrac{\sum X_i \sum Y_i}{n}}{\sum X_i^2 - \dfrac{(\sum X_i)^2}{n}} \qquad \ldots(5\text{-}12)$$

จากสมการ (5-8) ก็หาได้ว่า

$$a = \frac{\sum Y_i}{n} - b\frac{\sum X_i}{n}$$

$$= \overline{Y} - b\overline{X} \qquad \ldots(5\text{-}13)$$

อนึ่งสมการ (5-11) หรือ (5-12) นั้น ที่แท้จริงคือ โควาเรียนซ์ หารด้วยวาเรียนซ์นั่นเอง ซึ่งมีค่าเหมือนกับ

$$b = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2} \qquad \ldots(5\text{-}14)$$

จึงสรุปได้ว่าเราอาจใช้สมการ (5-11), (5-12) หรือ (5-14) สมการใดสมการหนึ่ง เพื่อหาค่า b ก็ได้ และใช้สมการ (5-13) เพื่อหาค่า a

จากข้อมูลในตาราง 5.3.1 เราอาจคำนวณค่าที่จำเป็นลงในตาราง 5.3.2 แล้วแทนค่าที่เหมาะสมลงในสมการ (5-11) ก็จะได้

$$b = \frac{(10)(89{,}514) - (1{,}646)(542)}{10(271{,}272) - (1{,}646)^2} = 0.88$$

เมื่อแทนค่าในสมการ (5-13) ก็จะได้

$$a = \frac{542}{10} - (0.88)\frac{1{,}646}{10}$$

$$= -90.65$$

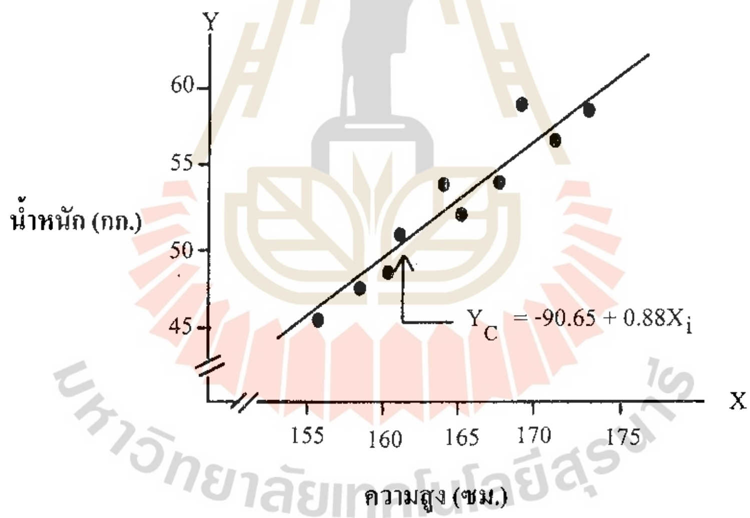
ดังนั้นสมการรีเกรซชันสำหรับความสัมพันธ์ระหว่างความสูงและน้ำหนัก คือ

$$Y_c = -90.65 + 0.88X_i \qquad \ldots(5\text{-}15)$$

เราอาจใช้สมการนี้เพื่อหาค่า Y_c สำหรับค่า X_i ทุกค่า ตัวอย่างเช่น เมื่อ $X = 156$ ก็หาได้ว่า $Y_c = -90.65 + (0.88)(156) = 46.63$ เป็นต้น เมื่อคำนวณ Y_c สำหรับ X_i ทุกค่าก็จะได้ค่า Y_c ดังแสดงในสดมภ์รองสุดท้ายของตาราง 5.3.2

ตาราง 5.3.2 แสดงค่าที่จำเป็นสำหรับการคำนวณหาค่า b และ a

X_i	Y_i	X_iY_i	X_i^2	Y_i^2	Y_c	$Y_i - Y_c$
156	46	7,176	24,336	2,116	46.63	-0.63
157	47	7,379	24,649	2,209	47.51	-0.51
160	51	8,160	25,600	2,601	50.15	0.85
162	52	8,424	26,244	2,704	51.91	0.09
163	54	8,802	26,569	2,916	52.79	1.21
165	54	8,910	27,225	2,916	54.55	-0.55
167	55	9,185	27,889	3,025	56.31	-1.31
170	61	10,370	28,900	3,721	58.95	2.05
172	60	10,320	29,584	3,600	60.71	-0.71
174	62	10,788	30,276	3,844	62.47	-0.47
รวม 1,646	542	89,514	271,272	29,652		0.02
เฉลี่ย 164.6	54.2					



รูป 5.3.1 แสดงรีเกรซชันของน้ำหนักที่มีต่อความสูงของนักศึกษา

ในกรณีของการหาเส้นรีเกรซชัน เรากล่าวไว้ครั้งแรกว่าเส้นรีเกรซชันจะวิ่งผ่านจุด $\overline{X}, \overline{Y}$ เสมอ ส่วนอีกจุดที่ต้องหาเพื่อลากเส้นรีเกรซชั่นนั้น เราคำนวณโดยใช้สมการ (5-15) นั่นเอง เช่น อาจเป็นจุดที่ $X = 170, Y = 58.95$ ก็ได้ อย่างไรก็ดีในที่นี้เราคำนวณไว้ทุกจุดที่เส้นรีเกรซชันต้องวิ่งผ่านแล้ว เมื่อเราได้จุดที่ต้องการแล้วก็อาจลากเส้นรีเกรซชันดังแสดงไว้ในรูป 5.3.1 อนึ่งเราอาจใช้สมการ (5-15) ทำนายค่า Y_c สำหรับค่า X_i อื่น ๆ ที่ไม่แสดงไว้ก็ได้ เช่น เราอาจต้องการทราบว่านักศึกษาที่

สูง 168 ซม. จะหนักเท่าไร แต่เราใช้สมการนี้ทำนายได้ดีเฉพาะในขอบเขตความสูงที่ศึกษา (คือ 156-174 ซม.) เท่านั้น เพราะความสูงนอกขอบเขตนี้ เราไม่แน่ใจว่าความสัมพันธ์จะเป็นแบบเส้นตรง หรือไม่

ในกรณีที่ข้อมูลมีค่าสูง การคำนวณหาค่า b จากสมการ (5-12) ไม่สะดวก เราอาจใช้สมการ (5-14) เช่น ตัวอย่างในตาราง 5.3.3 ซึ่งหาสมการรีเกรซชันของความดันโลหิตต่ออายุของสตรีจำนวน 5 คน

ตาราง 5.3.3 แสดงความดันโลหิตของสตรีที่มีอายุต่าง ๆ กัน 5 คน และวิธีหาสมการรีเกรซชัน

อายุ X_i	ความดันโลหิต Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
35	114	−20	−27	400	540
45	124	−10	−17	100	170
55	143	0	2	0	0
65	158	10	17	100	170
75	166	20	25	400	500
รวม 275 เฉลี่ย 55	705 141	0	0	1,000	1,380

$$b = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{1,380}{1,000} = 1.38$$

$$a = \bar{Y} - b\bar{X} = 141 - (1.38)(55) = 65.1$$

สมการรีเกรซชัน คือ $Y_c = 65.1 + 1.38X_i$

5.4 วาเรียนซ์จากรีเกรซชัน

เมื่อใช้สมการ $Y_c = a + bX_i$ เพื่อคำนวณค่า Y_c ดังแสดงผลไว้ในตาราง 5.3.2 และนำค่า Y_c ไปวาดเป็นเส้นตรงไว้ในรูป 5.3.1 จะเห็นได้ว่าแต่ละความสูงนั้นจะมีความแตกต่างระหว่าง Y_i (ค่าที่สำรวจได้) และ Y_c ค่าที่ได้จากการคำนวณ ความแตกต่างดังกล่าวนี้แสดงไว้ในสมการ (5-6) และในสดมภ์สุดท้ายของตาราง 5.3.2 ค่าดังกล่าวนี้แสดงถึงขนาดของความเบี่ยงเบนของค่าจากการคำนวณ โดยใช้สมการและค่าที่สังเกตได้จริง เมื่อนำค่านั้นไปยกกำลังสองแล้วบวกกัน ก็อาจหาวาเรียนซ์ของรีเกรซชันโดยใช้สมการ

$$s_{y.x}^2 = \frac{\sum (Y_i - Y_c)^2}{n-2}$$

$$= \frac{\sum e_i^2}{n-2} \qquad \ldots(5\text{-}16)$$

วาเรียนซ์นี้เรียกว่า วาเรียนซ์ของ Y_i เมื่อ Y_i ขึ้นต่อ X ทั้งนี้หาได้ว่า

$$\sum e_i^2 = (-0.63)^2 + (-0.51)^2 + \ldots + (-0.47)^2 = 9.7978$$

ดังนั้น

$$s_{y.x}^2 = \frac{9.7978}{8}$$

$$= 1.22$$

การหา $\sum e_i^2$ โดยวิธีการข้างบนนั้นเราต้องคำนวณ Y_c ทุกค่า ซึ่งทำให้เสียเวลาและอาจคลาดเคลื่อนได้เมื่อต้องตัดค่าหลังจุดทศนิยมทิ้ง จึงอาจหาโดยวิธีลัดซึ่งแสดงวิธีคำนวณดังต่อไปนี้

เมื่อ $a = \overline{Y} - b\overline{X}$ ก็หาได้ว่า $Y_c = a + bX_i = \overline{Y} + b(X_i - \overline{X})$ ดังนั้น

$$\sum e_i^2 = \sum (Y_i - Y_c)^2$$

$$= \sum [Y_i - \overline{Y} - b(X_i - \overline{X})]^2$$

$$= \sum (Y_i - \overline{Y})^2 - 2b\sum (X_i - \overline{X})(Y_i - \overline{Y}) + b^2 \sum (X_i - \overline{X})^2 \qquad \ldots(5\text{-}17)$$

$$= (Y_i - \overline{Y})^2 - \frac{[\sum (X_i - \overline{X})(Y_i - \overline{Y})]^2}{\sum (X_i - \overline{X})^2} \qquad \ldots(5\text{-}18)$$

หรือ

$$= \left[\sum Y_i^2 - \frac{(\sum Y_i)^2}{n}\right] - \frac{\left[\sum X_i Y_i - (\sum X_i)(\sum Y_i)/n\right]^2}{\sum X_i^2 - \left[(\sum X_i{}^2/n)\right]} \qquad \ldots(5\text{-}19)$$

ในการเปลี่ยนสมการจาก (5-17) เป็น (5-18) นั้นกระทำได้เพราะ b มีค่าดังสมการ (5-14) ดังนั้น เมื่อแทนค่า b ทั้งหมดที่มีอยู่ในสมการ (5-17) จึงได้สมการ (5-18) หรือ (5-19) เราอาจนำสมการเหล่านี้ไปคำนวณหา $\sum e_i^2$ ได้โดยไม่ต้องหาสมการเส้นตรง เช่นจากข้อมูลในตาราง 5.3.2 หาได้ว่า

$$\sum e_i^2 = \left[29{,}652 - \frac{542^2}{10}\right] - \frac{\left[89{,}514 - \frac{(1{,}646)(542)}{10}\right]^2}{271{,}272 - \frac{1{,}646^2}{10}}$$

$$= 9.79$$

ซึ่งค่าที่หาโดยใช้วิธีนี้ จะเที่ยงตรงดีกว่าวิธีการหาโดยการหักลบทีละค่าดังใช้ค่า Y_i และ Y_c ในตาราง 5.3.2 เพราะไม่จำเป็นต้องตัดค่าหลังจุดทศนิยมทิ้ง นอกจากนี้ ขอให้ดูวิธีการวิเคราะห์ความปรวนแปรโดยใช้สมการ (5-27), (5-28) และ (5-29)

วิธีการหาสมการรีเกรซชัน ดังที่กล่าวมาแล้วข้างบน ถ้าจำเป็นต้องคำนวณตัวเลขที่มีค่าสูง ๆ แม้มีความเที่ยงตรงสูง ก็มีความยุ่งยาก วิธีการที่ง่ายในการคำนวณ โดยเฉพาะเมื่อไม่ใช้เครื่องคิดเลข ก็ใช้สมการ (5-13) , (5-14) เพื่อหาสมการรีเกรซชัน ดังแสดงในตาราง 5.3.3

5.5 วาเรียนซ์ของ b และ a

ในขณะนี้อาจกล่าวได้ว่า e_i มีการกระจายแบบปกติ มีค่าเฉลี่ยเท่ากับศูนย์ และวาเรียนซ์เท่ากับ $\sigma^2_{y.x}$ ในทำนองเดียวกันก็พูดได้ว่า b ก็มีการกระจายแบบปกติ มีค่าเฉลี่ยเท่ากับ β และมีวาเรียนซ์ σ^2_b ดังนี้

$$\sigma^2_b = \sigma^2_{y.x} / \sum\left(X_i - \overline{X}\right)^2$$

การได้มาซึ่งวาเรียนซ์ของ b นี้อาจอธิบายได้ง่าย ๆ ว่า ถ้าเราสุ่มตัวอย่างจากประชากรหนึ่ง ๆ หลายครั้ง แต่ละครั้งคำนวณหา b ปรากฏว่า b จะมีค่าขึ้นลง จึงอาจกล่าวได้ว่า b อาจแปรปรวนได้เช่นกัน แต่ก็มีค่าอยู่ใกล้ β ของประชากรนั้น ในการประมาณวาเรียนซ์ของประชากรนั้นเราใช้วาเรียนซ์ของตัวอย่าง ซึ่งแทนด้วย s^2_b สมการในการหา s^2_b แสดงได้ดังนี้ จาก

$$b = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2}$$

$$= \frac{\sum(X_i - \overline{X})Y_i}{\sum(X_i - \overline{X})^2}$$

ดังนั้น

$$s^2_b = (\mathrm{Var}) \frac{\sum(X_i - \overline{X})Y_i}{\sum(X_i - \overline{X})^2}$$

$$= \frac{\sum(X_i - \overline{X})^2}{\sum(X_i - \overline{X})^4}\,[\mathrm{Var}(Y_i)]$$

$$= \frac{s^2_{y.x}}{\sum(X_i - \overline{X})^2} \qquad \ldots(5\text{-}20)$$

ในทำนองเดียวกัน a ก็มีการกระจายแบบปกติ มีค่าเฉลี่ยเท่ากับ α และมีวาเรียนซ์ σ_a^2 แต่เราอาจประมาณให้โดยใช้ s_a^2 จากตัวอย่าง ดังนี้

$$s_a^2 = \frac{s_{y.x}^2(\sum X_i^2)}{n\sum(X_i - \overline{X})^2} \qquad \ldots(5\text{-}21)$$

เราอาจใช้ s_b^2 และ s_a^2 เพื่อทดสอบสมมุติฐานเกี่ยวกับ b และ a ตามลำดับ เช่น ถ้าเราได้ค่า b มาแล้ว แต่ยังสงสัยว่าจะมีขนาดมากพอที่จะสรุปว่า Y มีรีเกรซชันต่อ X หรือไม่ นั่นก็คือทดสอบ $H_0: \beta=0$ โดยมี $H_1: \beta \neq 0$ โดยใช้สถิติทดสอบดังนี้

$$t = \frac{b - \beta}{\sqrt{s_b^2}} \qquad \ldots(5\text{-}22)$$

เช่น จากตัวอย่างเรื่องความสูง และน้ำหนักของนักศึกษานั้นพบว่า $b = 0.88$, $s_{y.x}^2 = 1.22$, และ $\sum(X_i - \overline{X})^2 = 340.4$ ถ้าให้สมมุติฐานเป็น $H_0: \beta = 0$ และ $H_1: \beta \neq 0$ ก็ได้

$$\begin{aligned} t &= \frac{0.88 - 0}{\sqrt{1.22/340.4}} \\ &= 14.69^{**} \end{aligned}$$

โดยเปรียบเทียบค่านี้กับค่าในตาราง t ที่ df = n − 2 = 8 ซึ่งพบว่าผลมีนัยสำคัญยิ่งในทางสถิติ ซึ่งแสดงระดับความแตกต่างโดยใช้เครื่องหมายดอกจัน 2 จุด จึงสรุปได้ว่า $\beta \neq 0$ คือน้ำหนักของคนขึ้นอยู่กับความสูง และเราอาจหาช่วงเชื่อมั่นของ β โดยใช้สมการ

$$b \pm t_{\alpha/2} s_b \qquad \ldots(5\text{-}23)$$

5.6 วาเรียนซ์ของ Y_c

เมื่อเราสุ่มตัวอย่างจากลักษณะที่มีความสัมพันธ์กันมาหลายชุด สมการรีเกรซชันที่คำนวณจากแต่ละตัวอย่างย่อมไม่เท่ากัน เมื่อใช้สมการเหล่านี้ทำนาย Y_c ของแต่ละค่า X ก็ย่อมไม่เท่ากัน อาจกล่าวได้ว่า Y_c มีการกระจายแบบปกติ มีค่าเฉลี่ย $\mu_{y.x}$ และมีวาเรียนซ์ σ_{yc}^2 จากสมการ

$$Y_c = a + bX_i = \overline{Y} + b(X_i - \overline{X})$$

ดังนั้น

$$\begin{aligned} s_{yc}^2 &= \operatorname{Var}\left[\overline{Y} + b(X_i - \overline{X})\right] \\ &= \operatorname{Var}(\overline{Y}) + (X_i - \overline{X})^2 \operatorname{Var}(b) \end{aligned}$$

$$= \frac{s_{y.x}^2}{n} + (X_i - \bar{X})^2 \frac{s_{y.x}^2}{\sum (X_i - \bar{X})^2}$$

$$= s_{y.x}^2 \left[\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right] \qquad \ldots(5\text{-}24)$$

ทั้งนี้ค่า X_i ที่เป็นตัวตั้งในสมการ (5-24) คือ X_i ที่ใช้ทำนาย Y_c ค่านั้น ๆ เมื่อจะทดสอบสมมุติฐานเกี่ยวกับ $\mu_{y.x}$ ก็ใช้สถิติ t ดังนี้

$$t = \frac{Y_c - \mu_{y.x}}{\sqrt{s_{yc}^2}}$$

โดยเปิดค่า t จากตารางที่ df = n - 2 ส่วนช่วงเชื่อมั่นของ $\mu_{y.x}$ หาได้จากสมการ

$$Y_c \pm t_{\alpha/2} s_{yc}$$

เช่น จากตัวอย่างในเรื่องความสูงของนักศึกษา ที่ X = 160, Y_c = 50.15 ดังนั้นหาได้ว่า

$$s_{yc}^2 = 1.22 \left[\frac{1}{10} + \frac{(160 - 164.6)^2}{340.4} \right]$$

$$= 0.20$$

ดังนั้น $s_{yc} = \sqrt{0.20} = 0.45$ ช่วงเชื่อมั่น 95 เปอร์เซ็นต์ของ $\mu_{y.x}$ คือ

$$50.15 \pm 2.306(0.45) = 49.11\ ,\ 51.19$$

ซึ่งอาจเขียนช่วงเชื่อมั่น 95 เปอร์เซ็นต์ ของ $\mu_{y.x}$ เมื่อ X = 160 คือ

$$49.11 < \mu_{y.x} < 51.19$$

5.7 วาเรียนซ์ของ Y_i

ในบางกรณีเราอาจใช้สมการรีเกรซชันเพื่อทำนาย Y_i (ไม่ใช่ Y_c) ของแต่ละค่า X คือต้องการทราบว่าแต่ละค่า X นั้น มี Y_i ในขอบเขตเท่าใด ในกรณีเช่นนี้ การทำนายคือการหาช่วงเชื่อมั่นของ Y_i นั่นเอง นอกจากนี้เราอาจต้องการทำนาย Y_i ซึ่งเป็นเรื่องของอนาคต เช่น ร้านค้าแห่งหนึ่งต้องการทำนายปริมาณการขายสินค้าในปีถัดไป การทำนายเช่นนี้ใช้สมการรีเกรซชันนั่นเอง แต่ตัวทำนายถือเป็น Y_i ค่าหนึ่ง วาเรียนซ์ของ Y_i หาจากส่วนเบี่ยงเบนจาก Y คือ

$$\text{Var} = (Y_c - Y_i) = \text{Var}(Y_c) + \text{Var}(Y_i)$$

ให้ s_{ya}^2 แทน $Var(Y_c - Y_i)$ ดังนั้น

$$s_{ya}^2 = Var(Y_c) + Var(Y_i)$$

ทั้งนี้ $Var(Y_c)$ และ $Var(Y_i)$ แสดงไว้แล้วในสมการ (5-24) และ (5-16) ตามลำดับ ดังนั้น

$$s_{ya}^2 = s_{y.x}^2\left[\frac{1}{n} + \frac{(X_i - \overline{X})^2}{\sum(X_i - \overline{X})^2}\right] + s_{y.x}^2$$

$$= s_{y.x}^2\left[1 + \frac{1}{n} + \frac{(X_i - \overline{X})^2}{\sum(X_i - \overline{X})^2}\right] \qquad \text{...(5-25)}$$

เมื่อจะหาช่วงเชื่อมั่นของ Y_i ก็ใช้สมการ

$$Y_c \pm t_{\alpha/2} s_{ya} \qquad \text{...(5-26)}$$

ซึ่งเปิดตาราง t ที่ df = n - 2

ตัวอย่างเช่น จากตาราง 5.3.2 เมื่อ X = 160, $Y_c = 50.15$ หาได้ว่า

$$s_{ya}^2 = 1.22\left[1 + \frac{1}{10} + \frac{(160 - 164.6)^2}{340.4}\right] = 1.42$$

$$s_{ya} = \sqrt{1.42} = 1.19$$

ที่ df = n - 2 = 10 - 2 = 8 ค่า $t_{0.025} = 2.306$ ดังนั้นช่วงเชื่อมั่นของ Y_i ที่ 95 เปอร์เซ็นต์คือ

$$50.15 \pm (2.306)(1.19)$$

หรืออาจเขียนว่า

$$50.15 - (2.306)(1.19) < Y_i < 50.15 + (2.306)(1.19)$$

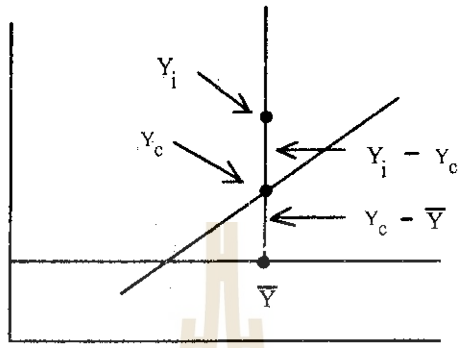
$$47.41 < Y_i < 53.65$$

5.8 การแยกความปรวนแปรของ Y_i

ในการศึกษาเรื่องรีเกรซชันนั้น เราอาจดำเนินต่อไปถึงขั้นที่แยกความปรวนแปรของ Y_i (หรืออีกนัยหนึ่งคือค่า $Y_i - \overline{Y}$) ออกเป็นส่วนย่อย ๆ 2 ส่วน ดังนี้

$$Y_i - \overline{Y} = (Y_i - Y_c) + (Y_c - \overline{Y}) \qquad \text{...(5-27)}$$

ซึ่งอาจแสดงโดยใช้รูป ดังนี้



ซึ่งจะเห็นได้ว่า $Y_i - Y_c = e_i$ นั่นเอง ซึ่งจัดเป็นความปรวนแปรที่ไม่ทราบสาเหตุแน่นอน ส่วน $Y_c - \overline{Y}$ คือค่าเบี่ยงเบนที่เกิดจากรีเกรซชัน หรือเกิดจากการที่ Y ขึ้นต่อ X และอาจพิสูจน์ได้ว่า $Y_c - \overline{Y} = b(X_i - \overline{X})$ ดังนี้

$$Y_c = a + bX_i$$

แต่

$$a = \overline{Y} - b\overline{X}$$

เมื่อแทนค่า a ในสมการบนก็ได้

$$Y_c = \overline{Y} - b\overline{X} + bX_i$$

ดังนั้น

$$Y_c - \overline{Y} = b(X_i - \overline{X}) \qquad \ldots(5\text{-}28)$$

ดังนั้นจากสมการ (5-27) อาจเขียนว่า

$$Y_i = \overline{Y} + e_i + b(X_i - \overline{X})$$

$$e_i = (Y_i - \overline{Y}) - b(X_i - \overline{X})$$

เมื่อยกกำลังสองของค่าเบี่ยงเบนเหล่านี้ แล้วบวกกัน ก็จะได้

$$\sum e_i^2 = \sum (Y_i - \overline{Y})^2 - b\sum (X_i - \overline{X})(Y_i - \overline{Y})$$

ซึ่งอาจเขียนเสียใหม่ว่า

$$\underset{\text{TSS}}{\sum (Y_i - \overline{Y})^2} = \underset{\text{SSR}}{b\sum (X_i - \overline{X})(Y_i - \overline{Y})} + \underset{\text{SSE}}{\sum e_i^2} \qquad \ldots(5\text{-}29)$$

ซึ่งคล้ายกับสมการ (5-18) และ (5-19) นั่นเอง ดังนั้นเห็นได้ว่า เราสามารถแยกความปรวนแปรที่มีอยู่ในข้อมูลออกเป็นส่วน ๆ ตามแหล่งที่มา การแยกความปรวนแปรนี้ เราเรียกว่า การวิเคราะห์วาเรียนซ์[8]

จากสมการ (5-29) ให้ชื่อแหล่งของความปรวนแปร ดังนี้ คือ

(1) e_i^2 เรียกว่า Error SS (หรือ SSE)

(2) $\sum(Y_i - \overline{Y})^2$ คือความปรวนแปร หรือ Total SS (หรือ TSS)

(3) $b\sum(X_i - \overline{X})(Y_i - \overline{Y})$ เป็นความปรวนแปรเนื่องจาก Y ขึ้นต่อ X เราจึงเรียกว่า Regression SS (หรือ SSR)

ดังนั้นจึงสรุปได้ว่า TSS = SSR + SSE ทั้งนี้ df ของ TSS = n - 1, df ของ SSR = 1 ดังนั้น df ของ SSE = n - 2 จึงอาจแสดงตารางวิเคราะห์วาเรียนซ์ได้ดังตาราง 5.8.1

ตาราง 5.8.1 แสดงวิธีการวิเคราะห์ของข้อมูลที่มีรีเกรซชันต่อกัน

Sources of variation	Degrees of freedom (df)	Sum of squares (SS)	Mean square (MS)
Regression	1	SSR	SSR/1
Error	n – 2	SSE	SSE/n – 2
Total	n – 1	TSS	

ซึ่ง MS ได้จากการหาร SS แต่ละบรรทัดด้วย df ในบรรทัดเดียวกัน

จากตัวอย่างเกี่ยวกับรีเกรซชันของน้ำหนักต่อความสูงของนักศึกษา สามารถวิเคราะห์วาเรียนซ์ ได้ดังนี้

$$\text{TSS} = \sum Y_i^2 - \sum(Y_i)^2/n = 29{,}652 - (542)^2/10 = 275.60$$

$$\begin{aligned}\text{SSR} &= b\left[\sum X_iY_i - (\sum X_i)(\sum Y_i)/n\right] \\ &= 0.88[89{,}514 - (1{,}646)(542)/10] \\ &= 264.70\end{aligned}$$

$$\begin{aligned}\text{SSE} &= \text{TSS} - \text{SSR} \\ &= 275.60 - 264.70 \\ &= 10.90\end{aligned}$$

การที่ SSE ที่หาโดยวิธีนี้แตกต่างจากค่าก่อน ๆ เนื่องมาจากการตัดจุดทศนิยมในการคำนวณค่า b นั่นเอง เมื่อนำผลการคำนวณลงตารางวิเคราะห์วาเรียนซ์ ก็จะได้ดังตาราง 5.8.2

ตาราง 5.8.2 ตารางวิเคราะห์วาเรียนซ์รีเกรซชันของความสูงและน้ำหนัก

Sources of Variation	df	SS	MS	F
Regression	1	264.70	264.70	194.63 **
Error	8	10.90	1.36	
Total	9	275.60		

ค่า 1.36 คือ $s^2_{y.x}$ นั่นเอง ส่วนค่า F ได้จาก

$$F = \frac{\text{MS Regression}}{\text{MS Error}}$$

การวิเคราะห์วาเรียนซ์มีวัตถุประสงค์เพื่อทดสอบสมมุติฐาน $H_0 : \beta = 0$ ดังที่เคยทดสอบโดยใช้ค่า t มาแล้วในตอน 5.4 นั่นเอง ทั้งนี้เปรียบเทียบค่า F นี้กับค่า F ในตาราง ผ.9 โดยมี df 1 และ 8 ก็พบว่า $F = 194.63 > F_{(0.01)}(1, 8) = 11.26$ จึงปฏิเสธ H_0 เราใส่เครื่องหมายดอกจัน 2 จุด เพื่อแสดงว่าเราปฏิเสธ H_0 ที่ระดับความแตกต่าง 0.01

5.9 แบบฝึกหัด

1. จงอธิบายถึงข้อกำหนด (assumption) ที่เกี่ยวข้องกับสมการรีเกรซชันของประชากร
2. จงแสดงให้เห็นว่าสมการรีเกรซชันของตัวอย่างนั้นอาจเขียนว่า

   $Y_c = a + bX_i$ หรือ $Y_c = \overline{Y} + b(X_i - \overline{X})$ ก็ได้
3. จงพิสูจน์ให้เห็นว่า

   ก. $\sum(X_i - \overline{X})(Y_i - \overline{Y}) = \sum(X_i - \overline{X})Y_i$

   ข. $\sum(X_i - \overline{X})(Y_i - \overline{Y}) = \sum X_i Y_i - (\sum X_i)(\sum Y_i)/n$

   ค. $b^2\sum(X_i - \overline{X})^2 = b\sum(X_i - \overline{X})(Y_i - \overline{Y})$
4. จงอธิบายถึงแหล่งที่มาหรือสมมุติฐานที่ทำให้เกิดค่าดังต่อไปนี้ ในรีเกรซชัน :

   a, b, $s^2_{y.x}$, s^2_b, s^2_{yc} และ s^2_a
5. จากการวัดความสูงของถั่วเหลืองที่มีอายุต่างกันได้ผลดังนี้

อายุ (สัปดาห์)	1	2	3	4	5	6	7
ความสูง (ซม.)	5	13	16	23	33	38	40

   ก. เลขชุดใดเป็นค่า X และ Y ทำไมคิดว่าเป็นเช่นนั้น

   ข. นำค่า X และ Y มาพลอตลงบนกระดาษกราฟ โดยให้ค่า X เดินตามแกน X และ Y เดินตามแกน Y

ค. คำนวณค่า a และ b และแสดงสมการรีเกรชชันของตัวอย่างนี้

ง. คำนวณหา $Y_c, \sum e_i^2$ และ $s_{y.x}^2$ จงเปรียบเทียบกับ $s_{y.x}^2$ ที่หาโดยใช้สมการ SEE = TSS − SSR

จ. จงหา s_b^2 , ทดสอบสมมุติฐาน $H_0: \beta = 0$ และหาช่วงเชื่อมั่นที่ 95 เปอร์เซ็นต์

ฉ. จงแสดงตารางวิเคราะห์วาเรียนซ์ ทดสอบสมมุติฐาน $H_0: \beta = 0$

6. จากการสอบของนักศึกษา 9 คน ปรากฏว่าได้คะแนนจากการสอบกลางเทอม (X) และสอบไล่ (Y) ดังนี้ (%)

X	77	50	71	72	81	94	96	99	67
Y	82	66	78	34	47	85	99	99	68

ก. จงหาสมการรีเกรชชัน

ข. จงทำนายคะแนนสอบไล่ของนักศึกษาที่สอบกลางเทอมได้ 85%

7. จากการศึกษาถึงการละลายของสารเคมีชนิดหนึ่ง (Y) ในน้ำ ปรากฏว่าปริมาณการละลายขึ้นอยู่กับอุณหภูมิดังนี้

X (°C)	0	15	30	45	60	75
Y (กรัม)	8	12	25	31	44	48
	6	10	21	33	39	51
	8	14	24	28	42	44

ก. จงหาสมการรีเกรชชัน

ข. จงประมาณการละลาย (เป็นกรัม) ในน้ำอุณหภูมิ 50°C

8. ข้อมูลกลุ่มหนึ่งมีค่าดังนี้

X	0	1	2	3	4
Y	130	145	150	165	170

จงคำนวณ TSS, SSR, SSE และทดสอบ $H_0: \beta = 0$

9. จากแบบฝึกหัดที่ 6 จงคำนวณหาช่วงเชื่อมั่น 0.95 ของ

ก. α

ข. β

ค. ของ $\mu_{y.x}$ เมื่อ X = 80

คำในบท

(1) cause and effect relationship, (2) independent variable, (3) dependent variable, (4) regression coefficient, (5) least square, (6) least square line, (7) covariance, (8) analysis of variance

บทที่ 6
สหสัมพันธ์

6.1 คำนำ

ในบทที่ 5 เราได้ศึกษาถึงการรีเกรซชันเส้นตรง และในการศึกษานั้นมีการกำหนดไว้ว่าตัวเลขชุดหนึ่ง (X) เป็นค่าอิสระและคงที่ คือ กำหนดไว้ว่ามีค่าเป็นเท่าไร แล้วสุ่มหาตัวแปรอีกค่าหนึ่ง (Y) ที่เกี่ยวข้อง หรือขึ้นอยู่ หรืออยู่ใต้อิทธิพลของค่านั้น

ในบทนี้จะได้ศึกษาถึงความสัมพันธ์ระหว่างค่าสังเกตในอีกแบบหนึ่ง คือตัวแปรทั้ง 2 ค่านั้นต่างก็ได้มาจากการสุ่ม ไม่มีการกำหนดค่าใดค่าหนึ่งไว้เป็นการล่วงหน้า คือสุ่มตัวอย่างมาแล้ว ก็ทำการวัดค่าทั้งสองพร้อม ๆ กันไป ประชากรอันเป็นที่มาของตัวอย่างนี้คือ ประชากรที่มีการกระจาย 2 ทาง[1] คือเลขหรือข้อมูลแต่ละชุดมีการกระจายของมันเอง ในประชากรชนิดนี้เราไม่มีการกำหนดว่าค่าหนึ่งขึ้นอยู่กับอีกค่าหนึ่ง เช่น ความสูงของบุตรชาย-หญิง ของครอบครัวต่าง ๆ มักมีความสัมพันธ์กัน แต่เราก็ไม่อาจกล่าวลงไปได้ว่า ความสูงของบุตรชายขึ้นอยู่กับความสูงของบุตรหญิง หรือความสูงของบุตรหญิงขึ้นอยู่กับความสูงของบุตรชาย ค่าสังเกตที่เป็นคู่ ๆ ซึ่งมีความผันแปรที่สัมพันธ์กัน คือ เพิ่ม-ลดด้วยกัน หรือค่าหนึ่งเพิ่มแต่อีกค่าหนึ่งลด เรียกว่าเป็นตัวเลขที่ผันแปรไปด้วยกัน การศึกษาความเกี่ยวข้องของ ตัวเลขแบบนี้ เราใช้วิธีวิเคราะห์สหสัมพันธ์[2] ทั้งนี้เพื่อให้รู้ว่าตัวเลขมีความสัมพันธ์กันแน่นแฟ้นเพียงใด นอกจากนั้นก็ทำให้รู้ว่าตัวเลขผันแปรไปด้วยกัน หรือไปในทางตรงกันข้ามกัน

การศึกษาเกี่ยวกับสหสัมพันธ์นั้น ถ้าหากมีตัวแปรเกี่ยวข้องเพียง 2 ชุด ก็เรียกว่าเป็นสหสัมพันธ์อย่างง่าย[3] แต่ถ้ามีตัวแปรเกี่ยวข้องมากกว่า 2 ชุด ก็เรียกว่าเป็น พหุสหสัมพันธ์[4] ในขั้นนี้เราศึกษาเฉพาะสหสัมพันธ์อย่างง่ายแต่เพียงอย่างเดียวเท่านั้น

6.2 ดรรชนีสหสัมพันธ์

ดังที่เราได้กล่าวถึงในสมการ (5-27) นั้น จะเห็นได้ว่า ความแตกต่างระหว่าง Y_i และ $\overline{Y}$ นั้นสามารถแบ่งออกได้เป็น 2 ส่วน คือ

$$Y_i - \overline{Y} = (Y_i - Y_c) + (Y_c - \overline{Y})$$

ซึ่งอาจพิสูจน์ต่อไปได้ว่า

$$\sum (Y_c - \overline{Y})^2 = \sum (Y_i - \overline{Y})^2 - \sum (Y_i - Y_c)^2$$

หารสมการนี้ด้วย $\sum(Y_i-\overline{Y})^2$ ก็ได้

$$\frac{\sum(Y_c-\overline{Y})^2}{\sum(Y_i-\overline{Y})^2} = \frac{\sum(Y_i-\overline{Y})^2-\sum(Y_i-Y_c)^2}{\sum(Y_i-\overline{Y})^2}$$

อัตราส่วนระหว่าง $\sum(Y_c-\overline{Y})^2$ และ $\sum(Y_i-\overline{Y})^2$ คือ coefficient of determination ซึ่งแทนด้วย r^2 ทั้งนี้ r^2 หมายถึงอัตราส่วนของความปรวนแปรอันสืบเนื่องมาจากอิทธิพลของตัวแปรอิสระ (X) ดังนั้น

$$r^2 = \frac{\sum(Y_c-\overline{Y})^2}{\sum(Y_i-\overline{Y})^2} = \frac{\sum(Y_i-\overline{Y})^2-\sum(Y_i-Y_c)^2}{\sum(Y_i-\overline{Y})^2} \qquad \ldots(6\text{-}1)$$

จะเห็นได้ว่าเมื่อ $Y_i=Y_c$ (คือเมื่อค่าที่สังเกตและค่าคาดหมายเท่ากัน) ก็หาได้ว่า $r^2=1$ แต่เมื่อ $Y_c=\overline{Y}$ (คือ X ไม่มีอิทธิพลต่อ Y เลย) ก็หาได้ ว่า $r^2=0$ ดังนั้นจึงกล่าวได้ว่า r^2 มีค่าอยู่ระหว่าง $0<r^2<1$

เมื่อถอดรากสองของ r^2 ก็จะได้ r ซึ่งเรียกว่าดรรชนีสหสัมพันธ์ของตัวอย่าง คือ

$$r = \sqrt{\frac{\sum(Y_c-\overline{Y})^2}{\sum(Y_i-\overline{Y})^2}} \qquad (1)$$

เราจำเป็นต้องดัดแปลงสมการข้างบนอีกเล็กน้อย โดยการยกกำลังสองสมการ (5-18) ดังนี้

$$\begin{aligned}\sum(Y_c-\overline{Y})^2 &= b\sum(X_i-\overline{X})(Y_i-\overline{Y})\\ &= \frac{\left[\sum(X_i-\overline{X})(Y_i-\overline{Y})\right]^2}{\sum(X_i-\overline{X})^2} \qquad (2)\end{aligned}$$

เมื่อแทนค่าจากสมการ (2) ใน (1) ก็ได้

$$r = \frac{\sum(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum(X_i-\overline{X})^2\sum(Y_i-\overline{Y})^2}} \qquad \ldots(6\text{-}2)$$

ซึ่งเหมือนกับสมการ

$$r = \frac{\sum X_iY_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sqrt{\left[\sum X_i^2-(\sum X_i)^2/n\right]\left[\sum Y_i^2-(\sum Y_i)^2/n\right]}} \qquad \ldots(6\text{-}3)$$

$$= \frac{\text{Covariance }(X, Y)}{s_x \, s_y} \qquad \ldots(6\text{-}4)$$

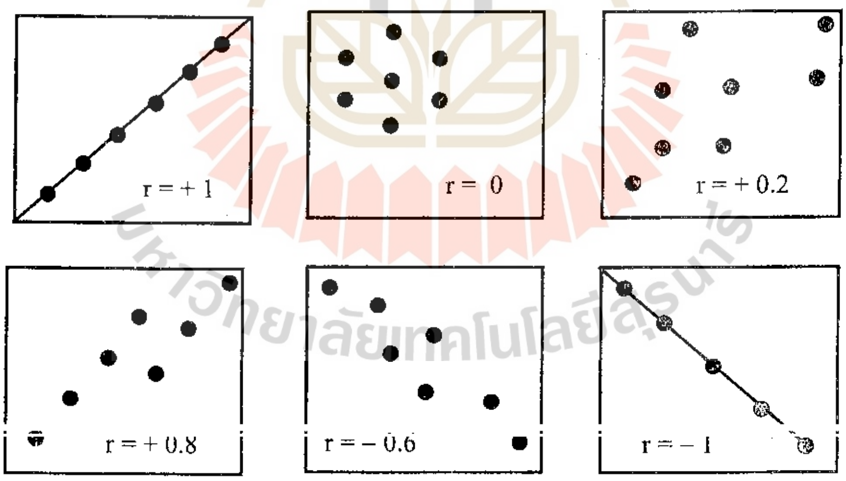
ดรรชนีสหสัมพันธ์ r เป็นค่าของตัวอย่าง ถ้าเป็นดรรชนีสหสัมพันธ์ของประชากรก็แทนด้วย ρ (อักษรกรีกอ่านว่า rho) เราอาจอธิบายคุณสมบัติของดรรชนีสหสัมพันธ์ r ได้ดังนี้

(1) เป็นตัวเลขไม่มีหน่วย

(2) มีค่าอยู่ระหว่าง 0 และ 1 ถ้าเลข 2 ชุดไม่มีความสัมพันธ์กันเลย คือมีโควาเรียนซ์เป็นศูนย์ ดังนั้น r (หรือ ρ) = 0 แต่ถ้าตัวเลขมีความสัมพันธ์กันอย่างสมบูรณ์ ก็จะให้ค่า r (หรือ ρ) = 1

(3) ดรรชนีสหสัมพันธ์อาจมีค่าเป็น "บวก" หรือ "ลบ" ก็ได้ เมื่อเลข 2 ชุด เพิ่มด้วยกันก็มีค่าเป็น "บวก" แต่เมื่อค่าหนึ่งเพิ่มขึ้นอีกค่าหนึ่งลดลงก็มีค่าเป็น "ลบ" ดังนั้นในกรณีที่เลข 2 ชุด มีสหสัมพันธ์ต่อกัน r (หรือ ρ) จะมีค่าอยู่ระหว่าง −1 และ +1

รูป 6.2.1 เป็นการกระจายของจุด X, Y ที่พลอตเป็นคู่ ๆ เพื่อแสดงให้เห็นว่าเมื่อ r มีขนาดต่าง ๆ กันนั้น จุดจะมีการกระจัดกระจายแบบใดบ้าง เมื่อ r = 0 การกระจายของจุดจะไม่มีแนวหรือทิศทางที่แน่นอน เมื่อ r = 1 เลข 2 ชุด ก็มีสหสัมพันธ์อย่างแน่นแฟ้น คือ เพิ่มไปด้วยกันเป็นขั้น ๆ ถ้า r มีค่าเป็นลบ การกระจายของจุดจะลดลงในด้านขวามือ ถ้า r มีค่าเท่ากัน ไม่ว่าจะเป็นบวกหรือลบ ก็ตาม ก็แสดงว่าความแน่นแฟ้นของความสัมพันธ์เท่ากัน



รูปที่ 6.2.1 การกระจายของค่า X และ Y จากตัวอย่างซึ่งให้ดรรชนีสหสัมพันธ์ต่าง ๆ กัน

ค่า r มิได้ขึ้นอยู่กับมาตราวัดของ X และ Y ดังนั้นอาจทำให้การคำนวณง่ายขึ้น ถ้าหากว่าได้มีการลบหรือหารข้อมูลดิบด้วยค่าบางค่า การลบหรือหารอาจจะทำเฉพาะ X หรือ Y หรือทั้ง 2 ค่าก็ได้ ทั้งนี้เพราะว่าเมื่อข้อมูลมีค่าสูง การคำนวณก็มีความยุ่งยาก เมื่อลดค่าให้ต่ำลงการคำนวณก็จะง่ายขึ้น โดยเฉพาะอย่างยิ่งเมื่อคำนวณ โดยไม่อาศัยเครื่องคิดเลข

ตัวอย่าง 6.2.1

เมื่อทำการสุ่มตัวอย่างครอบครัวที่มีบุตรชาย-หญิงมา 11 ครอบครัว แล้ววัดความสูงบุตรชาย-หญิงมาเป็นคู่ ๆ ดังตารางข้างล่าง (ความสูงเป็นนิ้ว) จากข้อมูลดังกล่าวนี้จงคำนวณ r

ครอบครัวที่	1	2	3	4	5	6	7	8	9	10	11
บุตรชาย (X)	71	68	66	67	70	71	70	73	72	65	66
บุตรหญิง (Y)	69	64	65	63	65	62	65	64	66	59	62

วิธีทำ $n = 11, \overline{X} = 69, \overline{Y} = 64, \sum(X_i - \overline{X})^2 = 74, \sum(Y_i - \overline{Y})^2 = 66,$

$\sum(X_i - \overline{X})(Y_i - \overline{Y}) = 39$ **ดังนั้นจากสมการ (6-2) ก็หาได้ว่า**

$$r = \frac{39}{\sqrt{(74)(66)}}$$

$$= 0.558$$

ตัวอย่าง 6.2.2

จากผลการสอบไล่วิชาภาษาอังกฤษและคณิตศาสตร์ของนักศึกษา เมื่อสุ่มมา 5 คน ก็ได้ผลดังนี้ และจงคำนวณค่า r

อังกฤษ , X	คณิตศาสตร์ , Y	X – 70	Y – 80
72	83	2	3
75	84	5	4
73	84	3	4
77	88	7	8
78	89	8	9

6.3 ความสัมพันธ์ระหว่าง b และ r

อันที่จริงแล้ว ดรรชนีสหสัมพันธ์ r คือค่าเฉลี่ยทางเรขาคณิตของ b และ b′ เมื่อ b และ b′ คือสัมประสิทธิ์ของรีเกรซชันของ Y ต่อ X และของ X ต่อ Y ตามลำดับ ดังนั้น

$$r = \sqrt{b \cdot b'} \qquad \ldots(6\text{-}5)$$

ซึ่งในข้อมูลที่สุ่มมาจากประชากรที่มีการกระจาย 2 ทางนั้น สมการรีเกรซชันอาจเขียนได้ทั้ง 2 แบบ ซึ่งคาดว่า Y มีรีเกรซชันต่อ X หรือ X มีรีเกรซชันต่อ Y ก็ได้ คือ

$$Y_c = a + bX_i$$
$$X_c = a' + b'Y_i$$

เมื่อ

$$b = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2},\ b' = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(Y_i - \overline{Y})^2}$$

ดังนั้น

$$b.b' = \frac{\left[\sum(X_i - \overline{X})(Y_i - \overline{Y})\right]^2}{\sum(X_i - \overline{X})^2 \sum(Y_i - \overline{Y})^2} = r^2$$

ดังนั้นจะได้ว่า

$$r = \sqrt{b \cdot b'} \qquad \ldots(6\text{-}5)$$

ความสัมพันธ์ดังสมการ (6-5) นั้น ต้องมีการสมมุติเสียก่อนว่าเราอาจใช้วิธีวิเคราะห์รีเกรซชันกับข้อมูลที่มีการกระจาย 2 ทางก็ได้ คืออาจหารีเกรซชันได้ทั้ง 2 ทาง ไม่ว่า Y ต่อ X หรือ X ต่อ Y ถึงแม้การสมมุตินั้นไม่สมเหตุสมผล เช่นเราไม่อาจกล่าวว่าความสูงของบุตรชายขึ้นอยู่กับความสูงของบุตรหญิง หรือความสูงของบุตรหญิงขึ้นอยู่กับความสูงของบุตรชายก็ตาม

ดังนั้นเราควรให้ข้อสรุปว่า ถ้า X และ Y มีความเกี่ยวข้องกัน คือโควาเรียนซ์ไม่เท่ากับศูนย์ แต่ไม่มีทางว่าจะขึ้นต่อกันแต่อย่างใด การหาความสัมพันธ์โดยใช้วิธีวิเคราะห์สหสัมพันธ์นับว่าเหมาะสมที่สุด

6.4 การทดสอบสมมุติฐาน $\rho = 0$

ค่า r อาจผันแปรขึ้นลงตามการสุ่ม คือเมื่อสุ่มตัวอย่างมาครั้งหนึ่ง คำนวณได้ r ค่าหนึ่ง เมื่อสุ่มตัวอย่างมาใหม่ก็คำนวณได้ r อีกค่าหนึ่ง ยิ่งกว่านั้นในประชากรที่ค่า X และ Y ไม่มีความสัมพันธ์กันเลย ($\rho = 0$) ถ้าสุ่มมาหลาย ๆ ครั้งก็ยังพบว่า r อาจมีค่าเป็นศูนย์ หรือมีค่าเป็นบวกและ

ลบก็ได้ ซึ่งการกระจายของค่า r เมื่อ $\rho = 0$ ขึ้นอยู่กับขนาดของตัวอย่าง การกระจายของค่า r มักเป็นแบบปกติ โดยเฉพาะอย่างยิ่งเมื่อ n มีค่าสูง และประชากรมี $\rho = 0$ แต่เมื่อ $\rho \neq 0$ การกระจายของค่า r จะมีรูปเบ้ คือถ้า ρ มีค่าเป็นบวกก็โย้ไปทางขวามือ ถ้า ρ เป็นลบก็โย้ไปทางซ้ายมือ

ในการทดสอบสมมุติฐาน $H_0 : \rho = 0, H_1 : \rho \neq 0$ นั้น อาจกระทำโดยการนำค่า r ที่คำนวณได้ไปเปรียบเทียบกับค่า r จากตาราง (ตาราง ผ.6) ในตารางนี้แสดงว่าค่าตัดสินที่ระดับ 5 และ 1 เปอร์เซ็นต์ ของค่า r จากประชากรที่มี $\rho = 0$ จงสังเกตว่าในการเปิดตารางใช้ $df = n - 2$ เสมอ ในการทดสอบนั้นเราต้องตัดเครื่องหมายบวก-ลบของค่า r ออกไปก่อน เพราะการกระจายของ r เป็นแบบสมมาตรนั่นเอง เครื่องหมายจึงไม่มีความสำคัญแต่อย่างใด

ในการทดสอบสมมุติฐาน $H_0 : \rho = 0$ อาจใช้สถิติทดสอบ t ก็ได้ ทั้งนี้เดิมทีเดียวนั้น ในการทดสอบค่า b เราใช้สถิติ $t = b/s_b$ แต่เราสามารถพิสูจน์ได้ว่า

$$t = \frac{b}{s_b} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

ดังนั้นในการทดสอบสมมุติฐาน $H_0 : \rho = 0, H_1 : \rho \neq 0$ ก็ใช้ค่า t จากสมการ

$$t = \frac{r - \rho}{\sqrt{\frac{1-r^2}{n-2}}} \qquad \ldots(6\text{-}6)$$

ซึ่งมี $df = n - 2$

ตัวอย่าง 6.4.1

จากตัวอย่างในเรื่องความสูงของบุตรชาย-หญิง จาก 11 ครอบครัว ซึ่งแสดงไว้ในตอน 6.2 คำนวณได้ว่า $r = 0.558$ จงทดสอบสมมุติฐาน $H_0 : \rho = 0$ ที่ระดับความแตกต่าง 0.05

วิธีทำ **(1) สมมุติฐาน** $H_0 : \rho = 0$

$H_1 : \rho \neq 0$

(2) กฎตัดสิน : ปฏิเสธสมมุติฐาน (H_0) ที่ระดับความแตกต่าง 0.05 ถ้า t ที่คำนวณได้ไม่อยู่ระหว่าง – 2.262 และ 2.262 ทั้งนี้

$$t = \frac{r - \rho}{\sqrt{\frac{1-r^2}{n-2}}}$$

(3) คำนวณ : $r = 0.558$, $\rho = 0$ (ตาม H_0) , $r^2 = 0.311$ และ $n-2 = 9$ ดังนั้น

$$t = \frac{0.558}{\sqrt{\frac{1-0.311}{9}}}$$

$$= 2.01$$

(4) $t = 2.01$ อยู่ระหว่างช่วง -2.262 และ 2.262 จึงยอมรับ H_0 ซึ่งแสดงว่าไม่มีสหสัมพันธ์ระหว่างความสูงของบุตรชาย-หญิง

มีข้อที่น่าสังเกตประการหนึ่งที่ควรจะกล่าวถึง คือ มีการสอดคล้องกันระหว่างค่า t จากตาราง ผ.4 และ r จากตาราง ผ.6 ทั้งนี้เปลี่ยนค่า r โดยใช้สมการ (6-6) แล้วได้ค่า t ที่ระดับความแตกต่างเดียวกัน เช่น ค่า r ที่ df 7 = 0.666 เมื่อแปลงค่าเป็น t ก็จะได้

$$t = (0.666)(\sqrt{7}) / \sqrt{[1-(0.66)^2]} = 2.345$$

ซึ่งเท่ากับค่า t จากตาราง ผ.4 ที่มี df = 7 ของระดับความแตกต่าง 0.05 พอดี ดังนั้นการทดสอบ $H_0 : \rho = 0$ อาจใช้วิธีเปิดตารางหรือใช้ค่า t ดังสมการ (6-6) ก็ได้

6.5 ทดสอบสมมุติฐาน $\rho = k \neq 0$

ในกรณีที่ $H_0 : \rho \neq 0$ แต่มีค่าเป็นอย่างอื่น การกระจายของ r ก็จะมีความเบ้ คือไม่กระจายแบบปกติเหมือนเมื่อ $\rho = 0$ ดังนั้นจะทดสอบโดยใช้ค่า t หรือใช้ตารางดังตอน 6.4 ย่อมไม่ได้

เมื่อ $\rho \neq 0$ แต่มีค่าเป็นอย่างอื่น เช่น $\rho = 0.5$ การทดสอบกระทำโดยการเปลี่ยนค่า r เป็น Z โดยใช้สมการ

$$Z_r = \frac{1}{2}\log_e\left[\frac{1+r}{1-r}\right] \qquad \ldots(6\text{-}7)$$

$$= \frac{1}{2}(2.3026)\log\frac{1+r}{1-r}$$

ซึ่งมีการกระจายแบบปกติ มีค่าเฉลี่ย

$$Z_\rho = \frac{1}{2}\log_e\left[\frac{1+\rho}{1-\rho}\right]$$

และวาเรียนซ์

$$\sigma_Z^2 = 1/(n-3)$$

ดังนั้นแทนที่จะทดสอบค่า r เรากลับทดสอบค่า Z_r คือใช้สถิติ

$$Z = \frac{Z_r - Z_\rho}{\sigma_Z} \qquad \text{...(6-8)}$$

การแปลงค่า r เป็น Z_r และ ρ เป็น Z_ρ อาจกระทำโดยใช้ตาราง ผ.7

ตัวอย่าง 6.5.1

สมมุติว่ามีตัวอย่าง ซึ่งมีจำนวนคู่ n = 12 และหาได้ว่า r = 0.7 จงทดสอบว่าตัวอย่างนี้มาจากประชากร ซึ่งมี $\rho = 0.5$

วิธีทำ (1) สมมุติฐาน : $H_0 : \rho = 0.5$

$H_1 : \rho \neq 0.5$

(2) กฎตัดสิน : ปฏิเสธสมมุติฐาน (H_0) ที่ระดับความแตกต่าง 0.05 ถ้า Z ที่คำนวณได้ ไม่อยู่ระหว่าง – 1.96 และ + 1.96 ไม่เช่นนั้นก็ยอมรับ H_0 ทั้งนี้

$$Z = \frac{Z_r - Z_\rho}{\sigma_Z}$$

(3) คำนวณ :

$$Z_r = \frac{1}{2}(2.3026)\log\left[\frac{1+0.7}{1-0.7}\right] = 0.867$$

$$Z_\rho = \frac{1}{2}(2.3026)\log\left[\frac{1+0.5}{1-0.5}\right] = 0.549$$

ซึ่งค่าดังกล่าวนี้อาจเปิดได้จากตาราง ผ. 7 ดังนั้นเมื่อแทนค่าในสมการก็ได้

$$Z = \frac{0.867 - 0.549}{\sqrt{1/(12-3)}} = 0.954$$

(4) Z = 0.954 : ซึ่งอยู่ระหว่าง – 1.96 และ + 1.96 จึงยอมรับสมมุติฐานที่ระดับความแตกต่าง 0.05

จงสังเกตว่า ค่า Z_r (หรือ Z_ρ) จากตารางนั้นจะมีค่าเป็นบวกอยู่เสมอ ดังนั้นในการแปลงค่า r ไม่มีผลใด ๆ ต่อเครื่องหมายบวก - ลบ คือนำเครื่องหมายบวก - ลบ จากค่า r ไปใช้กับค่า Z ได้ทันที

6.6 ทดสอบสมมุติฐาน $\rho_1 = \rho_2$

ในบางครั้งเราจำเป็นที่จะต้องทดสอบว่า ค่า r สองค่านั้นสุ่มมาจากประชากรเดียวกันหรือไม่ ในการทดสอบกระทำโดยการแปลงค่า r เป็น Z โดยใช้การคำนวณหรือตารางดังที่กล่าวมาแล้ว ในตอน 6.5 นั่นเอง ต่อจากนั้นก็ทดสอบว่าความแตกต่างระหว่างค่า Z นั้นมีนัยสำคัญหรือไม่ การทดสอบกระทำโดยใช้สมการ

$$Z = \frac{(Z_{r1} - Z_{r2}) - (Z_{\rho 1} - Z_{\rho 2})}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} \qquad \ldots(6\text{-}9)$$

ตัวอย่าง 6.6.1

สมมุติว่ามีการสุ่มตัวอย่างเป็นคู่ ๆ มา 2 ชุด ขนาดของตัวอย่าง $n_1 = n_2 = 30$ คำนวณได้ว่า $r_1 = -0.62$ และ $r_2 = -0.41$ จงทดสอบที่ระดับความแตกต่าง 0.05 ว่าตัวอย่างนี้มาจากประชากรเดียวกัน

วิธีทำ (1) สมมุติฐาน : $H_0 : \rho_1 = \rho_2$

$H_1 : \rho_1 \neq \rho_2$

(2) กฎตัดสิน : ปฏิเสธสมมุติฐาน (H_0) ที่ระดับความแตกต่าง 0.05 ถ้าค่า Z ที่คำนวณได้ไม่อยู่ระหว่าง -1.96 และ $+1.96$ มิฉะนั้นก็ยอมรับ H_0, ทั้งนี้สถิติทดสอบคือ

$$Z = \frac{(Z_{r1} - Z_{r2}) - (Z_{\rho 1} - Z_{\rho 2})}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$$

(3) คำนวณ :

ตัวอย่างที่	ขนาด	r	Z_r	$1/n-3$
1	30	-0.62	-0.725	0.037
2	30	-0.41	-0.436	0.037
			ความแตกต่าง = -0.289	ผลบวก = 0.074

$\sigma_Z = \sqrt{0.074} = 0.272$ ดังนั้น $Z = -0.289/0.272 = -1.06$

จากการทดสอบพบว่า $Z = -1.06$ อยู่ระหว่าง -1.96 และ $+1.96$ จึงยอมรับสมมุติฐาน (H_0) คือสรุปได้ว่า r ทั้งสองนี้สุ่มมาจากประชากรเดียวกัน

6.7 ช่วงความเชื่อมั่นของ ρ

ในเมื่อ Z_r มีการกระจายแบบปกติ ดังนั้นก็อาจหาช่วงเชื่อมั่นของ Z_ρ ได้โดยง่าย คือ แปลงค่า r ให้เป็น Z_r เสียก่อน แล้วหาช่วงเชื่อมั่นของ Z_ρ ต่อจากนั้นแปลงช่วงเชื่อมั่นของ Z_ρ เป็นช่วงความเชื่อมั่นของ ρ โดยใช้ตารางแปลงค่า Z เป็นค่า r (ตาราง ผ.8)

ตัวอย่าง 6.7.1

สมมุติว่าสุ่มตัวอย่างขนาด n = 12 มาตัวอย่างหนึ่ง แล้วคำนวณได้ว่า r = 0.7 จงหาช่วงเชื่อมั่นที่ 90 เปอร์เซ็นต์ของ ρ

วิธีทำ $E(Z_r) = Z$

$Var(Z_r) = s_r^2 = 1/(n-3)$

สมการสำหรับหาช่วงเชื่อมั่นของ Z_ρ **คือ**

$P(Z_r - Z_{\alpha/2}\sigma_r < Z_\rho < Z_r + Z_{\alpha/2}\sigma_r) = 1-\alpha$

ทั้งนี้ $Z_{\alpha/2} = Z_{0.05} = 1.645$ **และประมาณ** σ_r **และ** s_r

ซึ่ง $= 1/\sqrt{12-3} = 0.33$ **ดังนั้นช่วงเชื่อมั่นของ** Z_ρ **คือ**

$0.867 - (1.645)(0.33) < Z_\rho < 0.867 + (1.645)(0.33)$

$= 0.324 < Z_\rho < 1.410$

เมื่อแปลง Z_ρ **เป็น** ρ **ก็จะได้**

$= 0.309 < \rho < 0.889$

6.8 สหสัมพันธ์ของลำดับ

การคำนวณค่า r ตามวิธีที่กล่าวมาแล้ว สามารถกระทำได้เมื่อการกระจายของประชากรเป็นแบบปกติเท่านั้น ในกรณีที่ประชากรมีการกระจายแบบอื่น ถ้าต้องการหาว่า X และ Y มีสหสัมพันธ์หรือไม่ ก็อาจกระทำได้โดยใช้สหสัมพันธ์ของลำดับ[5] ซึ่งเสนอโดย Spearman ในปี ค.ศ.1904 วิธีการนี้ยึดถือเอาลำดับตามความมากน้อยของข้อมูล ไม่ขึ้นอยู่กับการกระจายของ X และ Y แต่อย่างใด เราเรียกสถิติแบบนี้ว่า สถิติที่ไม่ใช้พารามิเตอร์[6] หรือสถิติเกี่ยวกับการกระจายทุกรูปแบบ (distribution free statistic) ดรรชนีสหสัมพันธ์ ซึ่งคำนวณโดยใช้วิธีนี้ เรียกว่าดรรชนีสหสัมพันธ์ของลำดับ[7] และแทนด้วย r_s ซึ่งหาได้จากสมการ

$$r_s = 1 - \frac{6(\sum d^2)}{n(n^2-1)} \qquad \ldots(6\text{-}10)$$

ในเมื่อ d คือ ความแตกต่างระหว่างค่าบอกลำดับของ X และ Y ในคู่เดียวกัน ตัวอย่างเช่น สมมุติว่ามีการสุ่มนักศึกษามา 5 คน แล้วทำการสำรวจคะแนนจากการสอบในวิชาคณิตศาสตร์และวิชาชีววิทยาของแต่ละคน ปรากฏว่าได้ผลดังนี้

คณิตศาสตร์ X	ชีววิทยา Y	ลำดับ X	ลำดับ Y	ลำดับ d = X – Y	d^2
85	93	2	1	1	1
60	75	4	3	1	1
73	65	3	4	– 1	1
40	50	5	5	0	0
90	80	1	2	– 1	1

เราอาจต้องการที่จะตรวจสอบว่า คะแนนทั้งสองรายวิชานี้มีสหสัมพันธ์กันหรือไม่ ถ้ามีก็เพิ่มลดไปด้วยกันหรือคนละทาง ก็กระทำโดยการให้ลำดับตามความสำคัญมากน้อยแก่ข้อมูลแต่ละชุดแล้วคำนวณหา r_s ดังนี้

$$r_s = 1 - \frac{6(4)}{5(25-1)} = 0.8$$

สมการ (6-10) นี้ให้ค่า r เท่ากันกับการใช้สมการที่เคยกล่าวมาแล้ว คือสมการ (6-2) หรือ (6-3) ทุกประการ ดังนั้นเราอาจนำลำดับที่ให้ไว้นี้ไปหา r_s โดยใช้สมการนั้น ๆ ก็ได้ แต่สมการใหม่นี้สะดวกและรวดเร็วกว่ามาก

ดรรชนีสหสัมพันธ์นี้มีค่าระหว่าง – 1 และ +1 คือมีค่า –1 เมื่อมีการเพิ่มลดที่สวนทางกันอย่างสมบูรณ์ และมีค่า + 1 เมื่อเพิ่มไปด้วยกันทุก ๆ ขั้น ซึ่งเราอาจตรวจสอบได้ดังนี้

X	Y	d	d^2	X	Y	d	d^2
1	1	0	0	1	3	–2	4
2	2	0	0	2	2	0	0
3	3	0	0	3	1	2	4
		$r_s = 1$				$r_s = -1$	

ซึ่งจะเห็นได้ว่าเป็นความจริง เราอาจตรวจสอบความจริงนี้กับข้อมูลกี่คู่ก็ได้

การทดสอบสมมุติฐาน $H_0 : \rho = 0$ เราเพียงทดสอบความจริงว่าเมื่อ $\rho = 0$ การกระจายของ r_s มีค่าเฉลี่ยเป็น 0 และวาเรียนซ์ $\sigma_r^2 = 1^2/(n-1)$ ดังนั้นการทดสอบค่า r_s อาจกระทำได้โดยใช้สถิติ

$$Z = \frac{r_s - \rho}{\sigma_r} \qquad \ldots(6\text{-}11)$$

เพื่อทดสอบ $H_0 : \rho_s = 0, H_1 : \rho_s \neq 0$ เราปฏิเสธสมมุติฐาน ถ้าค่า Z ที่คำนวณได้ไม่อยู่ระหว่าง $-Z_{\alpha/2}$ และ $+Z_{\alpha/2}$

6.9 แบบฝึกหัด

1. จงอธิบายถึงความแตกต่างระหว่างรีเกรซชันและสหสัมพันธ์

2. จากข้อมูลในตัวอย่างที่ 1 (ตอนที่ 6.2) จงหาสมการรีเกรซชันของบุตรชาย (X) ต่อบุตรหญิง (Y) และ บุตรหญิง (Y) ต่อบุตรชาย (X) แล้วลากเส้นรีเกรซชันนั้น ๆ

3. จงพิสูจน์ให้เห็นว่า

ก. $SSE = TSS\ (1-r^2)$

เมื่อ $SSE = \sum (Y_i - Y_c)^2$, $TSS = \sum (Y_i - \overline{Y})^2$

ข. $t = b/s_b = r\sqrt{n-2}\,/\sqrt{1-r^2}$

ค. $b = r(s_y / s_x)$

ทั้งนี้เมื่อ s_y และ s_x คือส่วนเบี่ยงเบนมาตรฐานของ Y และ X ตามลำดับ

4. จากข้อมูลต่อไปนี้จงคำนวณหาดรรชนีสหสัมพันธ์ r

X	1	3	4	6	8	9	11	14
Y	1	2	4	4	5	7	8	9

5. ในการสอบวิชาภาษาอังกฤษและคณิตศาสตร์ของนักศึกษากลุ่มหนึ่ง เมื่อสุ่มมา 6 คน ได้ผลดังนี้

ภาษาอังกฤษ	70	92	80	74	65	83
คณิตศาสตร์	74	84	63	87	78	90

จงคำนวณหาดรรชนีสหสัมพันธ์โดยใช้คะแนนดิบ และเมื่อหักลบคะแนนภาษาอังกฤษและคณิตศาสตร์ ด้วย 65 และ 60 ตามลำดับ

6. สมมุติว่าสุ่มหาพี่น้องมา 10 คู่ แล้วสำรวจไอคิว เมื่อคำนวณหาดรรชนีสหสัมพันธ์ปรากฏว่าได้ $r = 0.6$ จงทดสอบสมมุติฐานที่ระดับความแตกต่าง 0.05 ว่าตัวอย่างนี้มาจากประชากรที่มี $\rho = 0$

7. จากข้อมูลดังต่อไปนี้

X	4	5	9	14	18	22	24
Y	16	22	11	16	7	3	17

ก. จงคำนวณหาดรรชนีสหสัมพันธ์ r

ข. จงทดสอบสมมุติฐานที่ระดับความแตกต่าง 0.05 ว่าตัวอย่างนี้มาจากประชากรที่มี $\rho = -0.8$ และ $\rho = 0$

8. จากการศึกษาถึงความสัมพันธ์ระหว่างจำนวนต้นต่อแปลง และผลผลิตของพืชชนิดหนึ่ง จำนวน 19 แปลง ปรากฏว่าได้ $r = 0.72$ จงทดสอบสมมุติฐานที่ระดับควมแตกต่าง 0.01 ว่าตัวอย่างนี้มาจากประชากรที่มี $\rho = 0.40$

คำในบท

(1) bivariate population , (2) correlation analysis , (3) simple correlation, (4) multiple correlation (5) rank correlation , (6) non-parametic statistics , (7) rank correlation coefficient.

บทที่ 7
การทดสอบโดยใช้ไค--สแควร์

7.1 คำนำ

ข้อมูลที่นำมาวิเคราะห์ในทางสถิตินั้นได้มา 2 วิธี คือ (1) ได้มาโดยวิธีชั่ง ตวง และวัด เช่น การให้ผลผลิตของพืช และการให้นมของโค เป็นต้น (2) ได้มาโดยการนับ เช่น เมื่อใช้ยาป้องกันเชื้อราบางชนิดแล้วพืชเป็นโรคโคนเน่ากี่ต้นและไม่เป็นกี่ต้น หรือเมื่อสำรวจเมล็ดถั่วเขียวที่ปลูกในฤดูแล้งแล้วจะให้เปอร์เซ็นต์เมล็ดแข็ง[1] เท่าใด ฯลฯ ข้อมูลประเภทหลังนี้เราอาจทดสอบว่า อัตราส่วนของสมาชิกที่สำรวจได้แตกต่างจากข้อสันนิษฐานหรือไม่ หรือต้องการทดสอบว่า อัตราส่วนของค่าสังเกตขึ้นอยู่กับปัจจัยบางชนิดหรือไม่ การทดสอบเพื่อให้ข้อสรุปถึงข้อมูลที่ได้จากการนับมักใช้สถิติไค-สแควร์ ความจริงแล้วไค-สแควร์[2] เป็นสถิติที่มีการใช้อย่างกว้างขวางมาก แต่ในบทนี้จะกล่าวถึงเฉพาะการใช้ประโยชน์ในการค้นคว้าวิจัยทางเกษตร และการใช้ประโยชน์แบบอื่น ๆ ที่พบแพร่หลายเท่านั้น

7.2 การทดสอบการสอดคล้อง[3]

เป็นที่พบเห็นเสมอว่า ผลที่ได้จากการทดลองหรือการปฏิบัติ ไม่สอดคล้องหรือไม่ตรงกับผลที่คาดหมายในทางทฤษฎี เช่น ในทางพันธุศาสตร์นั้น เมื่อผสมระหว่างพืชหรือสัตว์ที่แตกต่างกัน 2 ลักษณะ แต่ละลักษณะควบคุมโดยยีน 1 คู่ และยีนเด่นข่มยีนด้อยอย่างสมบูรณ์ อัตราส่วนของลูกผสมในชั่วที่สอง (F_2) ได้แก่ 9 : 3 : 3 : 1 เช่น มีลูกชั่วนี้ 480 ต้น ถ้ายึดถือตามทฤษฎีก็จะได้อัตราส่วนของค่าคาดหมาย[4] เป็น 270 : 90 : 90 : 30 แต่อัตราส่วนของค่าสังเกต[5] อาจเป็น 290 : 80 : 85 : 25 ก็ได้ ความคลาดเคลื่อน หรือความแตกต่างของค่าสังเกตจากค่าคาดหมายชุดนี้สูงพอที่จะทำให้เราตัดสินว่า อัตราส่วนของค่าสังเกตไม่กระจายตามอัตราส่วน 9 : 3 : 3 : 1 ได้หรือไม่ หรือความคลาดเคลื่อนมีเพียงเล็กน้อย ซึ่งยังจัดได้ว่ายังมีการกระจายแบบ 9 : 3 : 3 : 1 การตัดสินเช่นนี้เราใช้วิธีการทดสอบที่เรียกว่า การทดสอบการสอดคล้องของการกระจายของข้อมูลที่มีค่าสังเกตและค่าคาดหมาย โดยใช้ไค-สแควร์ ซึ่งมีสมการดังนี้

$$\chi^2 = \frac{(o_1 - e_1)^2}{e_1} + \ldots + \frac{(o_n - e_n)^2}{e_n}$$

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i} \qquad \ldots(7\text{-}1)$$

เมื่อ $i = 1, 2, ..., n$; o = ค่าสังเกตได้ และ e = ค่าคาดหมาย ในการทดสอบเปิดตารางไค-สแควร์ที่ $df = n - 1$ ทั้งนี้ n คือจำนวนชั้นของการกระจายของข้อมูล

ตัวอย่าง 7.2.1

ในการผสมระหว่างถั่วเขียวที่มีลำต้นสีม่วงซึ่งเป็นลักษณะข่ม กับต้นสีเขียว ซึ่งเป็นลักษณะด้อย ลักษณะสีดังกล่าวนี้ควบคุมโดยยีน 1 คู่ เมื่อสังเกตในชั่ว F_2 จำนวน 400 ต้น พบว่าให้ต้นสีม่วง 315 ต้น และสีเขียว 85 ต้น จงทดสอบที่ระดับความแตกต่าง 0.05 ว่าอัตราดังกล่าวนี้ไม่แตกต่างจากอัตราส่วนคาดหมาย (3 : 1)

วิธีทำ

(1) **สมมุติฐาน : H_0 : อัตราส่วนเป็น 3 : 1**

H_1 : อัตราส่วนไม่ใช่ 3 : 1

(2) **กฎตัดสิน : ปฏิเสธ H_0 ที่ระดับความแตกต่าง 0.05 ถ้าค่า χ^2 ที่คำนวณได้มีค่าสูงกว่า $\chi^2_{0.05}$ df1 = 3.84 ทั้งนี้**

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i}$$

(3) **คำนวณ : จากพืช 400 ต้น ถ้าเป็นอัตราส่วน 3 : 1 ก็จะได้ต้นสีม่วง 3/4 (400) = 300 ต้น และสีเขียว 1/4 (400) = 100 ต้น ดังนั้น**

$$\chi^2 = \frac{(315-300)^2}{300} + \frac{(85-100)^2}{100} = 0.75 + 2.25$$
$$= 3.00$$

(4) $\chi^2 = 3.00 < \chi^2_{0.05}$ df1 = 3.84 **จึงยอมรับ** H_0

จงสังเกตุในตัวอย่างนี้ว่า e_i ที่เป็นตัวหารจะเปลี่ยนแปลงไปตาม e_i ที่เป็นตัวตั้งและการทดสอบนี้เป็นการทดสอบแบบหางเดียวเสมอ

ความจริงแล้วข้อมูลที่มีการกระจายแบบไม่ต่อเนื่องดังตัวอย่างข้างบนนี้ เราควรแก้ไขให้มีการกระจายแบบต่อเนื่องเสียก่อน โดยนำเอา 0.5 ไปลบออก ก่อนลบออกถ้า $o_i - e_i$ เป็นค่าติดลบก็ทำให้เป็นผลบวกเสียก่อน คือสถิติทดสอบจะกลายเป็น

$$\chi^2 = \sum \frac{(|o_i - e_i| - 0.5)^2}{e_i} \qquad \ldots(7\text{-}2)$$

เมื่อแก้ไขแล้ว ค่าไค-สแควร์ลดลงกว่าเดิม อย่างไรก็ดีการแก้ไขให้เป็นการกระจายแบบต่อเนื่องจำเป็นเมื่อ df = 1 เท่านั้น

ตัวอย่าง 7.2.2

ในการโยนเหรียญ 3 อัน 200 ครั้ง ปรากฏว่าได้ด้านหัวครั้งละ 0, 1, 2 และ 3 เหรียญ เท่ากับ 18, 63, 84 และ 35 ครั้ง ตามลำดับ จงทดสอบที่ระดับความแตกต่าง 0.05 ว่า อัตราส่วนที่สังเกตและคาดหมายสอดคล้องกัน

วิธีทำ **ตัวอย่างนี้จะไม่แสดงขั้นตอนการคำนวณให้ละเอียด เพียงแต่สรุปวิธีการดังนี้**

จำนวนด้านหัว	0	1	2	3	รวม
ค่าสังเกต	18	63	84	35	200
ค่าคาดหมาย	25	75	75	25	200

$$\chi^2 = \frac{(18-25)^2}{25} + \frac{(63-75)^2}{75} + \frac{(84-75)^2}{75} + \frac{(35-25)^2}{25} = 8.96$$

$$\chi^2 = 8.96 > \chi^2_{0.05}\ \text{df}3 = 7.815$$

แสดงว่าค่าสังเกตและค่าคาดหมายไม่สอดคล้องกัน สรุปได้ว่าเหรียญมีน้ำหนักสองด้านไม่เท่ากัน

ในบางครั้งเรามีค่าสังเกตจากการสำรวจ แต่ไม่ทราบการกระจายในทางทฤษฎี เช่นนี้เราต้องประมาณค่าประชากรหรือพารามิเตอร์ที่จำเป็น แล้วคำนวณค่าคาดหมายต่อไป ถ้าเราต้องประมาณการพารามิเตอร์ m ค่า df ที่จะใช้ในการเปิดตารางไค-สแควร์ ได้แก่ n - m - 1

ตัวอย่าง 7.2.3

ในการโยนเหรียญ 5 อัน 1,000 ครั้ง สมมุติว่าการกระจายของจำนวนด้านหัวดังนี้

จำนวนด้านหัว	0	1	2	3	4	5
จำนวนครั้ง	38	144	342	287	164	25

สมมุติว่าเหรียญมีด้านหัวก้อยหนักไม่เท่ากัน จงคำนวณหาค่าคาดหมาย และทดสอบการสอดคล้องที่ระดับความแตกต่าง 0.05

วิธีทำ **ในการคำนวณหาค่าคาดหมายนั้นเราต้องทราบ p (โอกาสที่เหรียญขึ้นด้านหัว) เมื่อเราไม่ทราบเราก็ประมาณจากสมการ $\mu = Np$ แต่ก่อนอื่นเราต้องทราบ μ เสียก่อนคือ**

$$\mu = \sum fx / \sum f = [(38)(0) + (144)(1) + ... + (25)(5)]/1{,}000 = 2.47$$

ดังนั้น $p = \mu / N = 2.47/5 = 0.494$ **ต่อจากนั้นก็คำนวณหาการกระจายของไบโนเมียลโดยใช้สมการ (2-3) และได้ผลดังตารางข้างล่าง เมื่อจะหาค่าคาดหมายก็คูณด้วย 1,000**

จำนวนหัว (k)	P(k)	o_i	e_i	$(o_i - e_i)^2/e_i$
0	0.0332	38	33.2	0.694
1	0.1619	144	161.9	1.979
2	0.3162	342	316.2	2.105
3	0.3087	287	308.7	1.525
4	0.1507	164	150.7	1.174
5	0.0294	25	29.4	0.659
รวม	1.0000	1,000	1,000.1	8.136

จากตารางได้ค่า $\chi^2 = 8.136$ **ในตัวอย่างนี้เราคำนวณพารามิเตอร์ 1 ค่า คือ** p **ดังนั้น จึงเปิดตารางไค-สแควร์ที่** $df = 6 - 1 - 1 = 4$ **ที่ระดับความแตกต่าง 0.05 ซึ่งได้** $\chi^2_{0.05}\, df\, 4 = 9.488$ **จึงสรุปได้ว่าค่าสังเกตสอดคล้องกับค่าคาดหมาย**

ในบางครั้งค่าสังเกตและค่าคาดหมายต่ำมาก ในกรณีเช่นนี้มีการแนะนำว่า ควรจะรวมชั้นที่มีค่าต่ำ ๆ เข้าด้วยกัน โดยลดจำนวนชั้นลงไปด้วย ถือว่าถ้าค่าคาดหมายมีค่าอย่างน้อยเท่ากับ 1 แล้ว นับว่าใช้ได้

7.3 การทดสอบความเป็นอิสระของตัวอย่าง

ข้อมูลบางชนิดมีการจัดระเบียบทั้งในแบบแถวและสดมภ์ เช่น ในการสำรวจจำนวนต้นพืชที่เป็นโรค ซึ่งรอดตายโดยการฉีดสารป้องกันเชื้อราและไม่ฉีดอย่างละ 200 ต้น ดังนี้

	รอดตาย	ตาย	รวม
ฉีดยา	100	100	200
ไม่ฉีดยา	60	140	200
รวม	160	240	400

ตารางที่มีการจัดข้อมูลเช่นนี้เรียกว่าตารางคอนทินเจนซี ซึ่งอาจมีจำนวนแถว (r) และสดมภ์ (c) มากกว่านี้ก็ได้

ในตารางชนิดนี้ ข้อมูลในแต่ละช่องจะมีความสัมพันธ์กับอิทธิพลของแถวและสดมภ์ เช่นจากตัวอย่างข้างบนเห็นว่า การที่มีพืชรอดตายหรือไม่ เกิดจากการฉีดและไม่ฉีดสารเคมี ข้อมูลในแต่ละช่องเป็นเพียงค่าสังเกต (o_i) ส่วนค่าคาดหมาย (e_i) ก็คำนวณได้โดยใช้ความสัมพันธ์ระหว่างแถวและสดมภ์นั่นเอง วิธีการหาสมการเพื่อคำนวณค่าคาดหมายมีดังนี้ สมมุติว่าเรามีข้อมูลซึ่งจัดได้ r แถวและ c สดมภ์ และมีค่าสังเกต o_{11} , o_{12} , ..., o_{rc} ดังตารางข้างล่าง

ทั้งนี้ให้ C, R และ G เป็นผลรวมของสดมภ์ แถว และผลรวมทั้งหมดตามลำดับ ในการคำนวณค่าคาดหมายเพื่อบรรจุลงไปในแต่ละช่องนั้น ถือหลักของโอกาสหรือความน่าจะเป็นของผลบวกของแถวและของสดมภ์นั่นเองที่จะเป็นตัวชี้ให้เห็นถึงขนาดของค่าคาดหมาย เช่น ในสดมภ์ที่ 1 และแถวที่ i มีค่าคาดหมาย

$$e_{i1} = R_i / G$$

และในแถวที่ 1 สดมภ์ที่ j มีค่าคาดหมาย

$$e_{1j} = C_j / G$$

ทั้งนี้ $i = 1, 2, \ldots, r; j = 1, 2, \ldots, c$ เมื่อ r และ c เป็นจำนวนแถวและสดมภ์ตามลำดับ

แถวที่	สดมภ์ที่ (j)			รวม
(i)	**1**	**2**	**....c**	
1	o_{11}	o_{12}	....o_{1c}	R_1
2	o_{21}	o_{22}	....o_{2c}	R_2
.	.	.	.	.
r	o_{r1}	o_{r2}	....o_{rc}	R_r
รวม	C_1	C_2	C_c	G

ถ้าค่าคาดหมายนี้เป็นอิสระต่อกัน ก็คำนวณได้ว่าโอกาสของ e_{ij} เท่ากับ $(R_i/G) \times (C_j/G)$ ซึ่งเป็นอัตราส่วนของ 1 ถ้าให้เป็นอัตราส่วนของค่าสังเกตทั้งหมดก็คูณด้วย G ดังนั้น

$$e_{ij} = \frac{(R_i)(C_j)}{G} \qquad \ldots(7\text{-}3)$$

ซึ่งใช้เป็นสมการสำหรับคำนวณค่าคาดหมายในช่องที่ i และ j ทุกช่อง ต่อจากนั้นก็ทำการทดสอบสมมุติฐานโดยใช้สมการ (7-1) และเปรียบเทียบกับค่าไค-สแควร์ที่ $df = (r-1)(c-1)$

ตัวอย่าง 7.3.1
ได้มีการแบ่งไร่นาในเมือง Audubon รัฐไอโอวา ออกเป็น 3 กลุ่ม คือ กลุ่มที่เป็นเจ้าของเอง กลุ่มเช่า และกลุ่มที่เป็นเจ้าของและเช่าด้วย แต่ละกลุ่มมีดิน 3 ระดับ จงทดสอบว่าไร่นาทั้ง 3 กลุ่มนี้มีดินแตกต่างกันหรือไม่

ชนิดของดิน	เป็นเจ้าของ	เช่า	เป็นเจ้าของและเช่า	รวม
ดินเลว	36	67	49	152
ดินปานกลาง	31	60	49	140
ดินดี	58	87	80	225
รวม	125	214	178	517

วิธีทำ **ปัญหาเช่นนี้เป็นการทดสอบให้เห็นว่าแถวและสดมภ์มีอิสระต่อกันหรือไม่ เช่น ในกรณีนี้ก็ทดสอบว่าชนิดของดินขึ้นอยู่กับชนิดของไร่นาหรือไม่ ดังนั้นจะตั้งเป็นสมมุติฐานก็ได้ ดังนี้ H_0: แถวและสดมภ์เป็นอิสระต่อกัน, H_1 : H_0 ไม่เป็นความจริง ต่อจากนั้นก็คำนวณค่าคาดหมายแต่ละค่าโดยใช้ สมการ (7-3) เช่น**

$$e_{11} = \frac{(R_1)(C_1)}{G} = \frac{(152)(125)}{517}$$
$$= 36.75$$

จงสังเกตว่า e_{11} นี้คำนวณจากผลรวมของแถวที่ 1 และสดมภ์ที่ 1 เราสามารถคำนวณค่าคาดหมายอื่น ๆ โดยวิธีคล้ายคลึงกัน แล้วนำผลลงตารางดังนี้

ชนิดของดิน	เป็นเจ้าของ		เช่า		เป็นเจ้าของและเช่า	
	o_i	e_i	o_i	e_i	o_i	e_i
ดินเลว	36	36.75	67	62.92	49	52.33
ดินปานกลาง	31	33.85	60	57.95	49	48.20
ดินดี	58	54.40	87	93.13	80	77.47

แล้วทดสอบโดยใช้ไค-สแควร์ โดยใช้สมการดัดแปลงจากสมการ (7-1) คือ

$$\chi^2 = \sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

ดังนั้น

$$\chi^2 = \frac{(36-36.75)^2}{36.75} + \ldots + \frac{(80-77.47)^2}{77.47} = 1.54$$

$$df = (r-1)(c-1) = (3-1)(3-1) = 4$$

$$\chi^2 = 1.54 < \chi^2_{0.05} \quad df\,4 = 9.49$$

จึงยอมรับ H_0 คือสรุปว่าชนิดของดินและชนิดของไร่นาเป็นอิสระต่อกัน

ในกรณีของตารางคอนตินเยนซีขนาด 2 x 2 นั้นหาได้ว่า $df = (2-1)(2-1) = 1$ จึงจำเป็นต้องปรับให้เป็นค่าต่อเนื่องโดยนำ 0.5 ไปลบออก ดังนั้นสมการในการหาไค-สแควร์ได้แก่

$$\chi^2 = \sum \frac{(|o_{ij} - e_{ij}| - 0.5)^2}{e_{ij}} \qquad \ldots(7\text{-}4)$$

การคำนวณค่าไค-สแควร์ของตาราง 2 x 2 นั้นอาจใช้วิธีลัด เช่น จากตารางซึ่งมีค่าสังเกต a_1, a_2, b_1 และ b_2 ดังนี้

	1	2	รวม
A	a_1	a_2	N_A
B	b_1	b_2	N_B
รวม	N_1	N_2	N

เมื่อ N_1, N_2, N_A, N_B คือผลรวมของสดมภ์และแถว และ N เป็นผลรวมทั้งหมด ซึ่งสมการใช้หาไค-สแควร์แสดงได้ดังนี้

$$\chi^2 = \frac{N(a_1b_2 - a_2b_1)^2}{N_1N_2N_AN_B} \qquad ...(7\text{-}5)$$

เนื่องจากไค-สแควร์นี้มี df = 1 จึงอาจปรับให้เป็นค่าต่อเนื่องโดยใช้สมการ

$$\chi^2 = \frac{N(|a_1b_2 - a_2b_1| - N/2)^2}{N_1N_2N_AN_B} \qquad ...(7\text{-}6)$$

ตัวอย่าง 7.3.2

ในการทดลองใช้ฮอร์โมนเร่งการแตกรากกิ่งตอนพืชชนิดหนึ่งพบว่า จากการใช้ฮอร์โมนแก่พืช 100 กิ่ง มีการแตกรากในเวลากำหนด 75 กิ่ง ส่วนกิ่งที่ไม่ใช้ฮอร์โมนนั้น จาก 100 กิ่ง มีการแตกราก 65 กิ่ง จงทดสอบในระดับความแตกต่าง 0.05 ว่าการแตกรากขึ้นอยู่กับการใช้ฮอร์โมน

วิธีทำ **สมมุติฐานสำหรับตัวอย่างนี้คือ H_0: แถวและสดมภ์เป็นอิสระต่อกัน และ $H_1 : H_0$ ไม่เป็นความจริง เมื่อนำข้อมูลในตัวอย่างมาใส่ลงในตารางก็จะได้ผลดังนี้**

	แตกราก	ไม่แตกราก	รวม
ใช้ฮอร์โมน	**75**	**25**	**100**
ไม่ใช้	**65**	**35**	**100**
รวม	**140**	**60**	**200**

เมื่อคำนวณไค-สแควร์โดยใช้สมการ (7-6) ก็จะได้

$$\chi^2 = \frac{200[(75\times35)-(25)(65)-200/2]^2}{(100)(100)(140)(60)} = 1.928$$

ซึ่ง $\chi^2 = 1.928 < \chi^2_{0.05}$ df1 = 3.841 **จึงยอมรับ H_0 คือสรุปได้ว่าอัตราการแตกรากของกิ่งตอนไม่ขึ้นอยู่กับฮอร์โมน หรือฮอร์โมนไม่ช่วยให้มีการแตกรากเพิ่มขึ้นนั่นเอง**

7.4 การทดสอบดุลภาคของวาเรียนซ์[7]

การทดสอบดุลภาคของวาเรียนซ์โดยใช้ค่า F ดังแสดงในตอน 4.3 นั้น ใช้ได้เมื่อมีวาเรียนซ์เพียง 2 ค่าเท่านั้น แต่ในบางครั้งเราจำเป็นที่จะต้องตัดสินว่าวาเรียนซ์มากกว่า 2 ค่า ต่างกันหรือไม่ต่างกันเพียงใด Bartlett (ค.ศ.1943) ได้เสนอวิธีการทดสอบโดยใช้ไค-สแควร์ดังนี้

$$\chi^2 = M/C$$

$$M = 2.3026\left[\left(\sum v_i\right)\log s_p^2 - \sum v_i \log s_i^2\right]$$

$$C = 1 + \frac{1}{3(k-1)}\left[\sum \frac{1}{v_i} - \frac{1}{\sum v_i}\right]$$

เมื่อ

v_i = degrees of freedom (df)

s_p^2 = pooled variance = pooled SS / pooled df

k = จำนวนตัวอย่างหรือจำนวนวาเรียนซ์

ตัวอย่างของการทดสอบอาจแสดงโดยใช้วาเรียนซ์ของตัวอย่างในตาราง 7.4.1 ซึ่งเป็นวาเรียนซ์ของแฟลกซ์พันธุ์พ่อแม่ และลูกผสม F_1

ตาราง 7.4.1 แสดงการทดสอบดุลภาคของวาเรียนซ์มากกว่า 2 ค่า

ประชากร	v_i	s_i^2	$v_i s_i^2$	$\log s_i^2$	$v_i \log s_i^2$	$1/v_i$
P_1	81	58.57	4,744.17	1.76768	143.18208	0.01235
P_2	44	76.84	3,380.96	1.88559	82.96596	0.02273
F_1	13	79.67	1,035.71	1.90129	24.71677	0.07692
รวม	138		9,160.84		250.86481	0.11200

$$s_p^2 = \sum v_i s_i^2 / \sum v_i = 9{,}160.84/138 = 66.38$$

$$\left(\sum v_i\right)\log s_p^2 = (138)(1.82204) = 251.44152$$

$$M = 2.3026(251.44152 - 250.86481) = 1.3279$$

$$C = 1 + \frac{1}{3(2)}[0.11200 - \frac{1}{138}] = 1.01746$$

$$\chi^2 = 1.3279/1.01746 = 1.305 \quad \text{โดยมี} \quad df = k-1 = 2$$

$$\chi^2 = 1.305 \rangle \chi^2_{(.05)(2)} = 5.99$$

ซึ่งปรากฏว่าไม่แตกต่างกันในทางสถิติ อันแสดงว่าความปรวนแปรในพ่อแม่ และลูกผสม F_1 ไม่แตกต่างกันในทางสถิติแต่อย่างใดนั่นเอง

พึงสังเกตว่าในวิธีการข้างบนนั้นเราใช้ค่าล็อก[9] ฐาน 10 ถ้าใช้ค่าล็อกฐาน e ก็ไม่จำเป็นต้องคูณด้วย 2.3026 แต่อย่างใด นอกจากนั้นการหารค่า M ด้วยค่า C ก็ไม่จำเป็นถ้าหากว่า $\chi^2 = M$ ไม่ได้แตกต่างในทางสถิติอยู่แล้ว เพราะค่า C จะสูงกว่า 1 เสมอ เมื่อนำไปหารก็มีแต่จะทำให้ค่า M ต่ำลงกว่าเดิมเท่านั้น

7.5 การทดสอบการคล้ายคลึงของข้อมูล[10]

ข้อมูลบางอย่างอาจจะได้จากการทดลองหลาย ๆ ครั้ง แต่ละครั้งมีข้อมูลจำนวนน้อย ๆ ซึ่งเป็นการยากที่จะทำให้การทดสอบถูกต้องเที่ยงตรง แต่ถ้านำข้อมูลแต่ละตัวอย่างมารวมกันเพื่อทำให้ได้ค่าสังเกตจำนวนมาก ๆ แล้ว การทดสอบก็จะให้คำตอบที่ถูกต้องดีขึ้น อย่างไรก็ดีการรวมข้อมูลของตัวอย่างต่าง ๆ เข้าด้วยกันนี้ จะกระทำได้เมื่อข้อมูลมีอัตราส่วนที่เหมือนกันหรือคล้ายกัน คือ ควรได้มีการทดสอบความคล้ายคลึงเสียก่อน

ข้อมูลในตาราง 7.5.1 แสดงอัตราส่วนของลูกในชั่ว F_2 ของถั่วลันเตา ซึ่งให้เมล็ดสีเหลืองและสีเขียว แต่ละตัวอย่างได้จากพืช F_1 แต่ละต้น ตามกฎของเมนเดล เมื่อผสมระหว่างพืชที่แตกต่างกันหนึ่งลักษณะ ก็จะได้ลูก F_2 ในอัตราส่วน 3 : 1 เสมอ เมื่อคำนวณ ไค-สแควร์ของแต่ละตัวอย่างโดยใช้สมการ (7-1) ก็จะได้ผลดังแสดงไว้ในสดมภ์ขวาสุดของตาราง

ค่าไค-สแควร์ของแต่ละตัวอย่างมี df = 1 เมื่อเปรียบเทียบกับค่าไค-สแควร์ จากตารางปรากฏว่าไม่มีไค-สแควร์ค่าใดมีนัยสำคัญในทางสถิติ จึงถือว่าลูก F_2 ของแต่ละต้นให้อัตราส่วน 3 : 1

เมื่อเรามีตัวอย่างหลายตัวอย่าง ก็น่าที่จะนำค่าสังเกตของแต่ละตัวอย่างมารวมกันแล้วทดสอบจากผลรวมเพียงครั้งเดียว การรวมกันจนได้ค่าสังเกตขนาดใหญ่นี้ จะทำให้การทดสอบของเราถูกต้องดีขึ้น ทั้งนี้เพราะถ้าตัวอย่างมีขนาดเล็กแล้ว บางครั้งเราอาจปฏิเสธ H_0 แทนที่จะยอมรับ หรืออาจยอมรับ H_0 แทนที่จะปฏิเสธก็ได้ แต่การรวมข้อมูลเข้าด้วยกันนั้นใช่ว่าจะทำได้ทันที ทั้งนี้ก่อนการรวมนั้นต้องทดสอบเสียก่อนว่า ข้อมูลนั้นมีอัตราส่วนคล้ายคลึงกันหรือไม่ วิธีการทดสอบกระทำดังนี้

1. หาผลรวมของค่าไค-สแควร์ของทุกตัวอย่าง คือหา χ^2_s หาได้ว่า $\chi^2_s = 4.308$ ในขณะเดียวกันนั้น df ก็รวมกันด้วยซึ่งหาได้ว่า df = 3 แต่เราไม่อาจใช้ค่าไค-สแควร์นี้ทดสอบสมมุติฐานใด ๆ ยกเว้นเสียแต่ว่า χ^2_h ซึ่งคำนวณได้ดังข้อ 2. ไม่มีนัยสำคัญในทางสถิติเท่านั้น

2. หาค่าไค-สแควร์จากผลรวมของทุกตัวอย่าง คือหา χ^2_t ทั้งนี้ถือว่าอัตราส่วนยังคงเป็น 3 : 1 ดังนั้น

$$\chi^2_t = \frac{(96-94)^2}{94} + \frac{(36-38)^2}{38} = 0.148 (df = 1)$$

ตาราง 7.5.1 แสดงสีเมล็ดและจำนวนเมล็ดที่พบใน F_2 จากการผสมระหว่างถั่วลันเตา พ่อ-แม่ ที่มีเมล็ดสีเหลืองและสีเขียว

ต้นที่	ค่าที่สังเกต		ค่าคาดหมาย		χ^2
	สีเหลือง	สีเขียว	สีเหลือง	สีเขียว	
1	43	17	40	20	0.675
2	26	14	30	10	2.133
3	27	5	24	8	1.500
รวม	96	36	94	38	4.308

แล้วคำนวณหาค่าไค-สแควร์ของความแตกต่าง (χ_h^2) ดังนี้

	ไค-สแควร์	df
χ_s^2 (จาก 3 ตัวอย่าง)	4.308	3
χ_t^2 (จากยอดรวมของค่าสังเกต)	0.148	1
χ_h^2 (ความแตกต่าง)	4.160	2

χ_h^2 เป็นตัววัดขนาดของความขัดแย้ง (heterogeneity) ระหว่างข้อมูลแต่ละชุด ถ้าข้อมูลแต่ละตัวอย่างมีความคลาดเคลื่อนจากอัตราส่วนของ H_0 น้อยมาก หรือมีการคลาดเคลื่อนไปในทิศทางเดียวกันคือลักษณะหนึ่งสูงกว่าค่าคาดหมายเสมอ ส่วนอีกลักษณะต่ำกว่าค่าคาดหมายเสมอ เช่นนี้ χ_h^2 จะมีค่าต่ำ

จากตัวอย่างข้างบนพบว่า $\chi_h^2 = 4.160\,(\mathrm{df}\ 2)$ ไม่มีนัยสำคัญทางสถิติ เช่นนี้ก็แสดงว่าการทดสอบสมมุติฐานของอัตราส่วนดังกล่าวนี้อาจใช้ χ_s^2 หรือ χ_t^2 ได้

พึงสังเกตว่าถึงแม้ค่าไค-สแควร์ในตัวอย่างต่าง ๆ มี df = 1 แต่เราก็ไม่จำเป็นต้องแก้ไขให้เป็นข้อมูลที่มีการกระจายแบบต่อเนื่องตามวิธีของ Yates แต่อย่างใด ซึ่งเป็นข้อยกเว้นเมื่อเราจะนำข้อมูลไปรวมกันเพื่อให้เป็นตัวอย่างที่มีขนาดใหญ่

7.6 แบบฝึกหัด

1. ในการโยนลูกเต๋า 120 ครั้ง ปรากฏว่าขึ้นหน้า 1, 2, 3, 4, 5 และ 6 จำนวน 20, 22, 17, 18, 19 และ 24 ครั้งตามลำดับ จงทดสอบสมมุติฐาน (H_0) ที่ระดับความแตกต่าง 0.05 ว่าหน้าต่าง ๆ ของลูกเต๋ามีน้ำหนักเท่ากัน

2. ในการทดลองของเมนเดล (Mendel) เกี่ยวกับพันธุกรรมของรูปร่างและสีของเมล็ดถั่วลันเตา เมื่อผสมระหว่างถั่วลันเตาเมล็ดเรียบ-สีเหลือง กับถั่วลันเตาเมล็ดย่น-สีเขียว ซึ่งต่างก็เป็นพันธุ์แท้ เมื่อสังเกตลูกผสมชั่วที่สอง (F_2) พบว่าได้ลักษณะและจำนวนดังนี้

เมล็ดเรียบ-สีเหลือง	315 เมล็ด	เมล็ดย่น-สีเหลือง	101 เมล็ด
เมล็ดเรียบ-สีเขียว	108 เมล็ด	เมล็ดย่น-สีเขียว	32 เมล็ด

ตามกฎของเมนเดลการกระจายของลูกชั่วที่สองมีอัตราส่วน 9 : 3 : 3 : 1 จงทดสอบเพื่อแสดงให้เห็นว่าอัตราส่วนจากการทดลองนี้สอดคล้องกับอัตราส่วนตามกฎของเมนเดลจริง

3. จากการสำรวจครอบครัวที่มีบุตร 4 คน จำนวน 160 ครอบครัว เมื่อแยกเป็นพวกตามจำนวนบุตรชายก็ได้ผลดังนี้

จำนวนบุตรชาย	0	1	2	3	4
จำนวนครอบครัว	12	37	55	47	9

ก. ถ้าโอกาสที่จะมีบุตรชาย-หญิงเท่ากัน จงใช้สมการหาการกระจายไบโนเมียล เพื่อหาการกระจายของจำนวนบุตรชายทั้ง 160 ครอบครัว

ข. จงตั้งสมมุติฐาน และทดสอบที่ระดับความแตกต่าง 0.05 ว่าการกระจายของจำนวนบุตรชายเป็นการกระจายแบบไบโนเมียล

4. จากการสำรวจคนที่เป็นโรคมะเร็ง 180 คน ได้ผลดังนี้ :

	ไม่สูบบุหรี่	สูบเล็กน้อย	สูบมาก
เป็นมะเร็งที่ปอด	21	36	30
เป็นมะเร็งที่อื่น ๆ	48	26	19

จงทดสอบที่ระดับความแตกต่าง 0.05 ว่าการสูบบุหรี่และตำแหน่งที่เกิดของโรคมะเร็งไม่เกี่ยวข้องกัน

5. จากการสำรวจความสามารถในการเรียนของนักศึกษา 200 คน แล้วจัดแยกเป็นหมู่ตามอาชีพของผู้ปกครอง ปรากฏว่าได้ผลดังนี้

	เรียนเลว	เรียนปานกลาง	เรียนดี
กรรมกร	14	37	32
พ่อค้า	19	42	17
ข้าราชการ	12	17	10

จงทดสอบที่ระดับความแตกต่าง 0.05 ว่าความสามารถในการเรียนไม่ขึ้นอยู่กับอาชีพของผู้ปกครอง

6. จากการฉีดยาป้องกันโรคราสนิม (rust) แก่ถั่วเหลือง 200 ต้น เมื่อเปรียบเทียบกับที่ไม่ฉีดยา 400 ต้น ได้ผลดังนี้

	เป็นราสนิม	ไม่เป็น	รวม
ฉีดยา	135	65	200
ไม่ฉีดยา	165	235	400

จงทดสอบที่ระดับความแตกต่าง 0.05 ว่าการใช้ยามีผลต่อโรคราสนิมของถั่วเหลืองโดยใช้สมการ (7-6)

7. จากการชั่งน้ำหนักของนักเรียนที่มีอายุไล่เลี่ยกันจำนวน 10 กลุ่ม ปรากฏว่าได้วาเรียนซ์ (s^2) ดังนี้

กลุ่มที่	จำนวน (n)	s^2
1	10	51
2	15	78
3	21	91
4	23	52
5	15	101
6	11	36
7	31	41
8	15	76
9	3	64
10	6	93

จงทดสอบสมมุติฐาน (H_0) ว่าวาเรียนซ์เหล่านี้มีค่าเท่ากัน

8. ในการผสมระหว่างแมลงหวี่ *(Drosophila melanogaster)* โดยใช้ลักษณะมีปีกผสมกับลักษณะปีกกุด (Vestigial wing) จำนวน 3 คู่ เมื่อสังเกตดูลูก F_2 สมมุติว่าได้ผลดังนี้

คู่ที่	ปีกยาว	ปีกกุด
1	53	27
2	31	17
3	26	14

ก. จงทดสอบสมมุติฐานว่าแต่ละตัวอย่าง (แต่ละคู่ผสม) ให้อัตราส่วน 3 : 1

ข. ข้อมูลชุดต่าง ๆ มีอัตราส่วนคล้ายคลึงหรือแตกต่างกันเพียงใด

ค. อาจใช้ χ_s^2 เพื่อการทดสอบทางสถิติได้หรือไม่

ง. จงทดสอบสมมุติฐานโดยใช้ χ_t^2

จ. ทำไมจากการทดสอบโดยวิธีในข้อ ก. และ ง. จึงให้ผลสรุปที่แตกต่างกัน

คำในบท

(1) hard seed, (2) Chi-square, (3) goodness-of-fit test (4) expected frequency, (5) observed frequency, (6) contingency table, (7) homogeneity of variance, (8) flax, (9) logarithm, (10) test of homogeneity of samples.

บทที่ 8
หลักการวางแผนการทดลอง

8.1 คำนำ

มีปรากฏการณ์ที่เราเข้าใจกันดีอยู่อย่างหนึ่งคือ เหตุการณ์หนึ่ง ๆ ที่เกิดขึ้น ไม่แน่ใจว่าจะเกิดซ้ำเช่นเดิมอีกหรือไม่ ทั้งนี้เพราะมีปัจจัยภายนอกมากมายหลายชนิดที่เข้ามาเกี่ยวข้องหรือมีอิทธิพลต่อปรากฏ-การณ์นั้น ๆ ทั้งเป็นอิทธิพลที่ควบคุมได้และควบคุมไม่ได้ ยกตัวอย่างง่าย ๆ เช่น มีผู้พบว่าถั่วเหลืองต้นหนึ่งให้เมล็ดหนักถึง 80 กรัม ถ้าเรานำถั่วเหลืองต้นนี้ไปขยายพันธุ์ แล้วส่งเสริมให้กสิกรปลูกในอัตรา 32,000 ต้นต่อไร่ ก็จะให้ผลผลิตถึง 2,560 กก./ไร่ ถ้าปรากฏการณ์เช่นนี้เป็นความจริงในทุกหนทุกแห่ง ในทุกสถานการณ์และดินฟ้าอากาศ ก็ไม่จำเป็นต้องมีการใช้วิธีการทางสถิติที่สลับซับซ้อนในการวิจัย แต่ตามความเป็นจริงเมื่อนำถั่วเหลืองจากต้นนี้ไปปลูกในแปลงใหญ่ ถึงจะใส่ปุ๋ยอย่างเต็มที่ก็ให้ผลผลิตไม่เกิน 600 กก./ไร่ อะไรคือสาเหตุที่ทำให้เกิดความแตกต่างไปจากผลผลิตที่เราประมาณเอาไว้ในตอนแรก ที่พอจะคาดคะเนได้ น่าจะเกี่ยวข้องกับความอุดมสมบูรณ์ของดิน ปริมาณน้ำ และความชื้นที่พืชได้รับ ตลอดถึงธาตุอาหาร โรคแมลง ฤดูปลูก การปฏิบัติดูแลรักษา ฯลฯ สิ่งเหล่านี้ย่อมเป็นสาเหตุที่ทำให้ผลผลิตพืชปรวนแปรไปทั้งสิ้น

มีปัจจัยภายนอกมากมาย ที่เข้ามามีผลกระทบต่อการให้ผลผลิตของพืช-สัตว์ ในสภาพการปลูก การเลี้ยงดูชนิดหนึ่ง จะให้ผลผลิตระดับหนึ่ง แต่ในอีกสภาพหนึ่งให้ผลอีกระดับหนึ่ง จนเราไม่ทราบแน่นอนว่าลักษณะแท้จริงเป็นอย่างไร แต่ถ้าเราทำการทดลองซ้ำ ๆ หลายครั้ง ผลหรือลักษณะที่ปรากฏมากครั้งน่าจะเป็นลักษณะที่แท้จริงของพืช-สัตว์ชนิดนั้น

8.2 การวิจัยคืออะไร

การวิจัยคือ การศึกษา ค้นคว้า เก็บรวบรวมข้อมูล แล้วทำการวิเคราะห์ และตีความหมายจากผลการศึกษาค้นคว้า เพื่อให้สามารถอธิบายเกี่ยวกับกระบวนการ หรือวิธีการที่เป็นจริงเกี่ยวกับปรากฏการณ์นั้น ๆ การวิจัยอาจกระทำโดยการเก็บรวบรวมข้อมูลจากที่ปรากฏในธรรมชาติ นำมาพิจารณาหรือวิเคราะห์ แล้วนำผลจากการวิเคราะห์ไปอธิบายถึงความจริงเกี่ยวกับข้อมูล แต่ในปัจจุบันนี้ นักวิจัยเป็นจำนวนมากได้ใช้วิธีการทดลองเพื่อเก็บรวบรวมข้อมูล แทนที่จะเก็บรวบรวมจากที่ปรากฏในธรรมชาติแต่เพียงอย่างเดียว โดยใช้วิธีการทางสถิติเข้าช่วย คือใช้ความรู้ในทางสถิติทั้งในด้านการทดลองและแปลความหมายของข้อมูล ตัวอย่างเช่น การวิจัยทางเกษตร : การวิจัยทางเกษตรประกอบด้วยงานวิจัยระดับ

พื้นฐาน[1] และการวิจัยประยุกต์[2] การวิจัยระดับพื้นฐานคือการค้นคว้าหาความจริงทางวิทยาศาสตร์เกี่ยวกับพืช สัตว์ ดิน ปุ๋ย โรค แมลง หรือผลสนองตอบของสิ่งเหล่านั้นต่อปัจจัยต่าง ๆ ที่ใช้ในการทดลอง ในการวิจัยเรื่องหนึ่ง ๆ อาจประกอบด้วยการวิจัยระดับพื้นฐานมากมาย เช่น การวิจัยเพื่อพัฒนาข้าวโพดพันธุ์สุวรรณ 1 มีการวิจัยพื้นฐานเกี่ยวกับพันธุศาสตร์ โรค ดิน ปุ๋ย สรีรวิทยา และอื่น ๆ ส่วนการวิจัยประยุกต์คือ การวิจัยที่มีจุดมุ่งหมายเพื่อใช้ประโยชน์ อาจเป็นการวิจัยในปัญหาต่าง ๆ ในสภาพแวดล้อมที่เป็นจริง เพื่อการพัฒนาชุดเทคโนโลยีที่อาจนำไปใช้ประโยชน์ได้โดยตรง เช่น การวิจัยเกี่ยวกับระบบการปลูกพืช เพื่อจะนำไปแนะนำแก่กสิกรเป็นต้น โดยความจริงแล้วการวิจัยพื้นฐานและการวิจัยประยุกต์มีความต่อเนื่องกัน ในบางสถานะอาจมีส่วนเชื่อมต่อกันจนแยกไม่ออก

นอกจากที่กล่าวมาแล้ว การวิจัยอาจแบ่งเป็นหมวดหมู่ได้อีกหลายแบบ คือ แบ่งตามจำนวนสาขาวิชาการที่เกี่ยวข้องเป็น การวิจัยเอก (เอ-กะ) วิชา [3] และการวิจัยสหวิชา[4] การวิจัยอาจวิจัยตามเหตุการณ์ที่เกิดขึ้นแล้ว[5] หรือวิจัยโดยการทดลอง[6] การวิจัยตามเหตุการณ์ที่เกิดขึ้นแล้ว คือการสำรวจจากข้อมูลที่ปรากฏในธรรมชาติ แล้วนำมาให้เหตุผลหรือสรุปเกี่ยวกับสาเหตุ ในการวิจัยแบบนี้เราไม่สามารถควบคุมความคลาดเคลื่อนอันเกิดจากสาเหตุภายนอกได้ ผิดกับการวิจัยโดยการทดลอง ซึ่งเราสามารถควบคุมความคลาดเคลื่อนที่มีสาเหตุต่าง ๆ จากภายนอกได้ในระดับหนึ่ง

การวิจัยทุกชนิดต้องมีการวางสมมุติฐานไว้เป็นการล่วงหน้า สมมุติฐานคือข้อสันนิษฐานเกี่ยวกับค่าของประชากร หรือพารามิเตอร์[7] คือสันนิษฐานว่า ค่าที่ได้ควรจะเป็นเท่านั้นเท่านี้ แล้วทำการทดลองเพื่อพิสูจน์ว่าสมมุติฐานนั้นเป็นจริงหรือไม่ เช่น สันนิษฐานว่าพืช 10 พันธุ์ มีผลผลิตเท่ากัน แล้วเราทำการทดลองเพื่อหาคำตอบว่า ข้อสันนิษฐานดังกล่าวนั้นเป็นความจริงหรือไม่

การทดลองเป็นขั้นตอนสำคัญในการวิจัย ความจริงการทดลองก็คือกระบวนการในการเก็บรวบรวมข้อมูลภายใต้สภาวะที่มีการควบคุมหรือป้องกันความคลาดเคลื่อน หรือป้องกันผลกระทบจากภายนอก การทดลองอาจดำเนินในแปลงทดลองหรือในห้องปฏิบัติการ การทดลองในแปลงหรืออาจเรียกว่าการทดลองในสนาม ต้องใช้เทคนิคในการควบคุมความคลาดเคลื่อน เทคนิคดังกล่าวนั้นเราเรียกว่าแผนการทดลอง[8] ซึ่งมีอยู่มากมายหลายชนิดและเราจะทำการศึกษาต่อไป การทดลองทางเกษตรมีหลายชนิด เช่น ในการทดลองเกี่ยวกับพืช ก็มีการทดลองเปรียบเทียบพันธุ์ ทดลองใช้ปุ๋ยชนิดหรือระดับต่าง ๆ ทดลองใช้อัตราปลูก ทดลองกำจัดโรคแมลง ทดลองเกี่ยวกับผลของสภาพแวดล้อมหรือปัจจัยภายนอก เช่น แสงและความชื้น ที่มีต่อการเจริญเติบโตของพืช ฯลฯ การทดลองเกี่ยวกับสัตว์ เช่น การศึกษาสภาพการเลี้ยงดูแบบต่าง ๆ และศึกษาถึงผลของอาหารต่อการเจริญเติบโตดังนี้ เป็นต้น

8.3 พื้นฐานของการวางแผนการทดลอง

แผนการทดลอง คือ เทคนิคที่ใช้ในการเก็บรวบรวมข้อมูลวิธีหนึ่ง ในการทดลองเพื่อหาความแตกต่างระหว่างค่าเฉลี่ยนั้น เราเปรียบเทียบทีละหลายค่าก็ได้ คือทำการทดลองทีละหลาย ๆ ปัญหาหรือหลายทรีตเมนต์[9] เช่น ในการเปรียบเทียบพันธุ์พืชเราอาจทดลองครั้งละหลายพันธุ์ เปรียบเทียบสูตรอาหารสัตว์ครั้งละหลายสูตร การทดลองเช่นนี้ถ้ามีอิทธิพลภายนอกเข้ามากระทบ อาจทำให้เกิดความคลาดเคลื่อนได้มาก เช่น การทดลองเกี่ยวกับพืช ก็มีความคลาดเคลื่อนที่เกิดจากอิทธิพลต่าง ๆ เช่น : เกิดจากสภาพของดินที่มีความอุดมสมบูรณ์ไม่เท่ากันในทุก ๆ จุด เกิดจากการปฏิบัติรักษาที่ไม่สม่ำเสมอ เกิดจากความชื้นในดินที่ไม่เท่ากัน ฯลฯ การทดลองเช่นนี้อาจปรับปรุงให้ถูกต้องหรือมีประสิทธิภาพดีขึ้น โดยการใช้แผนการทดลองที่เหมาะสม

หลักในการวางแผนการทดลองประกอบด้วยขั้นตอนใหญ่ ๆ 3 ขั้นตอน คือ (1) การกำหนดปัญหาที่ต้องการจะศึกษาทดลอง (2) การกล่าวถึงวัตถุประสงค์ของการทดลองและสมมุติฐานที่ต้องการทดสอบ และ (3) การตัดสินใจเลือกใช้แผนการทดลอง การตั้งปัญหาที่จะศึกษา คือการกำหนดเรื่องและขอบเขตของเรื่องที่ต้องการทดลอง พร้อมกับกำหนดลักษณะที่ต้องการเก็บข้อมูลตลอดถึงขั้นตอนของการเก็บ ในการทดลองจะต้องกล่าวถึงวัตถุประสงค์ไว้ให้ชัดเจนว่า เราต้องการที่จะแสวงหาคำตอบในเรื่องใด วัตถุประสงค์ต้องมีความเฉพาะเจาะจงสำหรับการทดลองนั้น ๆ มิใช่วางไว้กว้าง ๆ เป็นเชิงปรัชญาจนไม่สามารถแสวงหาคำตอบที่แน่นอน เพื่อให้มีวัตถุประสงค์ที่รัดกุม เราอาจจะแยกเป็นข้อ ๆ และข้อที่สำคัญในระดับสูงจะต้องนำมากล่าวไว้ก่อน

การเลือกใช้แผนการทดลอง นับว่าเป็นขั้นตอนที่สำคัญในการวางแผนการทดลอง แผนการทดลองมีมากมายหลายชนิด แต่ละชนิดมีความเหมาะสมต่อปัญหาที่จะศึกษาทดลองแตกต่างกัน การเลือกใช้แผนการทดลองยังขึ้นอยู่กับชนิดของปัญหาที่ศึกษา และจำนวนของสาเหตุของความคลาดเคลื่อนที่เข้ามามีส่วนเกี่ยวข้องกับการทดลองนั้น เช่น ในการทดลองเกี่ยวกับพืช ถ้ามีปัญหาที่จะต้องศึกษามากก็ต้องใช้แปลงทดลองขนาดใหญ่ ความคลาดเคลื่อนเนื่องจากพื้นที่ทดลองก็มีมากขึ้น จึงจำเป็นต้องใช้แผนการทดลองชนิดที่ควบคุมความคลาดเคลื่อนได้ดี แต่ถ้ามีปัญหาเป็นจำนวนน้อย การทดลองน่าจะศึกษาให้ลึกถึงปัญหาอย่างอื่นไปด้วย ทั้งนี้เพื่อเพิ่มประสิทธิภาพและคุณค่าของการทดลอง การทดลองเช่นนี้ใช้แผนการทดลองที่สลับซับซ้อนขึ้น อย่างไรก็ดี แผนการทดลองที่ดีต้องดำเนินการทดลองได้ง่ายให้คำตอบที่ถูกต้องตามวัตถุประสงค์ และมีความประหยัด

เมื่อเลือกแผนการทดลองที่เหมาะสมแล้ว ขั้นตอนที่อาจกล่าวได้ว่าเป็นส่วนสำคัญของการวางแผนการทดลอง ได้แก่ การดำเนินการทดลอง การเก็บรวบรวมข้อมูล การวิเคราะห์ผลจากข้อมูล การแปลความหมายของผลการทดลอง และการสรุป

โดยสรุปแล้ว การใช้แผนการทดลองในการวิจัยมีวัตถุประสงค์หลักอยู่ 3 ประการ คือ (1) ให้สามารถควบคุมความคลาดเคลื่อนได้ (2) สามารถประมาณความคลาดเคลื่อนได้และ (3) สามารถแปลความหมายของผลการทดลองได้ถูกต้อง

8.4 ความหมายของคำหรือเทคนิคต่าง ๆ ในแผนการทดลอง

ในการวางแผนการทดลองนั้น มีองค์ประกอบหรือเทคนิคที่เราต้องทราบดังนี้

(1) ปัญหาทดลอง ปัญหาทดลองหรืออาจเรียกว่าทรีตเมนต์ หมายถึงสิ่งที่เราต้องการนำมาทดลองเปรียบเทียบเพื่อหาความแตกต่างหรือหาคำตอบ เช่น การทดลองเกี่ยวกับการใช้ปุ๋ยอัตราต่าง ๆ กับพืชผัก ปุ๋ยอัตราต่าง ๆ จัดเป็นปัญหาทดลองหรือทรีตเมนต์ เช่น ทรีตเมนต์ที่ 1 ใช้ปุ๋ย 50 กก./ไร่, ทรีตเมนต์ที่ 2, 3, 4, ... ใช้ปุ๋ย 100, 150 , 200, ...กก./ไร่ ตามลำดับ การทดลองเกี่ยวกับพืช เช่น การเปรียบเทียบลูกผสมข้าวฟ่างชนิดต่าง ๆ 10 ชนิด ลูกผสมแต่ละชนิดคือทรีตเมนต์ การทดลองเกี่ยวกับสัตว์ เช่น การเปรียบเทียบอาหารสัตว์ 5 สูตร สูตรอาหารเหล่านี้คือ ทรีตเมนต์

การทดลองบางชนิดเราอาจใช้ทรีตเมนต์ควบคุม หรือทรีตเมนต์เปรียบเทียบ[10] ทรีตเมนต์ดังกล่าวนี้หมายถึง พันธุ์ สูตรปุ๋ย หรือวิธีการอื่น ๆ ที่ยอมรับกันมาก่อน หรือใช้กันทั่วไป หรือเป็นวิธีการที่เรากำหนดขึ้นเป็นมาตรฐานเปรียบเทียบ เช่น ในการเปรียบเทียบพันธุ์ข้าวโพดเราอาจใช้พันธุ์สุวรรณ 1 เป็นพันธุ์เปรียบเทียบ ในการทดลองถ้าพบว่ามีพันธุ์ข้าวโพดที่ให้ผลผลิตดีกว่าพันธุ์สุวรรณ 1 ก็ใช้เป็นพันธุ์ส่งเสริมให้กสิกรปลูกต่อไป การเปรียบเทียบสูตรปุ๋ยของถั่วเหลือง เราอาจใช้ปุ๋ยสูตร 3-9-6 กก./ไร่ เป็นสูตรเปรียบเทียบ การเปรียบเทียบวิธีการเพาะปลูกใหม่ ๆ ทางเกษตรก็ใช้วิธีการที่ปฏิบัติกันมาแต่เก่าก่อนเป็นวิธีเปรียบเทียบ ทั้งนี้เมื่อพันธุ์พืช สูตรปุ๋ย และวิธีการเพาะปลูกที่กล่าวแล้วดีกว่า ทรีตเมนต์เปรียบเทียบเท่านั้น เราจึงนำผลการทดลองมาใช้ประโยชน์ต่อไป

(2) หน่วยทดลอง หน่วยทดลอง[11] หมายถึงสิ่งของ วัตถุ วัสดุ พืช สัตว์ พื้นที่ ฯลฯ ที่เรานำมาใช้เป็นตัวทดลองหรือเป็นหน่วยรองรับทรีตเมนต์ อาจพูดโดยสรุปว่า หน่วยทดลองคือตัวกลางที่ทดสอบเพื่อให้ได้คำตอบในปัญหาหรือทรีตเมนต์นั้น ๆ เช่น การทดลองเปรียบเทียบสูตรอาหารของสุกร สูตรอาหารหลาย ๆ สูตรที่เราต้องการเปรียบเทียบเรียกว่าทรีตเมนต์ สุกรที่ใช้ทดสอบหรือได้รับอาหารแต่ละสูตร เราเรียกว่าหน่วยทดลองหนึ่ง ๆ ในการทดลองเปรียบเทียบสารกำจัดแมลงศัตรูพืช สารชนิดต่าง ๆ เรียกว่าทรีตเมนต์ ส่วนพืชที่ใช้ทดลองเรียกว่าหน่วยทดลอง การทดลองใช้อุณหภูมิต่าง ๆ ในการเก็บรักษาผลไม้ อุณหภูมิต่าง ๆ เป็นทรีตเมนต์ ส่วนผลไม้ คือหน่วยทดลอง

ขนาดของหน่วยทดลองขึ้นอยู่กับชนิดของการทดลอง เช่น การทดลองอาหารของสุกร สุกรแต่ละตัวได้รับอาหาร 1 สูตร เรียกว่า 1 หน่วยทดลอง ในการทดลองเกี่ยวกับสัตว์ปีกแต่ละ 1 เล้า

อาจบรรจุได้ 20 ตัว แต่ละเล้าเรียกว่า 1 หน่วยทดลอง ในการเปรียบเทียบพันธุ์พืชแปลงหนึ่ง ไม่ว่าจะใหญ่หรือเล็กก็ตาม เรียกว่า 1 หน่วยทดลอง

หน่วยทดลองจะมีขนาดใหญ่หรือเล็ก ขึ้นอยู่กับความปรวนแปรภายในหน่วยทดลองนั้น ๆ เช่น ถ้าทดลองเกี่ยวกับอาหารไก่ ถ้าในธรรมชาตินั้นไก่แต่ละตัวมักมีการสนองตอบต่ออาหารสูตรเดียวกันแตกต่างกันมาก ในหน่วยทดลองนั้นก็ต้องใช้ไก่หลาย ๆ ตัว ในทางตรงกันข้าม ถ้าไก่แต่ละตัวมีการสนองตอบต่ออาหารเหมือน ๆ กัน ก็อาจใช้ไก่เพียงน้อยตัว โดยปกติในการทดลองเกี่ยวกับพืชหรือสัตว์ การใช้หน่วยทดลองที่มีขนาดใหญ่พอควร คือมีสมาชิกในหน่วยทดลองจำนวนที่เหมาะสม ก็สามารถลดความปรวนแปรในการทดลองลงได้ ซึ่งสังเกตได้จากการลดลงของสัมประสิทธิ์ของความปรวนแปร[12] จึงนับเป็นการเพิ่มความเที่ยงตรงของการทดลองได้วิธีหนึ่ง

(3) ความคลาดเคลื่อนในการทดลอง[13] ความคลาดเคลื่อนในการทดลอง คือความแตกต่างระหว่างปัญหาทดลองหรือทรีตเมนต์เดียวกัน ที่ได้รับการปฏิบัติที่เหมือนกัน ความคลาดเคลื่อนดังกล่าวนี้ จัดเป็นปัจจัยภายนอกที่กระทบต่อผลการทดลอง ทำให้ผลการทดลองที่ปรากฏออกมาแตกต่างจากความเป็นจริง เช่น การเปรียบเทียบพันธุ์พืช 3 พันธุ์ คือ ก, ข และ ค ในการทดลองชุดหนึ่ง ปรากฏว่าพันธุ์ ข ให้ผลผลิตในระดับสูงที่สุด แต่ในอีกชุดหนึ่งพันธุ์ ข กลับให้ผลผลิตต่ำที่สุด ความแตกต่างเช่นนี้เกิดจากอิทธิพลของสภาพแวดล้อมภายนอก ในกรณีเช่นนี้อาจจะเนื่องมาจากพื้นที่ปลูกไม่ดี ไม่สม่ำเสมอในเรื่องความอุดมสมบูรณ์ของดิน ฯลฯ ในการทดลองที่สองบังเอิญเราปลูกพันธุ์ ข ในแปลงที่มีความอุดมสมบูรณ์ต่ำจึงให้ผลผลิตต่ำไปด้วย กรณีเช่นนี้ก็เป็นการยากที่จะบอกว่าพันธุ์ใดให้ผลผลิต สูงกว่ากัน เพราะกลุ่มพันธุ์เหล่านี้แสดงสมรรถนะต่างกันในแต่ละการปลูกนั่นเอง ปรากฏการณ์เช่นนี้ เราเรียกว่าความคลาดเคลื่อนในการทดลอง

ตัวอย่างข้างบน แสดงให้เห็นถึงแหล่งของความคลาดเคลื่อนในการทดลองที่เนื่องมาจากดินปลูก โดยรวม ๆ แล้วความคลาดเคลื่อนของการทดลองมีสาเหตุจากหลายแหล่งดังนี้ :-

ก. ความคลาดเคลื่อนที่เกิดจากหน่วยทดลอง หน่วยทดลอง เช่น ดินที่ใช้ปลูกพืชอาจมีความอุดมสมบูรณ์ไม่เท่ากัน มีความแตกต่างกันในเรื่องของปริมาณธาตุอาหารในดิน ปริมาณอินทรีย์วัตถุ มีความแตกต่างในเรื่องความลาดเอียง ความชื้น ฯลฯ เหล่านี้อาจทำให้พืชเจริญเติบโตไม่เหมือนกัน ในการทดลองเกี่ยวกับพืช ถ้าพืชพันธุ์เดียวกันที่ปลูกในแปลงต่าง ๆให้ผลสูงต่ำแตกต่างกันมาก ก็ย่อมทำให้มีความคลาดเคลื่อนสูงขึ้น ในการทดลองเกี่ยวกับสัตว์ก็เช่นกัน ถ้าสัตว์ที่ได้รับทรีตเมนต์เดียวกันมีความแตกต่างกันในเรื่องของพันธุ์ อายุ ขนาด เพศ ฯลฯ ก็ย่อมทำให้ผลสนองตอบต่อทรีตเมนต์แตกต่างกันไปด้วย

ข. ความคลาดเคลื่อนที่เกิดจากสิ่งทดลอง สิ่งที่เราใช้ทดลอง เช่น พันธุ์พืช พันธุ์สัตว์ กระบวนการในการผลิตต่าง ๆ ฯลฯ โดยธรรมชาติแล้วมีความคลาดเคลื่อนอยู่ในตัว เช่น ในการทดลองเปรียบเทียบพันธุ์พืช แต่ละแปลงปลูกอาจให้ต้นพืชที่มีขนาดไม่เท่ากัน หรือพันธุ์พืชอาจมี

การแข่งขันกัน เช่น พืชพันธุ์ ก เมื่อปลูกใกล้พันธุ์ ข ให้ผลผลิตสูงแต่เมื่อใกล้พันธุ์ ค กลับให้ผลผลิตต่ำ ในการทดลองเกี่ยวกับอาหารสัตว์ อาจมีความคลาดเคลื่อนที่สัตว์ตัวหนึ่งอาจชอบหรือไม่ชอบกินอาหารสูตรดังกล่าวกว่าสัตว์อื่นที่ได้รับอาหารชนิดเดียวกัน

ค. ความคลาดเคลื่อนที่เกิดจากการปฏิบัติการทดลอง การปฏิบัติการทดลองทำให้เกิดความคลาดเคลื่อนได้เช่นกัน การปฏิบัติการที่ไม่สม่ำเสมอหรือไม่เหมือนกัน อาจทำให้ผลการทดลองแตกต่างกัน เช่น สุกรพันธุ์ ก แม้เป็นสัตว์ครอกเดียวกัน เพศเดียวกัน แต่ถ้าได้รับการดูแลเอาใจใส่ต่างกัน ก็ทำให้ผลสนองตอบต่อทรีตเมนต์ต่างกัน ในการทดลองเกี่ยวกับพืช พืชพันธุ์เดียวกันที่ปลูกในสองแปลง แต่มีการดูแลแตกต่างกัน เช่น แตกต่างกันในการให้น้ำ การใส่ปุ๋ย การปราบวัชพืช ฯลฯ อาจทำให้ผลผลิตแตกต่างกัน การปฏิบัติไม่สม่ำเสมอเหล่านี้ย่อมทำให้ความคลาดเคลื่อนในการทดลองเพิ่มขึ้นทั้งสิ้น

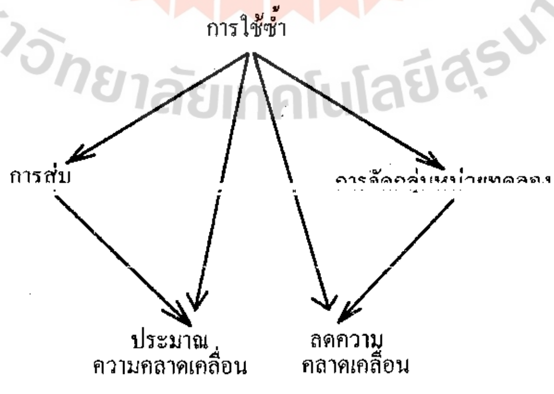
ต้นตอของความคลาดเคลื่อนอาจมาจากแหล่งอื่น ๆ อีกหลายแหล่ง ในการทดลองเราต้องพยายามควบคุมให้มีปรากฏให้น้อยที่สุด เช่น การปลูกพืช การเลี้ยงสัตว์ การเลือกแปลงทดลอง เลือกสัตว์ทดลอง การปฏิบัติดูแล ฯลฯ ให้เป็นไปโดยสม่ำเสมอ ไม่พยายามให้มีความแตกต่างกันและในการทดลอง เราพยายามใช้แผนการทดลองที่เหมาะสมเพื่อให้สามารถแยกความคลาดเคลื่อนให้เป็นสัดส่วนที่แน่นอน เพราะในกรณีที่แยกไม่ได้ ความคลาดเคลื่อนนี้ก็จะไปรวมอยู่ในแหล่งความปรวนแปรของการทดลอง ซึ่งเป็นความคลาดเคลื่อนที่ไม่ทราบสาเหตุ

(4) การซ้ำทดลอง[14] การซ้ำทดลองหรือการใช้ซ้ำ หมายถึงการทดลองหรือทดสอบทรีตเมนต์เดียวกันซ้ำหลาย ๆ ครั้ง การซ้ำทดลองเป็นการเพิ่มจำนวนค่าสังเกตในทรีตเมนต์เดียวกัน ทำให้สามารถประมาณความคลาดเคลื่อนในส่วนที่เกิดจากธรรมชาติของหน่วยทดลอง เช่น พืช 10 พันธุ์ ปลูกในแปลง ก ให้ผลผลิตเฉลี่ย 500 กก./ไร่ และ 10 พันธุ์เดียวกันนั้น ปลูกในแปลง ข ให้ผลผลิตเฉลี่ย 420 กก./ไร่ ความแตกต่างที่เกิดขึ้นนี้จัดเป็นผลของซ้ำ คือผลของดินปลูก หรือการปฏิบัติที่แตกต่างกันระหว่างซ้ำนั่นเอง การใช้จำนวนซ้ำมาก ๆ ย่อมเพิ่มความเที่ยงตรงในการทดลอง คือเราสามารถควบคุมความคลาดเคลื่อนในการทดลองได้ดีขึ้น และเมื่อมีซ้ำมากขึ้น ย่อมทำให้การหาค่าเฉลี่ยใกล้ความเป็นจริงมากขึ้นด้วย และการทำซ้ำ ๆ ทำให้เราได้ข้อมูลเกี่ยวกับทรีตเมนต์นั้น ๆ หลายค่า ซึ่งนำไปใช้ในการประมาณความคลาดเคลื่อนต่อไป

(5) การสุ่ม[15] การสุ่มจัดเป็นเทคนิคในการวางทรีตเมนต์ในหน่วยทดลอง เพื่อให้มีความเสมอภาคซึ่งกันและกัน เป็นการตัดปัญหาเรื่องความลำเอียง คือในแต่ละซ้ำนั้นให้แต่ละทรีตเมนต์ มีโอกาสตกในหน่วยทดลองใด ๆ ได้เท่ากัน เช่น ในการเปรียบเทียบสมรรถนะของเครื่องพิมพ์ดีดสองชนิด โดยใช้พนักงานพิมพ์ดีด 10 คน ให้แต่ละคนพิมพ์ทั้งสองเครื่อง ถ้าหากเมื่อเริ่มพิมพ์ทุกคนพิมพ์เครื่อง A ก่อน แล้วพิมพ์เครื่อง B ต่อ คือไม่มีการสุ่มลำดับของเครื่องที่เริ่มพิมพ์ เครื่องที่พิมพ์ครั้งแรกอาจเสียเปรียบ เพราะเมื่อเริ่มพิมพ์ทุกคนอาจเริ่มอย่างช้า ๆ เพราะมือยังรู้สึกขัดและยังไม่เคยชิน

กับข้อความที่พิมพ์ แต่ต่อไปก็จะเร็วขึ้น เครื่องที่สองจึงได้เปรียบ เพราะผู้สอบพิมพ์ได้คล่องขึ้น แถมยังอ่านข้อความที่พิมพ์ได้เร็วขึ้น แต่ถ้ามีการสุ่มลำดับของเครื่อง ปัญหาเช่นนี้ก็จะหมดไป ตัวอย่างนี้แสดงให้เห็นข้อดีของการสุ่มที่ชัดเจน อย่างไรก็ดีในการทดลองแต่ละครั้งควรทำการสุ่มใหม่ทุกครั้งไป ไม่ควรใช้ผลการสุ่มทดลองหนึ่งไปใช้กับอีกการทดลองหนึ่ง การสุ่มเป็นวิธีหรือขั้นตอนที่ทำให้สามารถประมาณความคลาดเคลื่อนได้อย่างมีประสิทธิภาพ

(6) การจัดกลุ่มหน่วยทดลอง กลุ่มของหน่วยทดลองที่รองรับทุกทรีตเมนต์ในซ้ำหนึ่ง ๆ เราเรียกว่าบลอก[16] ในการจัดบลอกพยายามให้หน่วยทดลองในบลอกหนึ่ง ๆ มีความสม่ำเสมอกันให้มากที่สุด หน่วยทดลองที่แตกต่างกันก็ให้อยู่ในอีกบลอกหนึ่ง เช่น ในการเปรียบเทียบสูตรอาหารสัตว์ 3 สูตร โดยใช้สุกรที่มีอายุต่าง ๆ กัน เราจัดสุกรที่มีอายุใกล้เคียงกันไว้ในบลอกเดียวกัน เช่น สุกรมี 4 ช่วงอายุ ก็จัดได้ 4 บลอก แล้วแต่ละบลอกได้รับอาหารทุกสูตร จึงเรียกว่าเป็นการทดลองหนึ่งซ้ำ ในการทดลองเปรียบเทียบพันธุ์พืช พบเสมอว่าพื้นที่ทดลองมีความไม่สม่ำเสมอในเรื่องของความอุดมสมบูรณ์ของดิน ความลาดเอียง ปัญหาเช่นนี้แก้ไขโดยการจัดเป็นบลอก ให้พื้นที่ในบลอกมีความสม่ำเสมอกัน ถึงจะแตกต่างระหว่างบลอกก็ไม่เป็นไร เพราะในการวิเคราะห์เรามีวิธีการที่จะแยกออกมาได้ ความแตกต่างระหว่างบลอกจะไม่เข้าไปกระทบความแตกต่างระหว่างพันธุ์พืช ยิ่งจัดให้สม่ำเสมอภายในบลอกเท่าใด และแตกต่างระหว่างบลอกเท่าใด เราได้รับความเที่ยงตรงในการทดลองเท่านั้น การจัดบลอกซึ่งทำให้สามารถเพิ่มประสิทธิภาพของการทดลองเช่นนี้ เรียกว่า การควบคุมความคลาดเคลื่อนในกลุ่ม[17] คือความคลาดเคลื่อนแบบนี้จะได้รับการควบคุมมิให้เข้าไปกระทบกระเทือนต่อผลของทรีตเมนต์ แต่กลับถูกแยกออกมาให้เป็นผลของบลอก



รูป 8.4.1 รูปแสดงความสัมพันธ์ระหว่างเทคนิคต่าง ๆ ที่ใช้ในการวางแผนการทดลอง

รูป 8.4.1 แสดงความสัมพันธ์ของเทคนิคต่าง ๆ ที่ใช้ในการวางแผนการทดลอง ซึ่งจะเห็นได้ว่า ไม่ว่าการใช้ซ้ำ การสุ่ม และการจัดกลุ่มของหน่วยทดลองออกเป็นแปลงย่อย ๆ ที่เรียกว่าเป็นการจัดบลอกนั้น ต่างก็มีวัตถุประสงค์คล้าย ๆ กัน คือการประมาณความคลาดเคลื่อนและการลดความคลาดเคลื่อนในการทดลอง

8.5 การควบคุมความคลาดเคลื่อนในการทดลอง

ในการทดลองเปรียบเทียบทรีตเมนต์ ถ้าทรีตเมนต์เดียวกันให้ผลแตกต่างกันก็แสดงว่ามีความคลาดเคลื่อนเกิดขึ้นในการทดลอง ทั้งนี้ความคลาดเคลื่อนย่อมเกิดขึ้นเสมอในการทดลองด้วยสาเหตุต่าง ๆ แต่ถ้าเกิดขึ้นน้อยเท่าไร หรือเราสามารถควบคุมได้เท่าใด ก็ทำให้สามารถจับความแตกต่างระหว่างทรีตเมนต์ได้ดีเท่านั้น วิธีการควบคุมความคลาดเคลื่อนมีได้หลายวิธี เช่น

(1) การใช้บลอก ดังที่กล่าวมาแล้วข้างต้น หลักของการจัดบลอกคือ จัดให้หน่วยทดลองที่เหมือนกันหรือคล้ายคลึงกันอยู่ในบลอกเดียวกัน เมื่อวิเคราะห์ผลการทดลองทำให้สามารถแยกความแตกต่างบางส่วนให้เป็นเรื่องของบลอก มิฉะนั้นความแตกต่างเหล่านี้จะไปรวมอยู่ในความคลาดเคลื่อนของการทดลอง ดังนั้นถ้าหน่วยทดลองมีความแตกต่างกันมากการจัดบลอกย่อมจำเป็นเสมอ ในทางตรงกันข้ามถ้าหน่วยทดลองมีความสม่ำเสมอหรือเหมือนกันมาก การจัดบลอกก็ไม่มีความจำเป็น

(2) การใช้ซ้ำ ในการทดลองนั้น การทำซ้ำ ๆ หลายครั้งเป็นการเพิ่มความถูกต้องและเที่ยงตรงในการทดลอง ไม่ว่าความคลาดเคลื่อนจากแหล่งใดก็ตาม ถ้ามีจำนวนซ้ำมากก็จะทำให้ความคลาดเคลื่อนนั้นลดน้อยลงไป เช่น ค่าสังเกต 3 ค่า คือ 65, 52, 70 มีวาเรียนซ์เท่ากับ 86.33 ถ้าสมมุติว่าผลเฉลี่ยของทรีตเมนต์จริง ๆ มีค่าใกล้เคียง 65 หากเพิ่มค่าสังเกตเป็น 6 ค่า คือ 63, 65, 67, 52, 70, 68 ก็ได้วาเรียนซ์เป็น 41.36 การที่มีวาเรียนซ์ต่ำลงเช่นนี้เพราะในการสุ่มข้อมูลเรามักได้ค่าที่ใกล้เคียงค่าเฉลี่ย เพราะในหลักวิชาสถิติที่เรียนมาแล้วแสดงให้เห็นว่า ข้อมูลที่มีค่าใกล้เคียงค่าเฉลี่ยจะมีจำนวนมากที่สุด เมื่อมีจำนวนมากโอกาสที่สุ่มได้มาก็มากไปด้วย จึงเห็นได้ว่ายิ่งมีค่าสังเกตมากขึ้นเท่าใด ยิ่งทำให้วาเรียนซ์ของค่าเฉลี่ยลดลงเท่านั้น เมื่อวาเรียนซ์ของค่าเฉลี่ยลดลง ก็สามารถจับความแตกต่างระหว่างค่าเฉลี่ยได้ดีขึ้น เพราะการจับความแตกต่างนั้น เราใช้วาเรียนซ์เป็นมาตรฐานรองรับหรือเป็นตัวหาร ดังกรณีของการทดสอบโดยใช้ค่า t หรือค่า F

จำนวนซ้ำที่ใช้ในการทดลองขึ้นอยู่กับขนาดของการทดลอง และธรรมชาติของหน่วยทดลอง ในการทดลองเกี่ยวกับพืช-สัตว์ ถ้าใช้มากซ้ำก็ย่อมเพิ่มทุนในการทดลอง โดยเฉพาะอย่างยิ่งการทดลองที่มีทรีตเมนต์มาก ๆ และหน่วยทดลองมีขนาดใหญ่ ในการทดลองเกี่ยวกับพืชซึ่งปลูกในแปลงทดลอง

นิยมใช้ 4 ซ้ำกันมาก พบว่าการใช้เกิน 4 ซ้ำ จะลดความคลาดเคลื่อนลงอย่างช้า ๆ จนไม่คุ้มกับการลงทุน แต่การทดลองแปลงเล็ก หรือทดลองในกระถาง ควรใช้ซ้ำ มาก ๆ เช่น 5 - 10 ซ้ำ โดยสรุปแล้ว การใช้ซ้ำมีประโยชน์ 2 อย่างคือ (1) ใช้ประมาณความคลาดเคลื่อนของการทดลอง และ (2) ใช้ลดความคลาดเคลื่อนของการทดลอง

(3) การระมัดระวังในการทดลอง ในการทดลอง ปัญหาหรือปัจจัยใดก็ตามที่ไม่ใช่เป็นเรื่องของการทดลอง ต้องทำให้สม่ำเสมอที่สุด เช่น การใส่ปุ๋ย การรดน้ำ การกำจัดวัชพืช การให้อาหารสัตว์ การขังกรง ฯลฯ ถ้าทำให้สม่ำเสมอทั้งการทดลองไม่ได้ เพราะการทดลองใหญ่เกินไป ก็พยายามให้มีความสม่ำเสมอภายในบลอกให้มาก การปฏิบัติที่ไม่สม่ำเสมอเป็นสาเหตุอันหนึ่งที่ทำให้ทรีตเมนต์เดียวกันมีความแตกต่างกันก็ได้

(4) การเพิ่มเทคนิคในการทดลอง ในการทดลองแต่ละชนิดเราอาจเพิ่มเทคนิคในการทดลอง การเก็บข้อมูล ฯลฯ เพื่อเพิ่มความถูกต้องเที่ยงตรงในการทดลอง เช่น ในการเปรียบเทียบพันธุ์พืชบางชนิด พืชอาจแข่งขันซึ่งกันและกัน ซึ่งอาจทำให้พืชบางพันธุ์ให้ผลผลิตต่ำกว่าความเป็นจริง ในการทดลองเช่นนี้เราควรใช้แถวคุม[18] แถวคุมคือแถวของพืชที่ปลูกเกินความต้องการไว้ เพื่อป้องกันไม่ให้แถวในที่ใช้ทดลองได้รับผลกระทบจากสภาพแวดล้อมภายนอก เช่น การทดลองเกี่ยวกับข้าวโพด เราอาจปลูก 4 แถว แต่วัดผลการทดลองจาก 2 แถวกลาง แถวคุมมีความจำเป็นในการเปรียบเทียบปุ๋ย พันธุ์พืช การกำจัดวัชพืช การทดลองเกี่ยวกับสารเคมีฆ่าแมลง และสารเคมีป้องกันโรคพืช เป็นต้น

เทคนิคอีกอันหนึ่งในการทดลอง คือการใช้ขนาดหน่วยทดลองที่เหมาะสม เช่น จำนวนสัตว์ที่ทดลองอาหารในแต่ละกรง ขนาดของพื้นที่ในการทดลองเกี่ยวกับพืชชนิดหนึ่ง ๆ

(5) การใช้วิธีวิเคราะห์ที่เหมาะสม ในการทดลองบางครั้งความแตกต่างของผลการทดลองมีต้นตอมาจากหน่วยทดลอง หรือสิ่งทดลองมีความแตกต่างกัน เช่น ในการทดลองอาหารสัตว์ สัตว์ที่ใช้ทดลองมีน้ำหนักไม่เท่ากัน ในการทดลองเกี่ยวกับพันธุ์พืช พืชแต่ละแปลงทดลองอาจมีจำนวนต้นไม่เท่ากัน ในการทดลองแบบนี้เราอาจสังเกตตัวแปรร่วม คือก่อนทดลองชั่งน้ำหนักสัตว์เอาไว้ หรือนับจำนวนต้นพืชแต่ละแปลงเอาไว้แล้วนำค่าสังเกตอันนี้มาวิเคราะห์ร่วมกับผลการทดลอง คือจำนวนต้นและน้ำหนักสัตว์หลังทดลอง การวิเคราะห์แบบนี้เรียกว่าการวิเคราะห์โควาเรียนซ์[19]

8.6 พื้นฐานของการวิเคราะห์วาเรียนซ์

เมื่อเราต้องการที่จะศึกษาปัญหาต่าง ๆ ทีละหลายปัญหา หรือทดสอบค่าเฉลี่ยพร้อมกันมากกว่า 2 ค่า วิธีการทางสถิติที่เรียนมาแล้วใช้ไม่ได้กับการศึกษา เช่นนี้ R.A.Fisher ได้คิดค้นวิธีการทางสถิติขึ้นมาใหม่ซึ่งเรียกว่า **แผนการทดลอง วิธีนี้ทำให้สามารถทดสอบสมมุติฐาน** $H_0 : \mu_1 = \mu_2 = ... = \mu_k$ ซึ่งเป็นวิธีทดสอบที่ใช้อัตราส่วนของวาเรียนซ์ เรียกว่า F-test นับว่าเป็นเรื่องแปลกที่เราทดสอบค่าเฉลี่ยโดยการใช้วาเรียนซ์ แต่ก็มีเหตุผลที่กระทำได้ ทั้งนี้เมื่อค่าเฉลี่ยของปัญหาที่เราศึกษาทดลองใกล้เคียงกันมากก็จะให้วาเรียนซ์ของค่าเฉลี่ย $s_{\bar{x}}^2$ ต่ำ ซึ่งจะนำไปสู่การยอมรับ H_0 ในทางตรงกันข้าม ถ้าค่าเฉลี่ยต่างกันมาก

ก็จะให้วาเรียนซ์ค่าเฉลี่ยสูง ซึ่งเป็นเหตุผลที่ทำให้ปฏิเสธสมมุติฐาน ตัวอย่างเช่น ทดลองเปรียบเทียบพันธุ์พืช 4 พันธุ์ แต่ละพันธุ์ปลูก 4 แปลง ดังข้อมูลในตาราง 8.6.1 ซึ่งเราอาจหาวาเรียนซ์ได้ดังนี้

ตาราง 8.6.1 ผลผลิตของพืช 4 พันธุ์ เปรียบเทียบกันโดยการปลูกพันธุ์ละ 4 แปลง (4 ซ้ำ)

พันธุ์พืช	ผลผลิต (กก./แปลง)				$\overline{X}$	วาเรียนซ์
A	3	2	3	4	3	$s_1^2 = 0.67$
B	5	3	5	3	4	$s_2^2 = 1.33$
C	2	3	2	1	2	$s_3^2 = 0.67$
D	6	6	4	4	5	$s_4^2 = 1.33$
					$s_{\bar{x}}^2 = 1.67$	เฉลี่ย = 1.00

จากตัวอย่างในตาราง 8.6.1 เราหาวาเรียนซ์ได้ 2 ค่า ซึ่งต่างก็สามารถใช้ประมาณวาเรียนซ์ของประชากร (σ^2) ดังนี้

(1) $s_{\bar{x}}^2 = 1.67$ ซึ่งคำนวณโดยใช้สมการ (1-5) โดยใช้ค่าเฉลี่ย 3, 4, 2 และ 5 วาเรียนซ์ดังกล่าวนี้จะสูงขึ้นตามความแตกต่างระหว่างค่าเฉลี่ย โดยความจริงแล้ว $s_{\bar{x}}^2 = s^2 / n$ ดังนั้นในตัวอย่างนี้ ก็หาได้ว่า $s^2 = 4 \times 1.67 = 6.68$ ซึ่งถือว่าเป็นความจริงได้ว่า s^2 นี้สูงหรือต่ำตามความแตกต่างระหว่างทรีตเมนต์บวกกับความคลาดเคลื่อนของการทดลอง

(2) ผลเฉลี่ยของวาเรียนซ์ คือ $(s_1^2 + ... + s_4^2)/4 = 1.00$ ซึ่งเกิดจากความคลาดเคลื่อนของการทดลองแต่เพียงอย่างเดียว

ในตัวอย่างนี้เห็นได้ว่า วาเรียนซ์ค่าแรก (6.68) มีวาเรียนซ์ค่าที่สองเป็นองค์ประกอบอยู่ด้วย การที่วาเรียนซ์ค่าแรกสูงกว่าค่าที่สอง ก็เนื่องจากผลของทรีตเมนต์นั่นเอง ดังนั้นถ้าอัตราส่วนระหว่างวาเรียนซ์สูงถึงระดับหนึ่งเราก็ปฏิเสธ H_0 ทั้งนี้อัตราส่วนหาได้จาก : สมการดังนี้

$$F = \frac{\text{ค่าประมาณ } \sigma^2 \text{ จากความแตกต่างระหว่างปัญหาทดลอง}}{\text{ค่าประมาณ } \sigma^2 \text{ จากความคลาดเคลื่อนของการทดลอง}} = \frac{6.68}{1.00} = 6.68$$

อัตราส่วนนี้เราเรียกว่า ค่า F ซึ่งเป็นพื้นฐานเบื้องต้นของการพัฒนาแผนการทดลองแบบต่าง ๆ ที่เราจะศึกษาต่อไป อนึ่งเมื่อเราได้ศึกษาในบทที่ 9 แล้ว ถ้ากลับมาวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 8.6.1 ก็จะทำให้เข้าใจคำอธิบายในตอนนี้ได้ดียิ่งขึ้น

8.7 สัญลักษณ์ของค่าสังเกตและการแยกความปรวนแปร

ข้อมูลจากการทดลอง ซึ่งกระทำโดยใช้แผนการทดลองแบบต่าง ๆ นั้น มีการจัดระเบียบทั้งแบบเป็นแถวและสดมภ์[20] และข้อมูลแต่ละค่าอาจแทนด้วยสัญลักษณ์ X ซึ่งมีเลขห้อยท้ายเพื่อบอกตำแหน่งดังนี้

	กลุ่มที่ 1	กลุ่มที่ 2	กลุ่มที่ 3	กลุ่มที่ n	เฉลี่ย
แถวที่ 1	X_{11}	X_{12}	X_{13}	-- X_{1n}	$\bar{X}_{1}.$
แถวที่ 2	X_{21}	X_{22}	X_{23}	-- X_{2n}	$\bar{X}_{2}.$
แถวที่ 3	X_{31}	X_{32}	X_{33}	-- X_{3n}	$\bar{X}_{3}.$
.	.	.	.	.	.
แถวที่ k	X_{k1}	X_{k2}	X_{k3}	-- X_{kn}	$\bar{X}_{k}.$
เฉลี่ย	$\bar{X}_{.1}$	$\bar{X}_{.2}$	$\bar{X}_{.3}$	$\bar{X}_{.n}$	$\bar{X}_{..}$

ทั้งนี้ให้ i เป็นหมายเลขแถว มีค่า 1, 2, .., k (อ่านว่า i มีค่าจาก 1 ถึง k) และ j เป็นหมายเลขสดมภ์มีค่า 1, 2, ... , n (อ่านว่า j มีค่าจาก 1 ถึง n) ส่วนผลบวกต่าง ๆ มีค่าดังนี้

- *ผลบวกของแถวที่ 1 คือ $X_{11} + X_{12} + \ldots + X_{1n} = \sum X_{1j}$
- ผลบวกของสดมภ์ที่ 1 คือ $X_{11} + X_{21} + \ldots + X_{k1} = \sum X_{i1}$
- ผลบวกของทุกค่า[21] คือ $X_{11} + X_{12} + \ldots + X_{kn} = \sum X_{ij}$
- ค่าเฉลี่ยของทุกค่า[22] คือ $\sum X_{ij}/kn = \bar{X}..$

ทั้งนี้แต่ละแถวอาจหมายถึงแต่ละทรีตเมนต์ก็ได้

ความปรวนแปร หรือความแตกต่างระหว่างค่าสังเกตต่าง ๆ บางค่าเฉลี่ยนั้นเราอาจใช้สมการง่ายๆ คือ

$$d = X_{ij} - \bar{X}..$$

ความปรวนแปรนี้มีสาเหตุมาจาก 2 แหล่ง คือ เกิดจากค่าสังเกตในแต่ละทรีตเมนต์ ($X_{ij} - \bar{X}_{i}.$) และระหว่างทรีตเมนต์ ($\bar{X}_{i}. - \bar{X}..$) ดังนั้น

$$\begin{aligned} d &= X_{ij} - \bar{X}_{i}. + \bar{X}_{i}. - \bar{X}.. \\ &= (X_{ij} - \bar{X}_{i}.) + (\bar{X}_{i}. - \bar{X}..) \end{aligned}$$

ดังนั้น

$$(X_{ij} - \bar{X}..) = (X_{ij} - \bar{X}_{i}.) + (\bar{X}_{i}. - \bar{X}..)$$

*** ผลบวกอย่างง่ายอาจเขียนให้สมบูรณ์ ดังนี้**

$$\sum X_{1j} = \sum_{j=1}^{n} X_{1j},\ \sum X_{i1} = \sum_{i=1}^{k} X_{i1},\ \sum X_{ij} = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}$$

ในหนังสือเล่มนี้จะใช้ผลบวกอย่างง่ายตลอดทั้งเล่ม

เมื่อยกกำลังสองและบวกกันทุกค่าก็จะได้

$$\underset{\text{ค่าที่ 1}}{\sum(X_{ij} - \bar{X}_{..})^2} = \underset{\text{ค่าที่ 2}}{\sum(X_{ij} - \bar{X}_{i.})^2} + \underset{\text{ค่าที่ 3}}{\sum(\bar{X}_{i.} - \bar{X}_{..})^2} \qquad ...(8\text{-}1)$$

ทั้งนี้

$$\sum(X_{ij} - \bar{X}_{i.})(\bar{X}_{i.} - \bar{X}_{..}) = 0$$

เพราะ $$\sum(X_{ij} - \bar{X}_{i.}) = 0$$

ความปรวนแปรในสมการ (8-1) มาจากแหล่งต่าง ๆ ดังนี้

(1) ค่าที่ 1 เกิดจากค่าเบี่ยงเบนของข้อมูลทุกหน่วยจากค่าเฉลี่ย จึงเรียกว่า Total Sum of Squares ซึ่งต่อไปนี้เรียกย่อ ๆ ว่า TSS การคำนวณ TSS นี้ เราถือว่า ค่าสังเกตตั้งแต่ X_{ij} ถึงค่าที่ X_{kn} มาจากประชากรเดียวกัน

(2) ค่าที่ 2 เกิดจากค่าเบี่ยงเบนระหว่างค่าต่าง ๆ ภายในกลุ่ม หรือภายในทรีตเมนต์ จากค่าเฉลี่ยของแต่ละทรีตเมนต์ ค่านี้เกิดจากความไม่เที่ยงตรงของการทดลองจึงเรียกว่า Error Sum of Squares ซึ่งต่อไปนี้จะเขียนย่อ ๆ ว่า SSE

(3) ค่าที่ 3 เกิดจากค่าเบี่ยงเบนระหว่างค่าเฉลี่ยของแต่ละทรีตเมนต์จากค่าเฉลี่ยทั้งหมด เป็นธรรมชาติที่แต่ละทรีตเมนต์ให้ค่าเฉลี่ยไม่เท่ากัน เราเรียกว่า Treatment Sum of Squares ซึ่งต่อไปนี้จะเขียนย่อ ๆ ว่า SSTr

ดังนั้นจากสมการ (8-1) เราอาจเขียนใหม่ว่า

$$\text{TSS} = \text{SSE} + \text{SSTr} \qquad ...(8\text{-}2)$$

ซึ่งเป็นการแยกความปรวนแปรทั้งหมด (TSS) ให้เป็นความปรวนแปรย่อย ๆ ตามแหล่งที่มา เมื่อหารค่าเหล่านี้ด้วย df[23] ก็จะเป็นค่าประมาณของวาเรียนซ์ (σ^2) โดยหาร TSS, SSE และ SSTr ด้วย

(1) df ของค่าสังเกตทั้งหมด ได้แก่ จำนวนค่าสังเกตทั้งหมดลบด้วย 1 หรือพูดว่า df ของ TSS เท่ากับ kn - 1 ดังนั้น

$$\hat{\sigma}_1^2 = \frac{\text{TSS}}{kn-1} = \text{Total MS(TMS)}$$

ทั้งนี้ $\hat{\sigma}^2$ (วาเรียนซ์แฮต) หมายถึงค่าประมาณวาเรียนซ์

(2) df ของทรีตเมนต์ ได้แก่จำนวนทรีตเมนต์ลบด้วย 1 หรือพูดว่า df ของ SSTr เท่ากับ k - 1 ดังนั้น

$$\hat{\sigma}_2^2 = \frac{SSTr}{k-1} = \text{MS Treatment (MSTr)}$$

(3) df ของค่าเบี่ยงเบนภายในกลุ่ม คือความแตกต่างระหว่าง df ของ TSS และ SSTr = (kn - 1) - (k - 1) = k (n - 1) ดังนั้น

$$\hat{\sigma}_3^2 = \frac{SSE}{k(n-1)} = \text{MSError (MSE)}$$

ในการทดสอบสมมุติฐานนั้น กระทำโดยใช้อัตราส่วนของวาเรียนซ์ ดังนี้

$$F = \frac{MSTr}{MSE} = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_3^2}$$

เมื่อเรากลับไปพิจารณาสมการในตอน 8.6 ที่ว่า

$$F = \frac{\text{ค่าประมาณ } \hat{\sigma}^2 \text{ จากความแตกต่างระหว่างปัญหาทดลอง}}{\text{ค่าประมาณ } \hat{\sigma}^2 \text{ จากความคลาดเคลื่อนของการทดลอง}}$$

ถ้า $H_0 : \mu_1 = \mu_2 = \ldots \mu_k$ เป็นความจริงก็หมายถึงว่า $\hat{\sigma}_2^2 = \hat{\sigma}_3^2$ แต่ถ้าบางค่าเฉลี่ยสูงหรือต่ำกว่าค่าอื่นก็หาได้ว่า $\hat{\sigma}_2^2 > \hat{\sigma}_3^2$ ในการทดสอบสมมุติฐานนั้นเรานำค่า F คำนวณได้นี้ไปเปรียบเทียบกับค่า F จากตารางที่ระดับความแตกต่าง 0.05 หรือ 0.01 ในการเปิดตาราง F นั้น ให้ $k-1$ เป็น df ของตัวตั้งและ $k(n-1)$ เป็น df ของตัวหาร วิธีการทดสอบอาจสรุปได้ดังนี้

(1) ถ้าค่า F ที่คำนวณได้มากกว่า $F_{0.05}$ ที่ df = $[k-1, k(n-1)]\infty$ ก็สรุปว่าค่าเฉลี่ยมีความแตกต่างกันในทางสถิติ ที่ระดับความแตกต่าง 0.05 หรือสรุปว่าความแตกต่างมีนัยสำคัญ[24] อาจแสดงเป็นสัญลักษณ์ด้วยเครื่องหมายดอกจันจำนวน 1 ดอก (*) ไว้ที่มีนสแควร์[25] ของทรีตเมนต์ หรือไว้ที่ค่า F ก็ได้

(2) ถ้าค่า F ที่คำนวณได้มากกว่า $F_{0.01}[k-1, k(n-1)]$ ก็สรุปว่าค่าเฉลี่ยมีความแตกต่างกันในทางสถิติที่ระดับความแตกต่าง 0.01 หรือสรุปว่าความแตกต่างมีนัยสำคัญยิ่ง[26] ซึ่งก็อาจแสดงเป็นสัญลักษณ์ด้วยเครื่องหมายดอกจัน จำนวน 2 ดอก (**) ไว้ที่มีนสแควร์ ของทรีตเมนต์หรือไว้ที่ค่า F ก็ได้

ในทั้ง 2 กรณีข้างบนนี้ เราปฏิเสธ H_0

(3) ถ้าค่า F ที่คำนวณได้ไม่สูงกว่า $F_{0.05}[k-1, k(n-1)]$ ก็สรุปว่าค่าเฉลี่ยไม่แตกต่างกันในทางสถิติ หรือความแตกต่างไม่มีนัยสำคัญในทางสถิติ[27] อาจไม่ต้องมีสัญลักษณ์ใด ๆ กำกับค่า F หรืออาจเขียนว่า "ns" ก็ได้

ทั้งนี้ค่าในวงเล็บในแต่ละข้อคือ ค่า df ของทรีตเมนต์และความคลาดเคลื่อนจากการทดลอง ซึ่งใช้ในการเปิดหาค่า F นั่นเอง

8.8 วิธีการคำนวณผลบวกของค่ายกกำลังสอง [28]

ตอนนี้เป็นตอนขยายเพิ่มเติมของตอน 8.7 ทั้งนี้เพื่อแสดงวิธีการสร้างสมการคำนวณ TSS และ SSTr ในรูปแบบที่สะดวกต่อการใช้ การดัดแปลงสมการสำหรับการคำนวณอาจกระทำได้ ดังนี้

1. Total Sum of Squares จากสมการ

$$\begin{aligned}
TSS &= \sum(X_{ij} - \bar{X}..)^2 \\
&= \sum(X_{ij}^2 - 2\bar{X}..X_{ij} + \bar{X}^2..)\frac{1}{2} \\
&= \sum X_{ij}^2 - 2\bar{X}..\sum X_{ij} + kn\bar{X}^2.. \\
&= \sum X_{ij}^2 - (2\sum X_{ij})^2/kn + (\sum X_{ij})^2/kn
\end{aligned}$$

ทั้งนี้เพราะ $\bar{X}.. = \sum X_{ij}/kn$ และ $\bar{X}^2.. = (\sum X_{ij}/kn)^2$ ดังนั้นจากสมการสุดท้ายข้างบนเราอาจแสดงไว้ว่า

$$TSS = \sum X_{ij}^2 - \frac{(\sum X_{ij})^2}{kn}$$

$\sum X_{ij}$ คือผลบวกของค่าสังเกตทั้งหมดในการทดลอง ส่วน kn คือจำนวนค่าสังเกตทั้งหมด เราเรียกว่า $(\sum X_{ij})^2/kn$ ว่าเป็นค่าปรับปรุง[29] ซึ่งต่อไปจะเรียกสั้น ๆ ว่า correction factor ย่อว่า "CF" ดังนั้นสมการในการคำนวณ TSS คือ

$$TSS = \sum X_{ij}^2 - CF \qquad ...(8\text{-}4)$$

2. Treatment Sum of Squares จากสมการ

$$\begin{aligned}
SSTr &= \sum(\bar{X}_i. - \bar{X}..)^2 \\
&= \sum_i\sum_j(\bar{X}_i. - \bar{X}..)^2 \\
&= \sum_i n_i(\bar{X}_i. - \bar{X}..)^2 \\
&= \sum_i n_i(\bar{X}_i^2. - 2\bar{X}..\bar{X}_i. + \bar{X}^2..) \\
&= \sum_i n_i\left[\left(\sum_j X_{ij}\right)/n_i\right]^2 - 2\left(\sum_i\sum_j X_{ij}\right)^2/kn + \left(\sum_i\sum_j X_{ij}\right)^2/kn
\end{aligned}$$

$$= \sum_i T_i^2 / n_i - \left(\sum_i \sum_j X_{ij} \right) / kn$$
$$= \sum_i T_i^2 / n_i - CF$$

ทั้งนี้ให้ $(\sum_j X_j)_i = T_i$ คือผลบวกของข้อมูลแต่ละทรีตเมนต์ และ n คือจำนวนค่าสังเกตภายในทรีตเมนต์ แต่ละทรีตเมนต์อาจมีค่าสังเกตไม่เท่ากันก็ได้ ดังนั้นแต่ละทรีตเมนต์มีค่า n ของตัวเองดังนี้

$$SSTr = \frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \ldots + \frac{T_k^2}{n_k} - CF$$
$$= \frac{\sum T_i^2}{n_i} - CF \qquad \ldots(8\text{-}5)$$

ในกรณีที่ข้อมูลทุก ๆ กลุ่ม มีค่าสังเกตเท่ากัน ก็อาจใช้สมการ

$$SSTr = \frac{T_1^2 + T_2^2 + \ldots + T_k^2}{n} - CF$$
$$= \frac{\sum T_i^2}{n} - CF \qquad \ldots(8\text{-}6)$$

สมการ (8-4) และสมการ (8-5) หรือสมการ (8-6) นี้ จัดเป็นตัวอย่างของสมการที่ใช้สำหรับคำนวณหาผลบวกของค่ายกกำลังสอง ในการทดลองที่มีความสลับซับซ้อนขึ้นไปกว่านี้ ก็มีสมการเพิ่มเติมไปจากนี้ อย่างไรก็ดีวิธีการสร้างสมการย่อมคล้ายคลึงกันนั่นเอง

8.9 แบบจำลองทางคณิตศาสตร์

ค่าสังเกต วัด ตรวจนับ ฯลฯ ในทางสถิติแต่ละค่าที่เราสมมุติว่าเป็น X_{ij} นั้น เหมือนกับไก่ที่มีขนห่อหุ้ม จะหนาจะบางขึ้นอยู่กับพันธุ์และเพศ ดังนั้นเราอาจเขียนเป็นสมการดังนี้

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad \ldots(8\text{-}7)$$

μ คือค่าเฉลี่ยของประชากร ซึ่งเป็นค่าเฉลี่ยของสมาชิกทั้งหมดที่มีในประชากรนั้น ๆ α_i คือผลของการทดลองที่เราใส่เข้าไปในประชากร เช่น ถ้าทดลองเลี้ยงวัวกับไก่ก็อาจเป็นการให้อาหารที่มีระดับโปรตีนต่าง ๆ กัน และ ε คือความคลาดเคลื่อนที่เกิดขึ้นในทางการทดลอง ทั้งนี้ i มีค่าจาก 1, 2, ..., k ; j = มีค่าจาก 1, 2, ..., n เราเรียกสมการ (8-7) ว่าแบบจำลองทางคณิตศาสตร์[30] ในแบบจำลองนี้เรามีข้อกำหนดบังคับไว้ดังนี้

(1) ความคลาดเคลื่อนจากการทดลองเป็นอิสระแก่กัน มีการกระจายที่เป็นแบบปกติ ที่มีพารามิเตอร์ 0, σ^2 คือมีค่าเฉลี่ย 0 และวาเรียนซ์ σ^2 คืออาจเขียนว่า $\varepsilon \sim NID(0, \sigma^2)$[31] ซึ่งเราจะอธิบายเพิ่มเติมถึงข้อกำหนดดังกล่าวนี้ในตอน 10.8 ต่อไป

(2) ผลของทรีตเมนต์เป็นแบบบวก[32] คือการเปลี่ยนแปลงจากทรีตเมนต์ไปอีกทรีตเมนต์หนึ่ง ความสัมพันธ์ระหว่างค่าเฉลี่ยของประชากร และค่าของทรีตเมนต์ (μ และ α_i) เป็นแบบบวก เช่น α_1 = ปุ๋ย 0 กก., α_2 = ปุ๋ย 10 กก., α_3 = ปุ๋ย 20 กก. ดังนั้นผลของปุ๋ยน่าจะเพิ่มในแบบบวก เช่น ให้ผลผลิต 200, 220 และ 240 กก./ไร่ ตามลำดับ เราจะอธิบายข้อกำหนดนี้เพิ่มเติมในตอน 10.8 ต่อไป

8.10 แบบฝึกหัด

1. ในการทดลองมีความคลาดเคลื่อนเกิดขึ้นเสมอ อยากทราบว่าความคลาดเคลื่อนคืออะไร เกิดขึ้นได้อย่างไร สามารถลดได้อย่างไร

2. เลขชุดหนึ่งมีการจัดระเบียบเป็นแถวและสดมภ์ เมื่อให้ค่าสังเกตเป็น X_{ij} ทั้งนี้ $i = 1, 2, \ldots, k$ (k = จำนวนแถว) $j = 1, 2, \ldots, n$ (n = จำนวนสดมภ์) จงแสดงสัญลักษณ์ของค่าดังต่อไปนี้

   ก. ค่าเฉลี่ยของแถว
   ข. ค่าเฉลี่ยของสดมภ์
   ค. ค่าที่ 5 ในสดมภ์ที่ 6
   ง. ค่าต่าง ๆ ในแถวที่ 2
   จ. ค่าต่าง ๆ ในสดมภ์ที่ 3

3. จงอธิบายถึงแหล่งที่มาของความปรวนแปรของข้อมูลที่ได้จากการใช้แผนการทดลอง
4. จงอธิบายถึงความหมายของคำต่อไปนี้

   ก. ปัญหาหรือทรีตเมนต์ (treatment)
   ข. ซ้ำ (replication)
   ค. การสุ่ม (randomization)

5. จากข้อมูลดังต่อไปนี้

A	5	4	8	6	3
B	9	7	8	6	9
C	3	5	2	3	7
D	2	3	4	1	4
E	7	6	9	4	7

จงคำนวณหา TSS, SSTr และ SSE และแสดงการทดสอบสมมุติฐาน

คำในบท

(1) basic research, (2) applied research, (3) monodisciplinary research, (4) interdisciplinary research, (5) *ex post facto* research, (6) experimental research, (7) parameter, (8) experimental design, (9) treatment (10) control or check treatment, (11) experimental unit, (12) coefficient of variation, (13) experimental error, (14) replicate or replication, (15) randomization, (16) block, (17) local control, (18) guard row, (19) analysis of covariance, (20) column, (21) grand total, (22) grand mean, (23) degrees of freedom, (24) significant, (25) mean square, (26) highly significant, (27) not significant, (28) sum of squares, (29) correction factor (CF), (30) mathematical model, (31) NID = normally independently distributed, (32) additivity.

บทที่ 9
แผนการทดลองแบบสุ่มสมบูรณ์

9.1 คำนำ

เราได้ทราบพื้นฐานของการวางแผนการทดลอง และการวิเคราะห์วาเรียนซ์มาแล้วจากบทที่ 8 จะเห็นได้ว่าการวางแผนการทดลองประกอบด้วย เทคนิคในทางปฏิบัติ และเทคนิคในทางสถิติมากมาย ตอนที่ 8.7 ได้มีการแสดงวิธีการแยกความปรวนแปรออกเป็นส่วน ๆ ตามแหล่งที่มา คือความปรวนแปรที่มาจากผลของทรีตเมนต์โดยตรง และที่เกิดจากความคลาดเคลื่อนในการทดลอง

ในบทนี้เราจะได้ศึกษาถึงแผนการทดลองชนิดแรก คือแผนการทดลองแบบสุ่มสมบูรณ์[1] หรืออาจเรียกย่อ ๆ ว่า “CRD” เป็นแผนการทดลองที่นำทรีตเมนต์เข้าทดลองกับหน่วยทดลองใดก็ได้อย่างอิสระ ซึ่งเป็นแผนการทดลองแบบที่ง่ายที่สุด คือง่ายในการดำเนินการทดลอง การจัดหาหน่วยทดลอง การเก็บข้อมูล การวิเคราะห์วาเรียนซ์ ตลอดถึงการอธิบายผลการทดลอง ใช้สำหรับการทดลองที่มีปัญหาเพียงกลุ่มเดียว[2] ปัญหาอื่นถึงมีเราก็ไม่ศึกษาร่วมกันในที่นี้ คือพยายามให้ปัญหาอื่นคงที่เอาไว้ เช่น การเปรียบเทียบพันธุ์พืช ก็ศึกษาเรื่องความแตกต่างของพันธุ์เพียงอย่างเดียวเท่านั้น ส่วนระยะปลูก ชนิดของปุ๋ย ฯลฯ ให้ใช้เหมือนกันสำหรับทุกพันธุ์ แผนการทดลองนี้เหมาะสำหรับการทดลองที่มีขนาดเล็ก มีทรีตเมนต์น้อย ต้องใช้หน่วยทดลองไม่มากนัก และแผนการทดลองนี้ เหมาะสมกับการทดลองที่หน่วยทดลองที่ใช้มีความสม่ำเสมอกัน การทดลองทางเกษตรส่วนมากหาหน่วยทดลองที่สม่ำเสมอได้ยาก จึงมักใช้แผนการทดลองชนิดนี้กับการทดลองในห้องปฏิบัติการ การทดลองในเรือนเพาะชำ การทดลองในกระถาง การทดลองที่ใช้อาหารเลี้ยงเชื้อ การทดลองทางชีววิทยา และวิศวกรรม เป็นต้น

9.2 เทคนิคในการใช้แผนการทดลองแบบสุ่มสมบูรณ์

แผนการทดลองแบบสุ่มสมบูรณ์ใช้เมื่อหน่วยทดลองมีความสม่ำเสมอดีมาก ไม่มีความแตกต่างระหว่างกันมากพอที่จะแยกเป็นกลุ่มหรือเป็นพวกเหมือนดังกล่าวไว้ในบทที่ 8 ใช้กับการทดลองในเรือนเพาะชำหรือในห้องปฏิบัติการ ใช้ในการทดลองเกี่ยวกับพืชหรือสัตว์ก็ได้ ทั้งนี้ถ้าหากว่า เราสามารถควบคุมแหล่งที่ทำให้เกิดความคลาดเคลื่อนได้ดี แผนการทดลองนี้มีกฎเกณฑ์บังคับน้อยที่สุด แต่ละปัญหาหรือทรีตเมนต์อาจทำกี่ซ้ำก็ได้ คือจำนวนซ้ำของแต่ละทรีตเมนต์ไม่จำเป็นต้องเท่ากัน

ในการทดลองเกี่ยวกับพืชนั้น ใช้ได้ดีเมื่อมีจำนวนทรีตเมนต์น้อย เช่น ในการเปรียบเทียบพันธุ์พืช ถ้ามีพืชหลายพันธุ์เกินไป ก็ต้องใช้เนื้อที่ทดลองขนาดใหญ่ เช่นนี้ทำให้ครอบคลุมพื้นที่ ๆ มีความแตกต่างกันมาก ซึ่งจะทำให้มีความคลาดเคลื่อนเพิ่มขึ้น ในกรณีเช่นนี้จึงควรใช้แผนการทดลองแบบอื่น ๆ ที่จะกล่าวถึงในบทต่อ ๆ ไป

ก. วิธีการสุ่ม หลักสำคัญในการทดลองนี้ต้องให้หน่วยทดลองต่าง ๆ มีโอกาสได้รับทรีตเมนต์ใดก็ได้ คือโอกาสที่ทรีตเมนต์ใดตกลงในหน่วยทดลองใดเท่ากัน ในการทดลองทางพืชนั้น ในขั้นต้น เราต้องจัดวางตำแหน่งทรีตเมนต์ในแปลงเสียก่อน ซึ่งกระทำโดยวิธีสุ่ม ทั้งนี้โดยเขียนแผนผังแปลงทดลองในแผ่นกระดาษ สมมุติว่าเราทดลองเปรียบเทียบพันธุ์พืช 4 พันธุ์ คือ A, B, C และ D ซึ่ง แต่ละพันธุ์ปลูก 3 แปลง ถ้าให้จำนวนพันธุ์เป็น k (เช่น $k = 4$) จำนวนแปลงสำหรับแต่ละพันธุ์เป็น n (เช่น $n = 3$) ดังนั้นจำนวนแปลงย่อยทั้งหมดคือ $k \times n = kn = 12$ แปลง ให้หมายเลขแปลงตั้งแต่ 1 ถึง 12 แล้วสุ่มเพื่อกำหนดตำแหน่งของแต่ละพันธุ์โดยใช้วิธีการดังต่อไปนี้

1. การสุ่มโดยใช้ตารางเลขสุ่ม ตารางเลขสุ่ม ปรากฏในหนังสือสถิติแทบทุกเล่ม เช่น ถ้าตัดมาบางส่วนจะมีลักษณะเช่นนี้ :

	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
00	54463	22662	65905	70639	79365	67382	29085	69831	47058	08186
01	15389	85205	18850	39226	42249	90669	96325	23248	60933	26927
02	85941	40756	82414	02015	13858	78030	16269	65978	01385	15345
03	61149	69440	11286	88218	58925	03638	52862	62733	33451	77455
04	05219	81619	10651	67079	92511	59888	84502	72095	83463	75577

ในการสุ่มเราหาจุดเริ่มต้นเสียก่อน ทั้งนี้โดยใช้ปลายดินสอชี้ลงไปในตาราง เมื่อถูกเลขใดแล้วก็ใช้เป็นเลขเริ่มต้น อ่านมาเป็นชุดจำนวน 12 ชุด (เท่ากับจำนวนแปลง) แล้วจัดลำดับชุดจากน้อยไปหามากเพื่อกำหนดตำแหน่งของแปลง เช่น จากตารางข้างบนนี้เราเริ่มที่เลข 6 (ลูกศรชี้) แล้วอ่านเป็นชุด ชุดละ 3 ตัวเลข การอ่านจะอ่านไปทิศทางใดก็ได้ เช่น อ่านลงมาด้านล่าง เมื่อหมดแล้วก็อ่านกลับขึ้นด้านบนจนครบ 12 ชุด แล้วจัดลำดับจากน้อยไปหามากดังนี้

ชุดเลขสุ่ม	659	188	824	112	106	516	868	140	503	057	063	922
ลำดับที่	9	6	10	4	3	8	11	5	7	1	2	12

เมื่อมีพืช 4 พันธุ์ แต่ละพันธุ์ปลูก 3 แปลง ดังนั้นพันธุ์ A ปลูกใน 3 แปลงแรก คือแปลงที่ 9, 6, 10 พันธุ์ B ปลูกใน 3 แปลงถัดไป คือแปลงที่ 4, 3, 8 พันธุ์ C และ D ปลูกในแปลงที่ 11, 5, 7 และ 1, 2, 12 ตามลำดับ ซึ่งอาจแสดงลำดับของแปลงดังรูป 9.2.1

2. การสุ่มโดยใช้จับฉลาก ในกรณีที่ไม่มีหรือไม่ต้องการใช้ตารางเลขสุ่ม ก็อาจสุ่มโดยการจับฉลาก คือ เขียนแผนผังและลงเลขที่แปลง (1 ถึง 12) เช่นเดียวกับที่กล่าวมาแล้วข้างบน ถ้าแต่ละพันธุ์ปลูก 3 แปลง ก็ให้เขียนชื่อพันธุ์ใส่แผ่นกระดาษเล็ก ๆ พันธุ์ละ 3 แผ่น ม้วนให้กลม

1 D	2 D	3 B	4 B
5 C	6 A	7 C	8 B
9 A	10 A	11 C	12 D

รูป 9.2.1 แผนผังของแปลงทดลองและการสุ่มวางทรีตเมนต์

ใส่ลงไปในภาชนะ เช่น ถ้วยแก้ว แล้วเขย่าให้เข้ากัน ต่อจากนั้นก็หยิบขึ้นมาอย่างสุ่ม ถ้าหยิบครั้งแรกได้พันธุ์ใดก็ให้ใช้กับแปลงแรก ในการหยิบครั้งถัดไปให้ใช้กับแปลงที่ 2 และ 3 จนถึงแปลงสุดท้าย ในการสุ่มแบบนี้ต้องทิ้งฉลากที่หยิบออกมาแต่ละครั้ง การทำเช่นนี้ก็ได้ลำดับแปลงอย่างสุ่มเช่นเดียวกัน

ข. การวิเคราะห์วาเรียนซ์ ในแผนการทดลองแบบ CRD ความแตกต่างระหว่างค่าสังเกตเกิดจากแหล่งต่าง ๆ 2 แหล่ง ดังสมการ (8-2) และมีจำนวน df ดังแสดงในตอน 8.7 ซึ่งอาจสรุปได้อีกครั้งดังนี้

แหล่งของความปรวนแปรคือ TSS = SSTr + SSE

จำนวน df คือ $(kn-1)$, $(k-1)$ และ $[\,k(n-1)\,]$ ตามลำดับ

สมการคำนวณ TSS, SSTr และ SSE แสดงไว้แล้วในตอน 8.8 เมื่อได้ค่าต่าง ๆ ที่ต้องการก็จัดลงตารางดังแสดงในตาราง 9.2.1 ซึ่งเรียกว่าตารางวิเคราะห์วาเรียนซ์ (analysis of variance, ANOVA)

ตาราง 9.2.1 ตารางวิเคราะห์วาเรียนซ์ใน CRD (1)

Sources of Variation (Sources)	Degrees of Freedom (df)	Sum of Squares (SS)	Mean square (MS)	F - value
Treatments	k - 1	$\frac{\sum T_i^2}{n} - CF$ = SSTr	$\frac{SSTr}{k-1}$ = MSTr	$\frac{MSTr}{MSE}$
Error	k(n - 1)	TSS − SSTr = SSE	$\frac{SSE}{k(n-1)}$ = MSE	
Total	kn - 1	$\sum X_{ij}^2 - CF$ = TSS		

(1) **ในหนังสือเล่มนี้ ตารางวิเคราะห์วาเรียนซ์จะใช้ชื่อส่วนต่าง ๆ ในตารางเป็นคำในภาษาอังกฤษ ทั้งหมดหัวสดมภ์ใช้คำย่อดังแสดงในวงเล็บ**

9.3 การวิเคราะห์วาเรียนซ์เมื่อแต่ละทรีตเมนต์มีค่าสังเกตเท่ากัน

ในการทดลองส่วนมากเรามักใช้หน่วยการทดลองเท่ากันในทุก ๆ ทรีตเมนต์ วิธีการวิเคราะห์วาเรียนซ์ได้แสดงโดยละเอียดแล้วในตอน 8.8 ข้อมูลที่นำมาใช้เป็นตัวอย่างแสดงการวิเคราะห์วาเรียนซ์เป็นข้อมูลที่ได้จากการทดลองเปรียบเทียบผลการปราบวัชพืช ที่มีต่อผลผลิตของถั่วเหลืองพันธุ์หนึ่ง ดังแสดงในตาราง 9.3.1 ซึ่งแต่ละทรีตเมนต์ทดลอง 4 แปลงย่อย การทดลองมีแบบจำลองทางคณิตศาสตร์ ดังสมการ (8-7)

ตาราง 9.3.1 ผลผลิตของถั่วเหลืองที่ทดสอบเทคนิคในการกำจัดวัชพืชในช่วงอายุต่าง ๆ กัน โดยทำการทดลองวิธีละ 4 ครั้ง หรือ 4 ซ้ำ

การทดลอง (treatments)	ซ้ำ(replications)				รวม	เฉลี่ย
	1	2	3	4		
A. ไม่กำจัดวัชพืช [1]	12	16	16	13	57	14.25
B. กำจัดเมื่ออายุ 30 วัน	24	20	22	26	92	23.00
C. กำจัดเมื่ออายุ 30, 60 วัน	28	21	34	32	115	28.75
D. กำจัดเมื่ออายุ 25, 50, 75 วัน	32	35	32	29	128	32.00
E. ใช้สารเคมี	32	30	33	26	121	30.25
รวม					513	

[1] **การทดลองที่ไม่กำจัดวัชพืชเป็นทรีตเมนต์เปรียบเทียบ**[3]

ข้อกำหนดสำหรับการวิเคราะห์วาเรียนซ์

ข้อมูลที่นำมาวิเคราะห์วาเรียนซ์เพื่อนำไปทดสอบความแตกต่างระหว่างทรีตเมนต์โดยใช้ F-test ต้องมีคุณสมบัติสอดคล้องกับข้อกำหนด คือ (1) ข้อมูลกระจายแบบปกติ และ (2) ทรีตเมนต์ต่าง ๆ มีวาเรียนซ์เท่ากัน

การวิเคราะห์วาเรียนซ์ดำเนินเป็นขั้น ๆ ดังนี้

ขั้นที่ 1 นำข้อมูลที่สำรวจได้จัดระเบียบเข้ากลุ่มดังแสดงไว้ในตาราง 9.3.1 คือข้อมูลของแต่ละทรีตเมนต์วางไว้ในแถวของทรีตเมนต์นั้น แล้วคำนวณหาผลรวมของทรีตเมนต์นั้น ๆ และผลรวมทั้งหมด

ขั้นที่ 2 หาค่า df (degrees of freedom) ของส่วนต่าง ๆ ดังแสดงในตาราง 9.2.1

ขั้นที่ 3 หาค่า SS (sum of squares) ชนิดต่าง ๆ คือ TSS, SSTr และ SSE ก่อนอื่นต้องคำนวณหาค่า CF (correction factor) ดังนี้

$$CF = \frac{(\sum X_{ij})^2}{kn}$$

$$= \frac{(513)^2}{20}$$

ต่อไปก็คำนวณค่า SS ต่าง ๆ ดังนี้

$$TSS = \text{Total sum of squares}$$

$$TSS = \sum X_{ij}^2 - CF$$

$$= 12^2 + 16^2 + ... + 26^2 - 13,158.45$$

$$= 14,169 - 13,158.45$$

$$= 1,010.55$$

$$SSTr = \text{Treatment sum of squares}$$

$$SSTr = \frac{\sum T_i^2}{n} - CF$$

$$= \frac{(57^2 + 92^2 + ... + 121^2)}{4} - 13,158.45$$

$$= 13,990.75 - 13,158.45$$

$$= 832.30$$

$$SSE = \text{Error sum of squares}$$

$$SSE = TSS - SSTr$$

$$= 1,010.55 - 832.30$$

$$= 178.25$$

ขั้นที่ 4 นำ df และค่า SS ต่าง ๆ ที่คำนวณได้ลงตารางวิเคราะห์วาเรียนซ์ ดังแสดงใน ตาราง 9.3.2

ขั้นที่ 5 คำนวณค่า MS (mean square) ต่าง ๆ ดังนี้

$$\text{Treatment MS (MSTr)} = \frac{SSTr}{\text{Treatment df}}$$

$$= \frac{832.30}{4}$$

$$= 208.08$$

$$\text{Error MS (MSE)} = \frac{\text{SSE}}{\text{Error df}} = \frac{178.25}{15} = 11.88$$

ทั้งนี้ไม่จำเป็นต้องคำนวณหา MS ของ Total แต่อย่างใด

ขั้นที่ 6 คำนวณหาค่า F (F-value) ของปัญหาที่ต้องการทดสอบ คือวิธีการกำจัดวัชพืช หรือทรีตเมนต์ วิธีการคำนวณคือ นำ MSE ไปหาร MSTr ดังนี้

$$F = \frac{\text{MSTr}}{\text{MSE}} = \frac{208.08}{11.88} = 17.52$$

แล้วนำผลการวิเคราะห์ทั้งหมดลงตารางดังตาราง 9.3.2

ตารางที่ 9.3.2 ผลการวิเคราะห์วาเรียนซ์ของข้อมูลในตาราง 9.3.1

Sources	df	SS	MS	F	F value 5%	F value 1%
Treatments	4	832.30	208.08	17.52**	3.06	4.89
Error	15	178.25	11.88			
Total	19	1,010.55				

**** = แตกต่างทางสถิติในระดับ 1 เปอร์เซ็นต์, CV = 13.44%**

ขั้นที่ 7 ทดสอบสมมุติฐาน ในการวิเคราะห์วาเรียนซ์นั้น เราต้องการทดสอบว่า ค่าเฉลี่ยต่าง ๆ เท่ากันหรือไม่ ซึ่งอาจตั้งได้หลายแบบ แต่ในขั้นนี้ตั้งสมมุติฐานดังนี้

$H_0 = \mu_A = \mu_B = \mu_C = \mu_D = \mu_E$

H_1 = ค่า μ ไม่เท่ากันอย่างน้อย 1 คู่

การทดสอบกระทำโดยเปรียบเทียบค่า F ที่คำนวณได้กับค่าที่เปิดจากตาราง F (ตาราง ผ.9) ที่ df ตัวตั้ง 4 และตัวหาร 15 พบว่าค่า F ที่คำนวณได้ (17.52) สูงกว่าค่าในตารางที่ระดับความ แตกต่าง 0.01 df 4, 15 ซึ่งเท่ากับ 4.89 ดังนั้นจึงปฏิเสธสมมุติฐาน คือสรุปว่าบางปัญหาหรือบางทรีตเมนต์ แตกต่างจากทรีตเมนต์อื่นอย่างมีนัยสำคัญยิ่ง หรืออาจเขียนสั้น ๆ ว่า

17.52** (เพราะ $F = 17.52 > F_{(.01)(4,15)} = 4.89$)

ซึ่งเห็นว่าเหนือเลข 17.52 ได้ใส่เครื่องหมายดอกจัน 2 จุด ซึ่งเป็นการแสดงว่าแตกต่างในระดับนัยสำคัญยิ่ง (ถ้าหากแตกต่างในระดับ 0.05 เรียกว่าแตกต่างในระดับนัยสำคัญ ซึ่งอาจกำกับด้วยเครื่องหมายดอกจันเพียง 1 จุด)

ขั้นที่ 8 คำนวณสัมประสิทธิ์ของความปรวนแปร หรือ CV (coefficient of variation) ดังนี้

$$CV(\%) = \frac{\sqrt{MSE}}{\overline{X}...} \times 100 \qquad ...(9\text{-}1)$$

$$= \frac{\sqrt{11.88}}{25.65} \times 100$$

$$= 13.44\%$$

"CV" คืออัตราส่วนระหว่างส่วนเบี่ยงเบนมาตรฐานและค่าเฉลี่ยนั่นเอง ค่าเฉลี่ย ($\overline{X}..$) คือ ผลรวมทั้งหมดหารด้วยจำนวนค่าสังเกต (kn) ใช้วัดขนาดของความปรวนแปรในการทดลองที่เกิดขึ้นโดยไม่ทราบสาเหตุ เป็นที่น่าสังเกตต่อไปว่า ค่า CV ของการทดลองในลักษณะชนิดเดียวกัน เช่น ผลผลิตนั้น ถ้าใช้เทคนิคที่เหมือนกันและวิเคราะห์วิธีเดียวกัน ไม่ว่าทดลองเมื่อใดหรือที่ใดก็จะใกล้เคียงกัน ถ้าหากว่าค่านี้สูงเกินค่าที่ควรจะเป็น ก็แสดงว่าต้องปรับปรุงวิธีการทดลองหรือวิธีการวิเคราะห์เพื่อให้แยกความปรวนแปรไม่ทราบสาเหตุออกมาให้ได้มากที่สุด ก็จะเหลือความ ปรวนแปรอันเป็นธรรมชาติของลักษณะนั้น ๆ

ขั้นที่ 9 คำนวณความคลาดเคลื่อนมาตรฐาน [4] เมื่อสิ้นสุดการวิเคราะห์วาเรียนซ์ ถ้าหากว่าผลของทรีตเมนต์มีความแตกต่างกันในทางสถิติ ก็ต้องดำเนินการในขั้นตอนต่อเนื่อง เพื่อเปรียบเทียบว่าทรีตเมนต์ใดแตกต่างจากทรีตเมนต์ใดบ้าง ในการเปรียบเทียบค่าเฉลี่ยต้องใช้ความคลาดเคลื่อนมาตรฐานชนิดใดชนิดหนึ่งใน 2 ชนิดดังนี้

(1) ความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย[5] คือ

$$s_{\overline{X}} = \sqrt{\frac{MSE}{n}} \qquad ...(9\text{-}2)$$

ค่า MSE = 11.88, n = ค่าสังเกตแต่ละทรีตเมนต์ = 4 ดังนั้น

$$s_{\overline{X}} = \sqrt{\frac{11.88}{4}}$$

$$= 1.72$$

(2) ความคลาดเคลื่อนมาตรฐานของความแตกต่าง[6] คือ

$$s_{\bar{d}} = \sqrt{\frac{2MSE}{n}} \quad ...(9\text{-}3)$$

$$= \sqrt{\frac{2(11.88)}{4}}$$

$$= 2.44$$

การเปรียบเทียบค่าเฉลี่ยจะอธิบายในบทที่ 12 ต่อไป ในการวิเคราะห์วาเรียนซ์นั้น ถ้าพบว่าทรีตเมนต์ไม่มีความแตกต่างกันในทางสถิติก็ไม่ต้องดำเนินการอะไรต่อไป

การวิเคราะห์วาเรียนซ์โดยการลดค่าสังเกต

ในการวิเคราะห์วาเรียนซ์กรณีที่ค่าสังเกตมีค่าสูง ๆ เราอาจนำค่าบางค่าไปลบออกเพื่อลดค่าลง ทำให้คำนวณได้ง่าย และถูกต้อง การทำเช่นนี้เรียกว่า การโคด (code) การลดลงไปของค่าสังเกตนั้น ไม่ทำให้ขนาดของวาเรียนซ์เปลี่ยนไป ดังนั้นผลของการทดสอบยังคงให้คำตอบเช่นเดิม เช่น ข้อมูลในตาราง 9.3.1 ถ้าลบด้วย 25 ทุกค่า ก็จะได้ข้อมูลดังตาราง 9.3.3

ตาราง 9.3.3 ผลการโคดข้อมูลในตาราง 9.3.1 โดยลบด้วย 25

ทรีตเมนต์	ซ้ำ				รวม
	1	2	3	4	
A	-13	-9	-9	-12	-43
B	-1	-5	-3	1	-8
C	3	-4	9	7	15
D	7	10	7	4	28
E	7	5	8	1	21
รวม					13

ซึ่งเห็นว่า เป็นเลขจำนวนต่ำ ๆ เมื่อทำการวิเคราะห์ก็จะได้ผลดังนี้

$$CF = \frac{13^2}{20} = 8.45$$

$$TSS = (-13)^2 + (-9)^2 + ... + 1^2 - 8.45 = 1,010.55$$

$$SSTr = \frac{1}{4}\left[(-43)^2 + (-8)^2 - \ldots\ldots + 21^2\right] - 8.45 = 832.30$$

$$SSE = TSS - SSTr$$

$$= 1{,}010.55 - 832.30 = 178.25$$

ซึ่งได้ผลเหมือนการวิเคราะห์ข้อมูลดิบดังตาราง 9.3.2 ทุกประการ แต่วิธีนี้ง่าย ทำได้รวดเร็ว ตัวเลขที่ใช้โคดจะมีค่าเท่าใดก็ได้ แต่ควรให้ได้ผลรวมต่ำที่สุด ดังนั้นมักอยู่ใกล้ ๆ ค่าเฉลี่ย

9.4 การวิเคราะห์วาเรียนซ์ของ CRD เมื่อมีค่าสังเกตไม่เท่ากัน

ในการทดลองโดยใช้ CRD ทรีตเมนต์อาจมีค่าสังเกตไม่เท่ากัน อาจจะไม่เท่ากันตั้งแต่เริ่มทดลอง หรือเมื่อทดลองแล้วเราไม่อาจเก็บข้อมูลบางหน่วยได้ เมื่อเป็นเช่นนี้ก็อาจวิเคราะห์วาเรียนซ์จากข้อมูลที่เหลือ วิธีการก็คล้ายคลึงกับที่กล่าวมาแล้วในตอน 9.3 เพียงแต่มีการดัดแปลงวิธีคำนวณบางค่า ในกรณีเช่นนี้ n_i ไม่เท่ากัน ให้ค่าสังเกตทั้งหมดเท่ากับ N ทั้งนี้ N คือผลรวมของจำนวนแปลงทดลองจากทุกทรีตเมนต์ ในการวิเคราะห์วาเรียนซ์อาจคำนวณค่าต่าง ๆ ดังนี้

$$\text{Correction factor (CF)} = \frac{(\sum X_{ij})^2}{N}$$

$N = n_1 + n_2 + \ldots.. + n_i$ เนื่องจาก n_i ไม่เท่ากัน ดังนั้นการหา SSTr ต้องใช้การหารเดี่ยว ๆ แล้วมารวมกันทีหลัง ดังนี้

$$SSTr = \frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \ldots. + \frac{T_k^2}{n_k} - CF$$

$$= \sum \frac{T_i^2}{n_i} - CF$$

ตัวอย่าง ในการทดลองเปรียบเทียบอาหารไก่ ซึ่งใช้แหล่งโปรตีนแตกต่างกัน 4 ชนิด ทดลองเลี้ยงไก่ โดยแยกไก่ ออกเป็น 4 กรง แต่ละกรงมีจำนวนไก่ไม่เท่ากัน แล้วบันทึกผลของอาหารเป็นน้ำหนักเพิ่มเป็นกรัมดังตาราง 9.4.1

ตาราง 9.4.1 ผลการทดลองเปรียบเทียบอาหารไก่

ชนิดของโปรตีน	นน. เพิ่ม (กรัม / ตัว)											รวม
โปรตีน A	58	68	69	60	73	64	71	67	75			615
โปรตีน B	41	47	54	65	32	73	44	48	64	54		522
โปรตีน C	54	44	51	56	47	61	59	49	41	61	73	596
โปรตีน D	44	51	69	59	59	55	66					403
รวม												**2,136**

ในตารางนี้แต่ละทรีตเมนต์มีค่าสังเกตไม่เท่ากัน จำนวนข้อมูลทั้งหมด (N) เท่ากับ 37 มีวิธีการ วิเคราะห์วาเรียนซ์ดังนี้

$$CF = \frac{(\sum X_{ij})^2}{N} = \frac{(2,136)^2}{37} = 123,310.70$$

$$\begin{aligned} TSS &= \sum X_{ij}^2 - C.F \\ &= 68^2 + 68^2 + \ldots + 66^2 - 123,310.70 \\ &= 127,598.00 - 123,310.70 \\ &= 4,287.30 \end{aligned}$$

$$\begin{aligned} SSTr &= \sum T_i^2 / n_i - CF \\ &= \frac{(615)^2}{9} + \frac{(522)^2}{10} + \frac{(596)^2}{11} + \frac{(403)^2}{7} - 123,310.70 \\ &= 124,767.05 - 123,310.70 \\ &= 1,456.35 \end{aligned}$$

$$\begin{aligned} SSE &= TSS - SSTr \\ &= 4,287.30 - 1,456.35 \\ &= 2,830.95 \end{aligned}$$

ส่วน df ต่าง ๆ หาได้ดังนี้

df ของ TSS = N–1 = 37–1 = 36

df ของ SSTr = k–1 = 4–1 = 3

df ของ SSE = N–1–(k–1) = N–k = 37–4 = 33

แล้วนำผลที่ได้ลงตารางวิเคราะห์วาเรียนซ์ดังตาราง 9.4.2

ตาราง 9.4.2 ผลการวิเคราะห์วาเรียนซ์ของน้ำหนักเพิ่มของไก่ จากตาราง 9.4.1

Sources	df	SS	MS	F (คำนวณ)	F(table) 5%	F(table) 1%
Treatments	3	1,456.35	485.45	5.66**	2.89	4.44
Error	33	2,830.95	85.79			
Total	36	4,287.30				

** = แตกต่างทางสถิติในระดับ 1 เปอร์เซ็นต์, CV = 16.05%

เมื่อเปรียบเทียบกับค่า F ในตารางคือ $F_{(0.01)(3, 33)} = 4.44$ จะเห็นได้ว่าชนิดของโปรตีนมีความแตกต่างอย่างมีนัยสำคัญยิ่ง และอาจหาค่าอื่น ๆ ดังนี้

$$CV = \frac{\sqrt{MSE}}{\overline{X}} \times 100 = \frac{\sqrt{85.79}}{2{,}136/37} \times 100 = 16.04\%$$

การคำนวณความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย $(S_{\bar{d}})$ และความคลาดเคลื่อนมาตรฐานของความแตกต่าง $(S_{\bar{d}})$ กระทำได้ดังนี้

$$s_{\overline{X}} = \sqrt{MSE\frac{1}{2}\left[\frac{1}{n_i} + \frac{1}{n_j}\right]} \qquad ...(9\text{-}4)$$

เมื่อ n_i และ n_j คือจำนวนค่าสังเกตในทรีตเมนต์ที่ i และ j เราใช้ $s_{\overline{X}}$ เพื่อคำนวณค่าเปรียบเทียบค่าเฉลี่ย เช่น โดยใช้วิธี Duncan's Multiple Range Test (DMRT) ในทางปฏิบัติ คูณ MSE ด้วยค่าต่าง ๆ เสียก่อน แล้วคูณด้วย $\sqrt{[(1/2)(1/n_i + 1/n_j)]}$ ทีหลัง และคำนวณความคลาดเคลื่อนมาตรฐานของความแตกต่าง $(s_{\bar{d}})$

$$s_{\bar{d}} = \sqrt{MSE\left[\frac{1}{n_i} + \frac{1}{n_j}\right]} \qquad ...(9\text{-}5)$$

n_i และ n_j คือ จำนวนค่าสังเกตในทรีตเมนต์ที่ต้องการเปรียบเทียบกัน

9.5 การวิเคราะห์วาเรียนซ์เมื่อมีเพียงสองทรีตเมนต์

การทดลองที่มีเพียง 2 ทรีตเมนต์ เช่น ในการเปรียบเทียบพันธุ์พืช 2 พันธุ์ ปุ๋ย 2 สูตร วิธีการเก็บรักษาผลไม้ 2 วิธี อาหารสัตว์ 2 สูตร ฯลฯ โดยแต่ละทรีตเมนต์ทำซ้ำหลายครั้ง เราอาจเปรียบเทียบผลการทดลองโดยใช้วิธีเปรียบเทียบ 2 ตัวอย่าง ดังที่กล่าวมาแล้วในบทที่ 4 คำตอบที่น่าสนใจแต่ถ้าเราจะวิเคราะห์การทดลองดังเช่นการทดลองแบบ CRD ก็ได้ ในกรณีเช่นนี้เราพบคำตอบที่น่าสนใจว่า $F = t^2$

ตัวอย่าง จากการทดลองเปรียบเทียบปุ๋ย 2 สูตร กับยางพาราปลูกใหม่ เริ่มใช้ปุ๋ยแต่ละสูตรตั้งแต่ยางอายุได้ 3 ปี เป็นต้นไป เมื่อยางอายุได้ 7 ปี ก็วัดเส้นรอบวงของลำต้นในระดับความสูง 150 ซม. จากยางซึ่งได้รับปุ๋ยสูตรละ 10 ต้น ดัง ตาราง 9.5.1

ตาราง 9.5.1 เส้นรอบวงของต้นยางอายุ 7 ปี เมื่อได้รับปุ๋ย 2 สูตร

ปุ๋ย	เส้นรอบวงของลำต้น (ซม.)									รวม
ปุ๋ย A	58	56	62	54	54	56	53	55	52	57	557
ปุ๋ย B	64	59	63	63	60	59	58	58	62	60	606

เมื่อทำการวิเคราะห์วาเรียนซ์ข้อมูลจากตารางข้างบนก็จะได้

$$CF = \frac{(1{,}163)^2}{20} = 67{,}628.45$$

$$TSS = 58^2 + 56^2 + \ldots + 60^2 - 67{,}628.45 = 238.55$$

$$SSTr = \frac{(557)^2 + (606)^2}{10} - 67{,}628.45 = 120.05$$

$$SSE = 238.55 - 120.05 = 118.50$$

แล้วนำค่า sum of squares ลงตารางวิเคราะห์วาเรียนซ์ดังตาราง 9.5.2

ตาราง 9.5.2 ผลการวิเคราะห์วาเรียนซ์ผลการใช้ปุ๋ยยางพารา ตาราง 9.5.1

Sources	df	SS	MS	F (คำนวณ)	F(table) 5%	F(table) 1%
Treatments	1	120.05	120.05	18.24**	4.41	8.28
Error	18	118.50	6.58			
Total	19	238.55				

** = แตกต่างทางสถิติในระดับ 1 เปอร์เซ็นต์ ; CV = 4.41%

เมื่อเปิดตาราง F ก็พบว่า $F > F_{(0.01)(1,18)} = 8.28$ จึงสรุปว่าผลของปุ๋ยแตกต่างอย่างมีนัยสำคัญยิ่ง

เมื่อเราทดลองนำข้อมูลไปวิเคราะห์หาค่า t ตามวิธีการในตอน 4.3 ก็จะได้ผลดังนี้

1. หา Sum of square ของปุ๋ย A [SS (A)]

$$SS(A) = [(58)^2 + (56)^2 + \ldots + (57)^2] - (557)^2/10 = 74.10$$

2. หา Sum of squares ของปุ๋ย B [SS (B)]

$$SS(B) = [(64)^2 + (59)^2 + \ldots + (60)^2] - (606)^2/10 = 44.40$$

ดังนั้น

$$s_n^2 = \frac{74.10 + 44.40}{10 + 10 - 2} = 6.58$$

ซึ่งจะเห็นว่า s_p^2 ที่หาได้จะมีค่าเท่ากับ MS Error ในตาราง 9.5.2 เมื่อคำนวณค่า t โดยใช้สมการ (4-12) โดยที่ค่าเฉลี่ยของปุ๋ย A และ B เท่ากับ 55.7 และ 60.6 ซม. ตามลำดับ ดังนั้น

$$t = \frac{55.7-60.6}{\sqrt{\frac{(2)(6.58)}{10}}} = \frac{-4.9}{1.147} = -4.272$$

เมื่อนำค่า t ที่คำนวณได้มายกกำลังสองก็จะเท่ากับว่า F ในตาราง 9.5.2 คือ $(-4.272)^2 = 18.24$ จึงสรุปได้ว่า การทดลองเปรียบเทียบ 2 ทรีตเมนต์ โดยใช้ CRD นั้น F จะเท่ากับ t^2 นั่นเอง หรือเมื่อเปิดตาราง F ในระดับความแตกต่างหนึ่ง ๆ ที่มี df ของทรีตเมนต์เท่ากับ 1 แล้ว ในระดับความแตกต่างเดียวกันนั้นก็ให้ค่า $\sqrt{F = t}$

9.6 แบบจำลองทางคณิตศาสตร์และค่าคาดหมายของวาเรียนซ์

แบบจำลองทางคณิตศาสตร์หรือโมเดลของค่าสังเกตทางสถิติ คือสมการที่แสดงส่วนประกอบย่อย ๆ ของค่านั้น ๆ ทั้งนี้เพราะข้อมูลที่เราบันทึกแต่ละค่าเกิดจากผลรวมของส่วนย่อย ๆ อย่างน้อย 2 ส่วน คือ ค่าเฉลี่ย (μ_i) และความคลาดเคลื่อน (ε)[7] ทั้งนี้ ε มีการกระจายแบบปกติมีค่าเฉลี่ยเท่ากับศูนย์และวาเรียนซ์เท่ากับ σ^2 [ซึ่งอาจเขียนว่า $\varepsilon_{ij} \sim \text{NID}(0, \sigma^2)$] ดังนั้นอาจเขียนได้ว่าค่าสังเกตแต่ละค่า (X_{ij}) มีส่วนประกอบย่อย ๆ ดังนี้คือ

$$X_{ij} = \mu_i + \varepsilon_{ij} \qquad \ldots(9\text{-}6)$$

ถ้าเป็นการทดลองแบบ CRD ก็อาจแยกต่อไปได้ว่า μ_i มีส่วนประกอบอีก 2 ส่วน คือค่าเฉลี่ยของประชากร (μ) และผลของทรีตเมนต์ (α_i)[8] ดังนั้นแบบจำลองทางคณิตศาสตร์ของข้อมูลจากแผนการทดลองแบบ CRD ได้แก่

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad \ldots(9\text{-}7)$$

ในสมการนี้เราให้ $i = 1, 2, 3, \ldots, k$ ซึ่งให้ k เป็นจำนวนทรีตเมนต์, $j = 1, 2, 3, \ldots, n$ ซึ่ง n เป็นจำนวนค่าสังเกตในแต่ละทรีตเมนต์, α_i = ผลของทรีตเมนต์ และให้ ε (epsilon) เป็นความคลาดเคลื่อนในการทดลอง จากสมการ (9-6) และ (9-7)หาได้ว่า

$$\alpha_i = \mu_i - \mu \qquad \ldots(9\text{-}8)$$

คือผลของทรีตเมนต์ที่ i เกิดจากค่าเฉลี่ยของทรีตเมนต์ที่ i ลบด้วยค่าเฉลี่ยของประชากร

เมื่อทราบแบบจำลองทางคณิตศาสตร์ของค่าสังเกตแล้ว ก็อาจใช้เป็นแนวกำหนดเพื่อหาค่าคาดหมายของวาเรียนซ์หรือ EMS[9] ต่อไป อย่างไรก็ดี ค่าคาดหมายของวาเรียนซ์มักแตกต่างกันตามธรรมชาติของปัญหา (ทรีตเมนต์) ที่ศึกษา ซึ่งธรรมชาติของปัญหาที่ศึกษามีได้ 2 แบบ คือ

(1) ปัญหาเป็นปัจจัยคงที่[10] ปัญหาเป็นปัจจัยคงที่ ถ้าหากว่าในการทดลองนั้นเราตั้งใจจะศึกษาเกี่ยวกับปัญหาใดปัญหาหนึ่งโดยเฉพาะ ผลที่ศึกษาจะใช้กับปัญหานั้นเท่านั้น เช่น ศึกษาเกี่ยวกับ

เครื่องจักรชนิดนั้น พืชพันธุ์นั้น ลูกผสมของพืชชนิดนั้น ฯลฯ และเมื่อเราจะทดลองครั้งต่อไปซ้ำอีกก็ยังคงสามารถใช้ทรีตเมนต์เดิมได้ เราไม่มีวัตถุประสงค์ที่จะให้ข้อสรุปครอบคลุมปัญหาในแบบเดียวกันที่ไม่ได้ศึกษา ปัญหาเช่นนี้เรียกว่าปัจจัยคงที่

(2) ปัญหาที่เป็นปัจจัยสุ่ม[11] ปัญหาที่เป็นปัจจัยสุ่ม ถ้าหากว่าเราสุ่มมาศึกษาในฐานะเป็นตัวแทนของประชากร เช่นเรานำมาทดลองเพียงไม่กี่ทรีตเมนต์จากประชากรที่มีสมาชิกจำนวนมากมาย ชุดที่นำมาศึกษาจึงเป็นเพียงตัวอย่างจากการสุ่ม[12] เมื่อเราศึกษาแล้วก็สามารถสรุปผลได้ทั่วไป คือครอบคลุมถึงประชากรอันเป็นที่มาของตัวอย่างด้วย และถ้าเราจะศึกษาทรีตเมนต์ชนิดนี้อีกครั้งก็อาจสุ่มได้ตัวแทนชุดใหม่เข้ามา ปัญหาเช่นนี้จัดเป็นปัจจัยสุ่ม

แบบจำลองของปัญหาที่เป็นปัจจัยคงที่อาจเรียกว่า **โมเดลที่ 1 (model I)** และแบบจำลองของปัญหาที่เป็นปัจจัยสุ่มเรียกว่า **โมเดลที่ 2 (model II)** ในการทดลองเราต้องทราบว่าปัญหาเป็นปัจจัยแบบใด เพราะแต่ละแบบมีการตั้งสมมุติฐานต่างกัน และมีวิธีการทดสอบผลการวิเคราะห์ที่ต่างกันด้วย ทั้งนี้แบบจำลองหรือโมเดลของ CRD แสดงไว้ในสมการ (9-7) ถึงแม้แบบจำลองทั้งสองจะมีสมการที่เหมือนกัน แต่มีความแตกต่างในเรื่องข้อกำหนดและสมมุติฐาน ดังนี้

(1) ในกรณีของปัญหาเป็นปัจจัยคงที่ (โมเดลที่ 1) มีข้อกำหนดว่า $\sum \alpha_i = 0$ และ ε กระจายแบบปกติ โดยมีพารามิเตอร์ $\mu = 0$ วาเรียนซ์เท่ากับ σ^2 และมีสมมุติฐานในการทดสอบว่า

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = ... = \alpha_k = 0$$

$$H_1 : \alpha \text{ บางค่าไม่เป็นศูนย์}$$

ทั้งนี้ในโมเดลนี้ ได้กำหนดผลของทรีตเมนต์ (d_i) ค่าสูงไว้เป็นบวก ค่าต่ำเป็นลบ เช่น $\alpha_1 = 2, \alpha_2 = 1, \alpha_3 = 0$ $\alpha_4 = -1$ และ $\alpha_5 = -2$ ดังนั้น $\sum \alpha_i = 0$

(2) ในกรณีของปัญหาที่เป็นปัจจัยสุ่ม (โมเดลที่ 2) ในแบบจำลองนี้ มีข้อกำหนดว่า $\alpha_i \sim NID^{(13)}\ (0, \sigma_\alpha^2)$ และ $\varepsilon_{ij} \sim NID(0, \sigma^2)$ และมีสมมุติฐานในการทดสอบว่า

$$H_0 : \sigma_\alpha^2 = 0$$

$$H_1 : \sigma_\alpha^2 \neq 0$$

วิธีการหาค่าคาดหมายของวาเรียนซ์[14]

ค่าคาดหมายของวาเรียนซ์หรือเรียกย่อ ๆ ว่า EMS คือค่าที่แสดงส่วนประกอบย่อย ๆ ของความแปรปรวนชนิดหนึ่ง ๆ ในการวิเคราะห์วาเรียนซ์เราจำเป็นต้องทราบ EMS ทั้งนี้เพื่อนำไปใช้ในการทดสอบสมมุติฐานต่อไป

(1) การหา EMS ของทรีตเมนต์เมื่อเป็นปัจจัยคงที่ จากสมการ (9-7) เมื่อรวมเฉพาะค่าสังเกตในทรีตเมนต์ใดทรีตเมนต์หนึ่ง เช่น ทรีตเมนต์ที่ i ก็ได้ผลดังนี้

$$\sum_j X_j = n\mu + n\alpha_i + \sum_j \varepsilon_{ij}$$

สมการนี้แสดงว่าในทรีตเมนต์ i นั้น มีการรวมกัน j ค่า และ j มีค่าตั้งแต่ 1 ถึง n ดังนั้นค่าเฉลี่ยของทรีตเมนต์ที่ i ก็คือผลรวมข้างบนหารด้วย n ซึ่งได้ผลดังนี้

$$\bar{X}_{i\cdot} = \mu + \alpha_i + \bar{\varepsilon}_{i\cdot} \qquad \ldots(9\text{-}9)$$

และจากสมการ (9-7) ถ้าหาผลรวมทั้งหมดในทุกแถว (i) และทุกสดมภ์ (j) ก็ได้

$$\sum_i \sum_j X_{ij} = kn\mu + \sum_i n\alpha_i + \sum_i \sum_j \varepsilon_{ij}$$

ซึ่งทราบแล้วว่า $\sum_i \alpha_i = 0$ ดังนั้นค่าเฉลี่ยของทั้งหมดก็คือค่านี้หารด้วย kn ซึ่งได้

$$\bar{X}_{\cdot\cdot} = \mu + \bar{\varepsilon}_{\cdot\cdot} \qquad \ldots(9\text{-}10)$$

จากบทที่ 8 ทราบแล้วว่า $SSTr = \sum_i n(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$ ดังนั้นเมื่อแทนค่า $\bar{X}_{i\cdot}$ และ $\bar{X}_{\cdot\cdot}$ ด้วยค่าทางขวามือของสมการ (9-9) และ (9-10) ก็จะได้

$$SSTr = \sum_i n(\mu + \alpha_i + \bar{\varepsilon}_{i\cdot} - \mu - \bar{\varepsilon})^2$$

$$SSTr = \sum_i n[\alpha_i + (\bar{\varepsilon}i - \bar{\varepsilon}_{\cdot\cdot})]^2$$

ในสมการนี้ α_i และ $(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})$ เป็นอิสระแก่กัน จึงหาได้ว่า

$$SSTr = \sum_i n\,\alpha_i^2 + \sum_i n(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot})^2$$

$$= \sum_i n\,\alpha_i^2 + (k-1)\,\sigma^2$$

ดังนั้น

$$E(MSTr) = [\sum_i n\,\alpha_i^2/(k-1)] + \sigma^2 \qquad \ldots(9\text{-}11)$$

$$= nK_a^2 + \sigma^2$$

ทั้งนี้ให้ K_a^2 เกิดจากความแปรปรวนเนื่องจากผลของทรีตเมนต์

(2) การหา EMS ของทรีตเมนต์เมื่อเป็นปัจจัยสุ่ม ในกรณีที่ทรีตเมนต์เป็นปัจจัยสุ่มมีวิธีการหา EMS แตกต่างจากข้างบนเพียงเล็กน้อย ทั้งนี้เพราะ $\sum_i \alpha_i \neq 0$ ดังนั้นค่าเฉลี่ยทั้งหมดก็คล้ายคลึงกับสมการ (9-9) ดังนี้

$$\bar{X}_{\cdot\cdot} = \mu + \bar{\alpha}_{\cdot} + \bar{\varepsilon}_{\cdot\cdot} \qquad \ldots(9\text{-}12)$$

และเมื่อนำสมการ (9-9) และ (9-12) มาแทนค่าในสมการ $SSTr = \sum_i n(\bar{X}_{i.} - \bar{X}_{..})^2$ ก็จะได้

$$SSTr = \sum_i n\left[(\alpha_i - \bar{\alpha}_.) + (\bar{\varepsilon}_{i.} - \bar{\varepsilon}_{..})\right]^2$$

$$= \sum_i n(\alpha_i - \bar{\alpha}_.)^2 + n\,(\bar{\varepsilon}_{i.} - \bar{\varepsilon}_{..})^2$$

$$= (k-1)n\sigma_a^2 + (k-1)\sigma^2 \qquad \ldots(9\text{-}13)$$

ดังนั้นเมื่อหารสมการ $(9-13)$ ด้วย $k-1$ ก็ได้

$$E(MSTr) = n\sigma_a^2 + \sigma^2$$

ทั้งนี้ให้ σ_a^2 เป็นวาเรียนซ์ของผลของทรีตเมนต์

(3) การหา EMS ของ error ไม่ว่าทรีตเมนต์เป็นปัจจัยคงที่หรือเป็นปัจจัยสุ่ม EMS error เป็นค่าเดียวกัน ดังนี้ จากค่าที่สองสมการ (8-1)

$$SSE = \sum_i \sum_j (X_{ij} - \bar{X}_{i.})^2 \qquad \ldots(9\text{-}14)$$

เมื่อแทนค่า X_{ij} ด้วยสมการ (9-7) และ $\bar{X}_i$ ด้วยสมการ (9-9) ก็ได้

$$SSE = \sum_i \sum_j (\mu + \alpha_i + \varepsilon_{ij} - \mu - \alpha_i - \bar{\varepsilon}_{i.})^2$$

$$= \sum_i \sum_j (\varepsilon_{ij} - \bar{\varepsilon}_{i.})^2$$

$$= \sum_i (n-1)\sigma^2 \qquad \ldots(9\text{-}15)$$

ซึ่ง $(n - 1)\sigma^2$ คือ sum of squares ของความคลาดเคลื่อนของแต่ละกลุ่มหรือทรีตเมนต์ เมื่อรวมค่านี้ของทุกกลุ่มเข้าด้วยกัน ก็จะได้ดังสมการข้างบน ซึ่งอาจเขียนเสียใหม่ว่า

$$SSE = k(n-1)\sigma^2$$

ดังนั้น

$$E(MSE) = k(n-1)\sigma^2 / k(n-1)$$

$$= \sigma^2 \qquad \ldots(9\text{-}16)$$

จากสมการข้างบนจะเห็นได้ว่า ทรีตเมนต์ไม่มีผลกระทบกระเทือนต่อค่าประมาณ σ^2 ดังนั้นไม่ว่าเรายอมรับหรือปฏิเสธสมมุติฐานก็ตาม MSE ก็เป็นค่าประมาณที่ไม่ลำเอียงของ σ^2 เสมอ

เมื่อนำค่าประมาณเหล่านี้ลงตาราง ก็จะได้ผลดังตาราง 9.6.1 ส่วนการทดสอบสมมุติฐานของข้อมูลในโมเดลที่ 1 และที่ 2 นั้น ปรากฏตามลำดับดังนี้

$$F = \frac{nK_a^2 + \sigma^2}{\sigma^2} \quad \text{และ} \quad F = \frac{n\sigma_a^2 + \sigma^2}{\sigma^2}$$

ซึ่งในทางปฏิบัติเราทดสอบโดยใช้อัตราส่วน $F = MSTr/MSE$ นั่นเอง

ตาราง 9.6.1 ค่า EMS ในแบบจำลองปัจจัยคงที่ (Model I) และแบบจำลองปัจจัยสุ่ม (Model II)

Sources	df	Expected Mean Squares (EMS)[1]	
		Model I	Model II
Treatments	$k-1$	$\sigma^2 + nK_t^2$	$\sigma^2 + n\sigma_t^2$
Error	$k(n-1)$	σ^2	σ^2
Total	$kn-1$		

[1] $nK_t^2 = \dfrac{\sum n\alpha_i^2}{k-1}$

9.7 การวิเคราะห์วาเรียนซ์เมื่อมีการสังเกตแปลงย่อย

ในการทดลองบางครั้ง ผู้ทดลองอาจเก็บข้อมูลลักษณะเดียวกันมาหลาย ๆ ค่า ข้อมูลเหล่านี้เรียกว่า ตัวอย่างย่อย[15] เช่น ในการทดลองเปรียบเทียบปุ๋ย 3 ระดับ (A, B, C) กับไม้ผล แต่ละระดับทดลอง 3 แปลง แต่ละแปลงเก็บผลผลิตมา 3 ต้น ในการเก็บผลการทดลอง ผลผลิตแต่ละต้นถือเป็น 1 ตัวอย่าง โดยปกติเรานำข้อมูลเหล่านี้มารวมกันและวิเคราะห์วาเรียนซ์เหมือนวิธีที่กล่าวมาแล้ว แต่เพื่อความแน่นอนอาจวิเคราะห์ โดยใช้ค่าสังเกตย่อย ๆ เช่น จากการใช้ปุ๋ยกับไม้ผลแต่ละแปลง เก็บผลผลิตมา 3 ต้น ดังตาราง 9.7.1

ในกรณี ของการสังเกตที่มีตัวอย่างย่อยแบบจำลองทางคณิตศาสตร์หรือโมเดลเป็น ดังนี้

$$X_{ijk} = \mu + \alpha_i + \varepsilon_{ij} + \delta_{ijk} \qquad \ldots(9\text{-}17)$$

เมื่อ $i = 1, 2, \ldots, a$; $j = 1, 2, \ldots, b$; $k = 1, 2, \ldots, n$ เมื่อ a, b และ n เป็นจำนวนทรีตเมนต์, จำนวนแปลงทดลองหรือซ้ำ และจำนวนแปลงย่อยตามลำดับ

ทั้งนี้ให้ μ เป็นค่าเฉลี่ยของประชากร, α_i เป็นผลของทรีตเมนต์, ε_{ij} เป็นความคลาดเคลื่อนเนื่องจากทรีตเมนต์ที่ i ทดลองในหน่วยทดลองที่ j ทั้งนี้ $\varepsilon_{ij} \sim NID(0, \sigma_\varepsilon^2)$ และ δ_{ijk} เป็นความคลาดเคลื่อนเนื่องจากทรีตเมนต์ที่ i ในหน่วยทดลองที่ j และในตัวอย่างย่อยที่ k ทั้งนี้ $\delta_{ijk} \sim NID(0, \sigma_\delta^2)$

ตาราง 9.7.1 การทดลองที่มีค่าสังเกตแปลงย่อย

ปุ๋ย	ผลผลิต (กก. / ต้น)			รวม
	แปลงที่ 1	แปลงที่ 2	แปลงที่ 3	
A	5 , 3 , 4	4, 4, 3	5, 6, 5	39
B	7, 5, 5	6, 7, 6	4, 6, 6	52
C	4, 4, 3	5, 5, 3	3, 4, 5	36
รวม				127

ขั้นตอนการวิเคราะห์มีดังนี้

$$CF = \frac{\sum (X_{ijk})^2}{abn} = \frac{(127)^2}{27} = 597.37$$

เมื่อ i = 1, 2, ..., a (a = จำนวนทรีตเมนต์) ; j = 1, 2, ..., b (b = จำนวนแปลงทดลองหรือซ้ำ) และ k = 1, 2, ..., n (n = จำนวนตัวอย่างย่อย)

$$\begin{aligned} TSS &= \sum X_{ijk}^2 - CF \\ &= 5^2 + 3^2 + 4^2 + 4^2 + \ldots + 3^2 + 4^2 + 5^2 - 597.37 \\ &= 635 - 597.37 = 37.63 \end{aligned}$$

$$\begin{aligned} SSTr &= \frac{T_1^2 + T_2^2 + T_3^2}{bn} - CF \\ &= \frac{(39^2 + 52^2 + 36^2)}{9} - 597.37 \\ &= 613.44 - 597.37 = 16.07 \end{aligned}$$

SS Experimental error (SS ex. error)

$$= \frac{(\text{ผลรวมแต่ละแปลงย่อย})^2}{n} - CF - SSTr$$

$$= \frac{(5+3+4)^2 + (7+5+5)^2 + \ldots + (3+4+5)^2}{3} - 597.37 - 16.07$$

$$= 620.33 - 597.37 - 16.07 = 6.89$$

SS Sampling error = TSS - SSTr - SS Exp. error

$$= 37.63 - 16.07 - 6.89 = 14.67$$

แล้วนำค่าที่คำนวณได้ลงตารางวิเคราะห์วาเรียนซ์ ดังตาราง 9.7.2 และแสดงการหา df ไว้ในตารางด้วยแล้ว

ตาราง 9.7.2 ผลการวิเคราะห์วาเรียนซ์ของข้อมูลในตาราง 9.7.1

Sources	df	SS	MS	F
Treatments	$a-1=2$	16.07	8.04	9.80**
Exp. Error	$a(b-1)=6$	6.89	1.15	1.40^{ns}
Sampling error	$ab(n-1)=18$	14.67	0.82	
Total	$anb-1=26$	37.63		

** = แตกต่างทางสถิติในระดับ 1 เปอร์เซ็นต์ , ns = not significant

ตาราง 9.7.3 ค่า EMS แบบจำลองของปัจจัยคงที่ (Model I) ปัจจัยสุ่ม (Model II) และปัจจัยผสม (Model III)

Sources	MSE	Model I	Model II	Model III
Treatments	M_3	$\sigma^2 + bnK_a^2$	$\sigma^2 + n\sigma_b^2 + bn\sigma_a^2$	$\sigma^2 + n\sigma_b^2 + bnK_a^2$
Exp. Error	M_2	$\sigma^2 + nK_b^2$	$\sigma^2 + n\sigma_b^2$	$\sigma^2 + n\sigma_b^2$
Sampling error	M_1	σ^2	σ^2	σ^2

การทดสอบ การทดลองเกี่ยวกับปัจจัยคงที่ ใช้ M_1 หาร M_3 และ M_2 ถ้าเป็นปัจจัยสุ่ม เริ่มจาก M_2/M_1 ถ้าไม่แตกต่างก็ใช้ M_2 หรือ M_1 ก็ได้ หาร M_3 เพื่อประเมินผลของทรีตเมนต์ ถ้า M_2/M_1 แตกต่างก็ใช้ M_2 ทดสอบ M_1 ต่อไป ในปัจจัยแบบผสมวิธีการทดสอบคล้ายกับการทดสอบปัจจัยสุ่มนั่นเอง

แผนการทดลองแบบเนสต์[16]

การศึกษาบางอย่างคล้ายกับการทดลองที่มีตัวอย่างย่อยมาก แต่มีวัตถุประสงค์ต่างกัน เช่น เราต้องการที่จะซื้อสินค้าชนิดหนึ่งจากบริษัทต่าง ๆ a บริษัท แต่ละบริษัทผลิตสินค้าส่งมาเป็นงวด ๆ จำนวน b งวด แต่ละงวดสุ่มมา n ตัวอย่าง การทดลองแบบนี้ เรียกว่าแผนการทดลองแบบเนสต์ หรือการจัดกลุ่มแบบจัดลำดับชั้น เป้าหมายของการทดสอบคือ ต้องการทราบว่า ความแปรปรวนหรือความแตกต่างต่างๆ เกิดขึ้นในขั้นใด เมื่อตรวจพบแล้วก็สามารถเลือกจากบริษัทที่ดีที่สุด หรือแนะนำให้ผู้ผลิตปรับปรุงในขั้นตอนใดเพื่อลดความแปรปรวน การทดลองมีแบบจำลองทางคณิตศาสตร์ ดังนี้

$$X_{ijk} = \mu + A_i + B_{j(i)} + \varepsilon_{ijk}$$

ทั้งนี้ I = 1, 2 ,..., a ; j = 1, 2,..., b;k = 1, 2 ,..., n และให้ A เป็นบริษัท, B เป็นงวดสินค้า $A_i, B_{j(i)}$, และ ε_{ijk} ต่างก็เป็นอิสระแก่กัน

ตาราง 9.7.4 คะแนนคุณภาพสินค้าแต่ละงวดที่ผลิตโดยบริษัทผลิตสินค้าต่าง ๆ

บริษัท	งวดสินค้า				รวม	รวม	รวม
A	1	5	3	4	12		
	2	4	4	3	11		
	3	5	6	5	16	39	
B	1	7	5	5	17		
	2	6	7	6	19		
	3	4	6	6	16	52	
C	1	4	4	3	11		
	2	5	5	3	13		
	3	3	4	5	12	36	127

จากตาราง 9.7.4 อาจวิเคราะห์วาเรียนซ์ได้ดังนี้

$$\text{Correction factor(CF)} = \frac{\left(\sum X_{ijk}\right)^2}{abn} = \frac{127^2}{27} = 597.37$$

$$\text{Total SS(TSS)} = 5^2 + 3^2 + \ldots\ldots + 5^2 - 597.37$$

$$= 635 - 597.37 = 37.63$$

$$\text{SS(Companies)} = \frac{39^2 + 52^2 + 36^2}{(3)(3)} - 597.37$$

$$= 613.44 - 597.37 = 16.07$$

ต่อไปก็ทำการวิเคราะห์คุณภาพสินค้าแต่ละงวดภายในบริษัท โดยแยกวิเคราะห์เป็นบริษัท ดังนี้

$$\text{SS Company A} = \frac{12^2 + 11^2 + 16^2}{3} - \frac{39^2}{9} = 4.67$$

$$\text{SS Company B} = \frac{17^2 + 19^2 + 16^2}{3} - \frac{52^2}{9} = 1.55$$

$$\text{SS Company C} = \frac{11^2 + 13^2 + 12^2}{3} - \frac{36^2}{9} = 0.67$$

ผลรวมจากทุกบริษัท = 4.67 + 1.55 + 0.67 = 6.89

ดังนั้นอาจหา error sum of squares (SSE) ดังนี้

$$SSE = 37.63 - 16.07 - 6.89 = 14.67$$

แล้วนำผลลงตารางดังตาราง 9.7.5 ดังนี้

ตาราง 9.7.4 ผลการวิเคราะห์วาเรียนซ์ของข้อมูลในตาราง 9.7.4

Sources	df	SS	MS	EMS(Model II)
Companies (A)	$a - 1 = 2$	16.07	8.04	$\sigma^2 + n\sigma_B^2 + bn\sigma_A^2$
Batches/companies (B/A)	$a(b-1) = 6$	6.89	1.15	$\sigma^2 + n\sigma_B^2$
Samples/Batch/Company	$Ab(n-1) = 18$	14.67	0.82	σ^2
Total	$Abn - 1 = 26$			

การทดสอบสมมุติฐาน

$$\sigma_A^2 = 0; \quad F = \frac{8.04}{1.15} = 6.99^* \quad \text{(เปรียบเทียบกับ } F_{0.01} \text{ df 2, 6)}$$

$$\sigma_B^2 = 0; \quad F = \frac{1.15}{0.82} = 1.40^{ns} \quad \text{(เปรียบเทียบกับ } F_{0.05} \text{ df 6, 18)}$$

ผลการทดสอบนี้คุณภาพสินค้าปรวนแปรหรือแตกต่างระหว่างบริษัท แต่ไม่มีความปรวนแปรภายในบริษัท

9.8 แบบฝึกหัด

1. จงแสดงสมการทางคณิตศาสตร์ค่าต่างๆ ที่ได้รับจากการทดลอง โดยใช้แผนการทดลอง CRD ดังนี้

ก. X_{ij} ค. $\overline{X}_{i.}$ จ. T_i

ข. $\sum X_{ij}$ ง. $\overline{X}_{..}$

2. เมื่อใดจึงควรใช้แผนการทดลอง CRD อธิบายพร้อมยกตัวอย่างการทดลองที่ควรใช้แผนการทดลองชนิดนี้

3. สมมุติว่ามีตำราทางสถิติเบื้องต้นอยู่ 3 ชนิด แต่ละชนิดมีรายละเอียดและเนื้อหาคล้ายคลึงกัน แต่มีวิธีการอธิบายปัญหาแตกต่างกัน ถ้าเรามีนักศึกษาอยู่ 18 คน และต้องการที่จะทดลองเปรียบเทียบผลของการใช้ตำราทั้ง 3 ชนิดนี้ ท่านคิดว่าเราควรทำการทดลองอย่างไร จงอธิบายถึงแผนการทดลองที่จะใช้ ทำไมจึงใช้แผนการทดลองดังกล่าว และจงอธิบายถึงวิธีการสุ่มเพื่อจัดกลุ่มนักศึกษา

4. จากการทดลองเปรียบเทียบผลผลิตพืช 3 พันธุ์ แต่ละพันธุ์ปลูก 4 แปลง ปรากฏว่าได้ผลดังนี้

พันธุ์ A	74	78	68	72
พันธุ์ B	93	83	89	87
พันธุ์ C	75	80	79	82

ก. จงแสดงแบบจำลองทางคณิตศาสตร์

ข. จงกล่าวถึงข้อกำหนดที่จำเป็นสำหรับการวิเคราะห์วาเรียนซ์และจงตั้งสมมุติฐาน

ค. จงวิเคราะห์วาเรียนซ์และทดสอบสมมุติฐานว่า ค่าเฉลี่ยของผลผลิตทั้ง 3 พันธุ์ ไม่แตกต่างกัน

5. จากการทดลองเปรียบเทียบผลผลิตพันธุ์ข้าว 7 พันธุ์ ปรากฏว่าได้ผลผลิต (กก./แปลง) ดังนี้

พันธุ์	ซ้ำที่ 1	ซ้ำที่ 2	ซ้ำที่ 3	ซ้ำที่ 4
A	2.13	1.97	2.87	2.73
B	2.77	1.59	2.87	2.90
C	2.20	2.21	2.20	2.69
D	2.83	2.07	2.60	3.41
E	2.51	2.01	2.49	2.98
F	2.81	1.94	3.15	2.87
G	3.27	2.69	3.42	3.56

ก. จงแสดงแบบจำลองทางคณิตศาสตร์

ข. จงบอกสมมุติฐาน

ค. วิเคราะห์วาเรียนซ์และทดสอบว่าพันธุ์ข้าวไม่แตกต่างกัน

6. ในการทดลองเปรียบเทียบวาเรียนซ์พันธุ์มะเขือเทศ 4 พันธุ์นั้น จำนวน 4 ครั้ง ในการทดลองพบว่า บางพันธุ์ในบางครั้งไม่สามารถเก็บเกี่ยวผลผลิตได้ ซึ่งได้ผลผลิตเป็น กก./แปลง ดังนี้

พันธุ์	พันธุ์			
	1	2	3	4
A	18	13	12	16
B	20	23	21	-
C	14	-	9	10
D	11	17	-	16

จงทดสอบในระดับความแตกต่าง 0.05 ว่าพันธุ์เหล่านี้ให้ผลผลิตเท่ากัน

คำในบท

(1) completely randomized design , (2) single factor , (3) control , (4) standard error, (5) standard error of mean , (6) standard error of difference , (7) experimental error , (8) treatment effect , (9) expected mean square (EMS) , (10) fixed effect , (11) random effect , (12) random sample , (13) NID = normally independently distributed , (14) expected mean square = EMS, (15) sub-sample, (16) nested design

บทที่ 10
แผนการทดลองแบบสุ่มภายในบลอก

10.1 คำนำ

แผนการทดลองแบบ CRD ที่กล่าวมาแล้ว ไม่เหมาะสำหรับการทดลองที่มีขนาดใหญ่เพราะการทดลองที่มีขนาดใหญ่จำเป็นต้องใช้หน่วยทดลอง[1] เป็นจำนวนมาก เมื่อเป็นเช่นนี้ก็จำเป็นอยู่เองที่ต้องใช้หน่วยทดลองที่มีความแตกต่างกัน เช่น ถ้าเป็นการทดลองเกี่ยวกับพืช ก็ต้องใช้พื้นที่ใหญ่ขึ้นเป็นต้น ซึ่งเป็นการเพิ่มความคลาดเคลื่อนและลดความเที่ยงตรงในการทดลองลงมา ทั้งนี้เพราะพื้นที่ยิ่งกว้าง อัตราความแตกต่างของดินในเรื่องความอุดมสมบูรณ์ของดิน ธาตุอาหารในดิน ความลาดเทของพื้นที่ ฯลฯ ย่อมมีมากกว่าพื้นที่ขนาดเล็ก วิธีการแก้ไขที่เป็นไปได้ก็คือการปรับปรุงวิธีการทดลองจาก CRD โดยแยกหน่วยทดลองออกเป็นกลุ่ม ๆ พวกที่เหมือนกันให้อยู่ในกลุ่มเดียวกัน พวกที่ต่างกันก็ให้อยู่คนละกลุ่ม เมื่อแยกแล้วก็จะมีความแตกต่างระหว่างกลุ่มที่ชัดเจน ต่อจากนั้นทำการทดลองทุกทรีตเมนต์ในแต่ละกลุ่ม กลุ่มดังกล่าวนั้นเรียกว่าบลอก[2] ดังนั้น บลอกก็คือกลุ่มของหน่วยทดลองที่มีลักษณะเหมือน ๆ กันนั่นเอง แผนการทดลองที่จัดบลอกเช่นนี้เรียกว่า แผนการทดลองแบบสุ่มภายในบลอก[3] ซึ่งต่อไปนี้เรียกย่อ ๆ ว่า แผนการทดลอง แบบ RCB

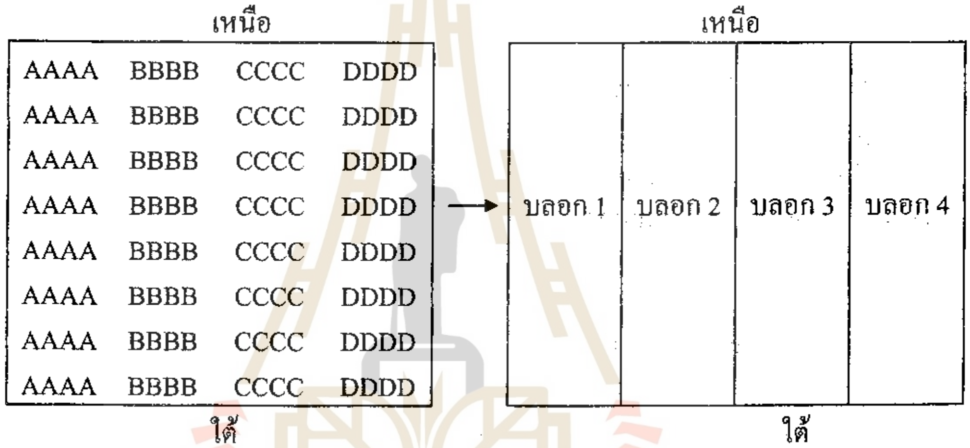
10.2 เทคนิคในการจัดบลอก

วัตถุประสงค์ของการจัดกลุ่มของหน่วยทดลอง ก็เพื่อที่จะลดความคลาดเคลื่อนของการทดลองอันเนื่องจากความไม่สม่ำเสมอของหน่วยทดลองนั่นเอง ในการทดลองเกี่ยวกับพืช เช่น การเปรียบเทียบพันธุ์พืชนั้น หน่วยทดลองคือแปลงปลูกพืช เราแบ่งแปลงปลูกออกเป็นบลอก ภายในบลอกให้ได้แปลงที่มีลักษณะเหมือนกัน คือดินมีความอุดมสมบูรณ์หรือความลาดเทใกล้เคียงกัน เช่น ถ้าหากความอุดมสมบูรณ์ของดินมีการเรียงระดับความสูงต่ำอยู่ในแนวทิศตะวันตก-ตะวันออก บลอกก็ควรตัดเรียงในแนวทิศเหนือ-ทิศใต้ ดังรูปที่ 10.2.1 แต่ถ้าเราไม่ทราบระดับความอุดมสมบูรณ์ของดินในจุดต่าง ๆ ก็ควรให้บลอกเป็นรูปสี่เหลี่ยมจตุรัสหรือใกล้เคียงสี่เหลี่ยมจตุรัสเป็นดีที่สุด ทั้งนี้เพราะเชื่อว่าแปลงที่มีรูปร่างเช่นนี้จะมีความสม่ำเสมอภายในแปลงดีที่สุด

ในการทดลองเกี่ยวกับสัตว์ เช่น ทดสอบการให้อาหารสุกร หน่วยทดลองคือสุกร ย่อมเป็นไปได้ว่าสุกรที่ใช้มีทั้งเพศและอายุแตกต่างกัน เพศและอายุที่แตกต่างกันนี้อาจจะสนองตอบต่อการให้อาหารแตกต่างกัน ดังนั้นในบลอกหนึ่งอาจประกอบด้วยเพศผู้จากครอกเดียวกัน เพศเมียจากครอก

นั้นอาจอยู่ในอีกบลอกหนึ่ง ถ้าจัดสองบลอกยังไม่พอก็นำครอกอื่น ๆ มาทำเช่นเดียวกัน เพื่อเพิ่มจำนวนบลอกให้มากขึ้น เช่นนี้จะเห็นได้ว่าหน่วยทดลองในแต่ละบลอกมีความสม่ำเสมอมาก แต่มีความแตกต่างกันระหว่างบลอก

ในการทดลองแบบอื่น ๆ ก็มีเทคนิคการจัดบลอกที่คล้ายคลึงกัน เช่น การทดลองยาแก้ปวดฟัน 4 ชนิดแก่คนไข้ คนไข้อายุต่าง ๆ กันน่าจะมีการสนองตอบต่อการใช้ยาต่างกัน ดังนั้น ในการทดลองแบ่งคนไข้ออกเป็นกลุ่ม ๆ คือ กลุ่มอายุน้อย อายุปานกลาง และอายุมาก กลุ่มอายุนี้จัดเป็นบลอกแบบหนึ่ง



รูปที่ 10.2.1 แสดงการเรียงลำดับของระดับความอุดมสมบูรณ์ของดิน สมมุติว่าดินกลุ่ม A อุดมสมบูรณ์มาก และ B, C รอง ๆ ลงไปจนถึง D เป็นดินที่อุดมสมบูรณ์น้อยที่สุด ดังนั้นถ้ามีการจัดบลอกก็จะได้ 4 บลอกตามกลุ่มของความอุดมสมบูรณ์ของดิน จะเห็นได้ว่าภายในแต่ละบลอกหน่วยทดลองมีความเหมือนกันสูง แต่ระหว่างบลอกแตกต่างกัน จะมากน้อยแล้วแต่ว่าเป็นบลอกใด

ในการทดลองโดยใช้แผนการทดลองแบบ RCB นี้ ในแต่ละบลอกมีทรีตเมนต์ต่าง ๆ อยู่ครบ แต่ละทรีตเมนต์ปรากฏหนึ่งครั้ง หรือหนึ่งแปลง การทดลองที่มีทุกทรีตเมนต์ และแต่ละทรีตเมนต์ เกิดขึ้นหนึ่งครั้งนี้เรียกว่า หนึ่งซ้ำ[4] ดังนั้นบลอกที่มีครบทุกทรีตเมนต์เรียกว่าซ้ำก็ได้ การวางทรีตเมนต์ลงในหน่วยรองรับที่อยู่ภายในบลอกกระทำโดยการสุ่ม โดยเหตุนี้เองเราเรียกแผนการทดลองนี้ว่า เป็นแผนการทดลอง **RCB หรือแบบสุ่มภายในบลอก** เช่น ในการทดลองเปรียบเทียบพันธุ์พืช 4 พันธุ์ ก็แบ่งแต่ละบลอกออกเป็น 4 แปลงย่อย เพื่อปลูกพืช 4 พันธุ์นั้น การที่จะปลูกพันธุ์ไหนลงในแปลงย่อยแปลงใดนั้นย่อมขึ้นอยู่กับผลของการสุ่ม

ในการปฏิบัติการทดลอง เราต้องให้มีความสม่ำเสมอภายในบลอกให้มากที่สุด เช่น ในการรดน้ำแปลงทดลองพืช ถ้าไม่อาจรดได้เสร็จในวันเดียว ก็อาจแบ่งรดวันละบลอก ในการใส่ปุ๋ย การปราบ วัชพืช

การให้อาหารสัตว์ และการบันทึกลักษณะต่าง ๆ ก็เช่นกัน ก็ให้พยายามทำภายในบลอกเดียวกันให้เสร็จก่อนจึงย้ายไปบลอกอื่น ๆ ต่อไป การปฏิบัติต่อการทดลองทุกทรีตเมนต์ภายในบลอกที่เหมือนกันนี้ นับว่ามีความสำคัญมาก เป็นการป้องกันไม่ให้ทรีตเมนต์แตกต่างกันเนื่องจากสาเหตุอื่น หรือพูดให้ถูกต้องก็คือเป็นการลดความคลาดเคลื่อนของการทดลองนั้น

ในแผนการทดลองแบบ CRD ซึ่งไม่มีการใช้บลอกนั้นเห็นว่า ความคลาดเคลื่อนที่เกิดขึ้นจากทุกสาเหตุปนอยู่ในความคลาดเคลื่อนของการทดลอง ดังนั้น ข้อดีของการจัดบลอกก็คือ ทำให้เราแยกความคลาดเคลื่อนส่วนหนึ่งออกมาให้เป็นเรื่องของบลอก ความคลาดเคลื่อนนี้จะไม่ปนอยู่กับความคลาดเคลื่อนในการทดลองอีกต่อไป จึงเป็นการเพิ่มความเที่ยงตรงในการทดลอง ความแตกต่างระหว่างบลอกไม่มีผลกระทบต่อความแตกต่างระหว่างค่าเฉลี่ยของทรีตเมนต์แต่อย่างใด

10.3 วิธีการสุ่มวางทรีตเมนต์

ในแผนการทดลองแบบ RCB นั้น มีการแบ่งกลุ่มหน่วยทดลองออกเป็นหลายบลอก แต่ละบลอกที่มีครบทุกทรีตเมนต์เรียกว่า 1 ซ้ำ จำนวนซ้ำในการทดลองอาจมี 2 – 6 ซ้ำ การจัดทรีตเมนต์ลงในแปลงย่อยหรือหน่วยย่อยภายในบลอกต้องกระทำโดยวิธีสุ่ม และการสุ่มต้องกระทำใหม่สำหรับทุกบลอก จะนำผลการสุ่มจากบลอกหนึ่งไปใช้กับอีกบลอกหนึ่งไม่ได้

การสุ่มและการจัดระเบียบทรีตเมนต์ในบลอกเป็นเทคนิคที่ควรทราบ ตัวอย่างเช่น ถ้าเราต้องการที่จะเปรียบเทียบพันธุ์พืช 6 พันธุ์ คือ A, B, C, D, E และ F โดยใช้แผนการทดลองแบบ RCB จำนวน 4 ซ้ำ การจัดแปลงและการสุ่มอาจดำเนินตามขั้นตอนดังนี้

(1) การแบ่งแปลงทดลอง แบ่งแปลงทดลองออกเป็น 4 บลอก แต่ละบลอกแบ่งเป็น 6 แปลงย่อย เพื่อใช้ปลูกพืช 1 ซ้ำ (6 พันธุ์) โดยปลูกพันธุ์ละ 1 แปลงย่อย ในส่วนของการเตรียมการอาจวาดแผนผังแปลงในแผ่นกระดาษก็ได้

(2) ทำการสุ่มเพื่อวางทรีตเมนต์ ก่อนสุ่มให้หมายเลขแทนพันธุ์ ดังนี้

หมายเลข:	1	2	3	4	5	6	(สดมภ์ที่ 1)
พันธุ์:	A	B	C	D	E	F	(สดมภ์ที่ 2)

ต่อไปก็สุ่มทีละบลอก สมมุติว่าเราสุ่มบลอกที่ 1 เพื่อจัดทำซ้ำที่ 1 ถ้าการสุ่มกระทำโดยใช้ตารางเลขสุ่ม ก็สุ่มเลขชุดละ 2 ค่ามา 6 ชุด คือจำนวนชุดเท่ากับจำนวนพันธุ์ สมมุติว่าได้ผลเป็นลำดับดังนี้

เลขสุ่ม: 08 96 72 45 18 42 (สดมภ์ที่ 3)

เมื่อเรียงตามลำดับจากน้อยไปหามากก็ได้

เลขสุ่ม:	08	96	72	45	18	42	(สดมภ์ที่ 3)
ลำดับที่:	1	6	5	4	2	3	(สดมภ์ที่ 4)

ทั้งนี้ลำดับในสดมภ์ที่ 3 เป็นลำดับของแปลงย่อยในบลอกนั่นเอง ถ้าลำดับนี้อยู่ในบลอกที่ 1 หรือเป็นซ้ำที่ 1 ก็ใส่เลข 1 ลงหน้าเลขในสดมภ์ที่ 3 แล้วใส่เลข 0 นำหน้าทุกค่าในสดมภ์ที่ 3 แทรกระหว่างกลางอีกค่าหนึ่งก็จะได้สดมภ์ที่ 5 ดังนี้

พันธุ์:	A	B	C	D	E	F	(สดมภ์ที่ 2)
ลำดับแปลง:	101	106	105	104	102	103	(สดมภ์ที่ 5)

การใส่เลข 0 ไว้ ก็เพื่อสำรองในกรณีที่จำนวนแปลงแต่ละบลอกมีตั้งแต่ 10 แปลงขึ้นไป ทั้งนี้เพราะอาจมีจำนวนทรีตเมนต์เกิน 10 นั่นเอง

สมมุติว่าเราทำการสุ่มทุกบลอกจนได้ครบ 4 ซ้ำก็ได้ผลเช่น ตาราง 10.3.1 ซึ่งเป็นการนำผลของบลอกที่ 2, 3 และ 4 และอื่น ๆ ซึ่งสุ่มได้ในภายหลังมารวมไว้นั่นเอง

ตาราง 10.3.1 แสดงวิธีการสุ่มเพื่อวางพันธุ์ (ทรีตเมนต์) ใน RCB จำนวน 4 ซ้ำ

เลขที่	พันธุ์	บลอก (ซ้ำ)			
ทรีตเมนต์	พืช	I	II	III	IV
1	A	101	206	303	402
2	B	106	204	305	403
3	C	105	202	304	401
4	D	104	203	302	406
5	E	102	205	301	404
6	F	103	201	306	405

จากตาราง 10.3.1 เห็นได้ว่ามีการปลูกพืชพันธุ์ A ในแปลงที่ 1, 6, 3 และ 2 ของซ้ำ (บลอก) ที่ 1, 2, 3 และ 4 ตามลำดับ ถ้าแสดงเป็นรูปหรือแผนผังการทดลองก็อาจได้ดังรูป 10.3.1 จงสังเกตว่าในแต่ละบลอกมีเลขที่ของแปลงเรียงเป็นลำดับ แต่จากผลของการสุ่มดังตาราง 10.3.1 ทำให้ลำดับของพันธุ์ในแต่ละบลอกแตกต่างกันไป

บลอก 1		2		3		4	
101	A	201	F	301	E	401	C
102	E	202	C	302	D	402	A
103	F	203	D	303	A	403	B
104	D	204	B	304	C	404	E
105	C	205	E	305	B	405	F
106	B	206	A	306	F	406	D

รูป 10.3.1 รูปแสดงแปลงทดลองทำการเปรียบเทียบพันธุ์พืช 6 พันธุ์ โดยใช้แผนการทดลอง RCB จำนวน 4 ซ้ำ

ในการสุ่มอาจกระทำโดยใช้วิธีการจับฉลากก็ได้ คือเขียนชื่อพันธุ์ลงในแผ่นกระดาษแล้วหยิบขึ้นมาทีละแผ่น ได้พันธุ์ใดก็บรรจุลงในแปลงตามลำดับก่อนหลังของการสุ่ม ซึ่งจะได้ผลคล้ายกับการใช้ตารางเลขสุ่มเช่นกัน วิธีนี้สะดวก ไม่จำเป็นต้องใช้ตารางใด ๆ เช่น ถ้าเป็นการทดลองเปรียบเทียบทรีตเมนต์ 10 ชนิด โดยใช้การทดลองขนาด 4 บลอก หรือ 4 ซ้ำ การสุ่มมี 4 ขั้นตอนดังนี้

1. เขียนหมายเลขหน่วยทดลอง หรือแปลงทดลองในแต่ละบลอกให้ครบทั้ง 4 บลอก โดยให้หมายเลข 3 หลัก เลขค่าแรก (1, 2,3 และ 4) เป็นหมายเลขบลอก ดังตารางข้างล่าง เลขทั้ง 2 หลักหลังแทนหมายเลขของหน่วยทดลองหรือแปลงทดลอง

2. เขียนชื่อทรีตเมนต์ลงในแผ่นกระดาษขนาดเท่า ๆ กัน เช่น ทรีตเมนต์ A, B, C, D, E, F, G, H, I และ J แล้วม้วนให้กลม ใส่ลงในภาชนะ เช่น ขัน หรือถ้วยปากกว้าง เขย่าหรือใช้มือคลุกเคล้าให้ปนกัน

3. ทำการสุ่มทีละบลอก เช่น สุ่มในบลอกที่ 1 หยิบฉลากอย่างสุ่มออกมาทีละใบ เช่น หยิบใบแรกขึ้นมาเปิดอ่านแล้วใส่เป็นทรีตเมนต์ของหน่วยทดลองหรือแปลงที่ 1 แล้วสุ่มต่อ ๆ ไปเพื่อนำลงในหน่วยทดลองที่ 2, 3,.... เป็นลำดับไป จนครบ 10 หน่วยทดลอง แต่ละครั้งของการสุ่มไม่ใส่ฉลากคืนในภาชนะ กระทำจนกว่าหมดบลอกที่ 1

4. กระทำเช่นนี้จนครบทุกบลอก ผลที่ได้อาจแสดงเป็นตัวอย่างในตาราง 10.3.2 ซึ่งในขั้นการทดลอง ก็นำทรีตเมนต์ต่าง ๆ ไปใช้กับหน่วยทดลองนั้น ๆ เช่น ทรีตเมนต์ A ใช้กับหน่วยทดลองที่ 8 , 10 , 4 และ 2 ในบลอกที่ 1, 2, 3, และ 4 ตามลำดับ

เพื่อความสะดวกในการดำเนินการทดลอง เราอาจจัดระเบียบจากตารางข้างบนเสียใหม่ โดยใช้ลำดับของทรีตเมนต์เป็นหลักดังนี้ ดังตาราง 10.3.3

ตาราง 10.3.2 แสดงผลการสุ่มเพื่อวางทรีตเมนต์ในหน่วยทดลอง เช่น แปลงปลูกพืชซึ่งมีการเรียงเป็นลำดับ

ลำดับหน่วยทดลอง							
บลอก 1	ทรีตเมนต์	บลอก 2	ทรีตเมนต์	บลอก 3	ทรีตเมนต์	บลอก 4	ทรีตเมนต์
101	B	201	E	301	I	401	B
102	J	202	D	302	E	402	A
103	H	203	C	303	G	403	I
104	C	204	B	304	A	404	J
105	D	205	I	305	D	405	H
106	F	206	G	306	F	406	D
107	I	207	H	307	C	407	F
108	A	208	F	308	J	408	E
109	E	209	J	309	H	409	C
110	G	210	A	310	B	410	G

ตารางที่ 10.3.3 **แสดงตำแหน่งของทรีตเมนท์ต่าง ๆ ในการทดลอง**

ทรีตเมนต์	หน่วยทดลอง			
	บลอก 1	บลอก 2	บลอก 3	บลอก 4
A	108	210	304	402
B	101	204	310	401
C	104	203	307	409
D	105	202	305	406
E	109	201	302	408
F	106	208	306	407
G	110	206	303	410
H	103	207	309	405
I	107	205	301	403
J	102	209	308	404

จากตารางดังกล่าวนี้ เห็นได้ว่าทรีตเมนต์หนึ่ง ๆ ปรากฏที่ใดบ้างในแต่ละบลอกหรือแต่ละซ้ำ

บางครั้งการสุ่มโดยใช้ฉลากอาจทำในวิธีกลับกันกับข้างบน เช่น ถ้ามี 10 ทรีตเมนต์ก็เขียนเลข 1, 2, 3,...จนถึง 10 ลงในแผ่นกระดาษแล้วสุ่มทีละบลอก เช่น ในบลอกแรกเราหยิบได้เลข 4, 1, 8, 10, 5, 9, 2, 6, 3 และ 7 แล้วใส่เลข 1 และ 0 นำหน้าก็จะได้แปลงทดลองของบลอกแรกสำหรับทรีตเมนต์ A, B, C, D, ... และ J ตามลำดับ และบลอกอื่น ๆ ก็สุ่มวิธีเดียวกันได้ผล ดังตาราง 10.3.4

10.4 การวิเคราะห์วาเรียนซ์

เมื่อมีการใช้บลอกในการทดลอง ก็สามารถแยก sum of squares ที่มีอยู่ทั้งหมด (TSS) ออกเป็นส่วน ๆ ได้ดังนี้

$$TSS = SSB + SSTr + SSE \qquad \ldots(10\text{-}1)$$

เมื่อ

SSB = Block sum of squares

SSTr = Treatment sum of squares

SSE = Error sum of squares

ตาราง 10.3.4 แสดงผลของการสุ่มโดยใช้ฉลาก

ทรีตเมนต์	บลอก 1	บลอก 2	บลอก 3	บลอก 4
A	104			
B	101			
C	108			
D	110		สุ่มเช่นเดียวกับ	
E	105		บลอกที่ 1จนครบ	
F	109		ทั้ง 4 บลอก	
G	102			
H	106			
I	103			
J	107			

วิธีการคำนวณหาค่าเหล่านี้คล้ายกับในแผนการทดลองแบบ CRD คือ

$$\text{Correction factor (CF)} = \frac{(\sum X_{ij})^2}{kn}$$

ในที่นี้ $i = 1, 2, \ldots, k$ (เมื่อ k = จำนวนทรีตเมนต์) และ $j = 1, 2, \ldots, n$ (เมื่อ n เป็นจำนวนบลอก) สำหรับ TSS และ SSTr นั้น คำนวณโดยใช้สมการที่กล่าวมาแล้ว คือสมการ (8-3) และ (8-4) ส่วน SSB นั้น คำนวณโดยใช้สมการ

$$SSB = \frac{(B_1^2 + B_2^2 + \ldots + B_n^2)}{k} - CF$$

$$= \frac{\sum B_j^2}{k} - CF \qquad \ldots(10\text{-}2)$$

โดยที่ B คือผลรวมของแต่ละบลอก เมื่อนำสมการ (10-2) ไปเพิ่มเติมลงในตาราง 9.2.1 ก็จะได้สมการแสดงการวิเคราะห์วาเรียนซ์ในการทดลองแบบ RCB ที่สมบูรณ์ ส่วน SSE นั้น คงหาได้โดยการหักลบ คือ

$$\mathbf{SSE = TSS - SSB - SSTr}$$

และ df ของความแปรปรวนเหล่านี้ได้แก่

Total df = จำนวนค่าสังเกตทั้งหมด - 1 = kn - 1

Block df = จำนวนบลอก - 1 = n - 1

Treatment df = จำนวนทรีตเมนต์ - 1 = k - 1

Error df = kn - 1 - (n - 1) - (k - 1) = (k - 1) (n - 1)

ในการวิเคราะห์วาเรียนซ์นั้นอาจแสดงเป็นตัวอย่าง โดยใช้ข้อมูลในตาราง 10.4.1 ตารางนี้เป็นแบบอย่างสำหรับการจัดข้อมูลใน RCB คือ มีระเบียบในแนวตั้งหรือบลอกและในแนวนอนหรือทรีตเมนต์ ซึ่งเรียกว่าเป็นการจัดระเบียบแบบสองทาง การทดลองดังกล่าวมีแบบการจำลองทางคณิตศาสตร์ดังนี้

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \qquad \ldots(10\text{-}3)$$

$i = 1, 2, \ldots, k$; $j = 1, 2, \ldots, n$ ซึ่ง k และ n เป็นจำนวนทรีตเมนต์และบลอกตามลำดับ α และ β เป็นผลของทรีตเมนต์และบลอก ทั้งที่มีข้อกำหนดว่า ทั้ง α และ β มีผลในแบบบวก[5] และ $\varepsilon_{ij} \sim NID(0, \sigma^2)$

ตาราง 10.4.1 ผลผลิตสดของหญ้าอาหารสัตว์ เมื่อได้รับปุ๋ยไนโตรเจนในอัตราต่าง ๆ กัน (ตัน/ไร่)

ปุ๋ย	บลอก				รวม
	I	II	III	IV	
A ไม่ใส่ปุ๋ย (control)	4	6	5	8	23
B 15 กก./ไร่	9	8	7	10	34
C 30 กก./ไร่	9	10	7	8	34
D 45 กก./ไร่	10	12	8	9	39
E 60 กก./ไร่	13	12	8	12	45
F 75 กก./ไร่	12	10	12	9	43
รวม	57	58	47	56	218

เมื่อนำข้อมูลในตาราง 10.4.1 มาวิเคราะห์วาเรียนซ์ก็ได้ผลดังนี้

$$\text{Correction factor (CF)} = \frac{(\sum X_{ij})^2}{kn} = \frac{(218)^2}{24} = 1{,}980.17$$

$$\begin{aligned} TSS &= \sum X_{ij}^2 - CF \\ &= 4^2 + 6^2 + \ldots + 9^2 - 1{,}980.17 \\ &= 127.83 \end{aligned}$$

$$SSB = \frac{(\sum B_j)^2}{k} - CF$$

$$= \frac{1}{6}(57^2+58^2+47^2+56^2)-1{,}980.17$$

$$= 12.83$$

$$\text{SSTr} = \frac{(\sum T_i)^2}{n} - \text{CF}$$

$$= \frac{1}{4}(23^2+34^2+\ldots+43^2)-1{,}980.17$$

$$= 78.83$$

$$\text{SSE} = \text{TSS}-\text{SSB}-\text{SSTr}$$

$$= 127.83-12.83-78.83$$

$$= 36.17$$

เมื่อคำนวณค่า df โดยวิธีการที่กล่าวถึงข้างบนเรียบร้อยแล้วก็นำค่าที่ได้ทั้งหมดลงตารางวิเคราะห์วาเรียนซ์ดังตาราง 10.4.2 แล้วคำนวณหาค่า MS ดังนี้

$$\text{MS (mean square)} = \frac{\text{Sum of square}}{\text{df}}$$

ต่อจากนั้นก็คำนวณหาค่า F

ตาราง 10.4.2 ผลการวิเคราะห์วาเรียนซ์ของข้อมูลจากตาราง 10.4.1

Sources	df	SS	MS	F	F 5%	F 1%
Blocks	3	12.83	4.28	1.78^{ns}	3.29	5.42
Treatments	5	78.83	15.77	6.54**	2.90	4.50
Error	15	36.17	2.41			
Total	23	127.83				

ns, ** = ไม่แตกต่างทางสถิติ และแตกต่างที่ระดับ 1 เปอร์เซ็นต์ตามลำดับ, CV = 17.10%

ในขั้นต่อไปก็เปรียบเทียบค่า F นี้ กับค่า F จากตาราง สำหรับบลอกเป็นค่า F ที่ df 3 และ 15 ส่วนทรีตเมนต์ก็เปิดที่ df 5 และ 15 การตั้งสมมุติฐานทำได้หลายรูปแบบดังอธิบายไว้ในตอน 9.7 เมื่อทดสอบแล้วพบความแตกต่างทางสถิติอย่างมีนัยสำคัญ ก็ดำเนินการหาความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ยและของความแตกต่าง เพื่อนำไปใช้ประโยชน์ในการเปรียบค่าเฉลี่ยต่อไป ดังนี้

$$s_{\bar{x}} = \sqrt{\frac{\text{MSE}}{n}} = \sqrt{\frac{2.41}{4}} = 0.78 \quad \text{ตัน/ไร่}$$

$$s_{\bar{d}} = \sqrt{\frac{2\text{MSE}}{n}} = \sqrt{\frac{2(2.41)}{4}} = 1.09 \quad \text{ตัน/ไร่}$$

และอาจคำนวณค่า CV ได้ดังนี้

$$CV(\%) = \frac{\sqrt{MSE}}{\overline{X}_{..}} \times 100 = \frac{\sqrt{2.41}}{9.08} \times 100 = 17.10\%$$

การทดสอบสมมุติฐานเกี่ยวกับบลอกอาจไม่จำเป็น ทั้งนี้เพราะเราทราบแหล่งของความคลาดเคลื่อนดีแล้ว ปกติบลอกมักมีความแตกต่างกันในทางสถิติ แต่ในการทดลองบางอย่างบลอกอาจเป็นปัญหาอีกชนิดหนึ่งที่ต้องทดสอบ ตัวอย่างเช่น โรงงานแห่งหนึ่งต้องการทดสอบความสามารถในการทำงานของคนงาน 5 คน ในการทดสอบกระทำโดยใช้เครื่องจักร 3 ชนิด ทั้งนี้ชนิดของเครื่องจักรจัดเป็นบลอก ในกรณีเช่นนี้เราอาจต้องการทราบด้วยว่าเครื่องจักรเหล่านี้มีประสิทธิภาพแตกต่างกันหรือไม่

10.5 แบบจำลองทางคณิตศาสตร์

แบบจำลองทางคณิตศาสตร์ของ RCB คล้ายคลึงกับของ CRD นั่นเอง เพียงแต่เราเพิ่มผลของการใช้บลอกเข้าไปในสมการเท่านั้น ดังนั้นถ้าให้ β เป็นผลของการใช้บลอก ก็อาจดัดแปลงสมการ (9-7) เป็นแบบจำลองทางคณิตศาสตร์ของ RCB ดังนี้

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \qquad ...(10\text{-}3)$$

เมื่อให้ $i=1, 2,...,k$ (k = จำนวนทรีตเมนต์) ; $j = 1, 2, ..., n$ (n = จำนวนบลอก) ; α และ β คือผลของทรีตเมนต์และผลของการใช้บลอกตามลำดับ ข้อกำหนดที่สำคัญของแบบจำลองมีดังนี้

(1) ผลของทรีตเมนต์และผลของบลอกต่างก็เป็นแบบบวก คือถ้า μ_{ij} เป็นค่าเฉลี่ยของค่าสังเกตแต่ละค่าก็หาได้ว่า

$$\mu_{ij} = \mu + \alpha_i + \beta_j$$

ซึ่งอาจแสดงให้เห็นว่า ความแตกต่างระหว่างทรีตเมนต์ต่าง ๆ ในแต่ละบลอกเท่ากัน เช่น ความแตกต่างระหว่าง α_i ในบลอกที่หนึ่งจะเท่ากับความแตกต่างระหว่าง α_i ในบลอกที่สอง ดังนี้

$$(\mu + \alpha_2 + \beta_1) - (\mu + \alpha_1 + \beta_1) = (\mu + \alpha_2 + \beta_2) - (\mu + \alpha_1 + \beta_2)$$

ในทำนองเดียวกันหาได้ว่า ความแตกต่างระหว่างบลอกที่รับทรีตเมนต์ต่าง ๆ ก็เท่ากันด้วย

(2) ความคลาดเคลื่อน ε_{ij} มีการกระจายแบบปกติ มีค่าเฉลี่ยเท่ากับ 0 และมีวาเรียนซ์เท่ากับ σ^2 เมื่อใช้แบบจำลองหรือโมเดลที่เป็นปัจจัยคงที่ (หรือโมเดลที่ 1) ก็มีข้อกำหนดว่า $\sum\alpha_i = \sum\beta_j = 0$ ในทำนองเดียวกับสมการ (9-9) เราหาได้ว่า

$$\overline{X}_{i.} = \mu + \alpha_i + \bar{\varepsilon}_{i.}$$
$$\overline{X}_{.j} = \mu + \beta_j + \bar{\varepsilon}_{.j}$$

แล้วสามารถหา EMS สำหรับปัจจัยคงที่และปัจจัยสุ่ม โดยใช้วิธีการคล้ายคลึงกับวิธีที่ใช้ในตอน 9.7 นั่นเอง ซึ่งแสดง EMS เหล่านี้ดังตาราง 10.5.1

ตาราง 10.5.1 ค่า EMS ในแบบจำลองปัจจัยคงที่ (model I) และแบบจำลองปัจจัยสุ่ม (model II)

Sources	df	Expected Mean Square (EMS)	
		Model I	Model II
Blocks	n − 1	$\sigma^2 + kK_b^2$	$\sigma^2 + k\sigma_b^2$
Treatment	k − 1	$\sigma^2 + nK_a^2$	$\sigma^2 + n\sigma_a^2$
Error	(k−1)(n−1)	σ^2	σ^2
Total	kn − 1		

นอกเหนือจากข้อกำหนดซึ่งกล่าวไว้เกี่ยวกับแบบจำลองในตอน 9.7 แล้ว แบบจำลองปัจจัยสุ่มของ RCB นี้ ต้องมีข้อกำหนดเพิ่มเติมว่า β_j มีการกระจายแบบปกติ โดยมีค่าเฉลี่ยเป็น 0 และวาเรียนซ์เท่ากับ σ^2 ด้วย

ในการทดลองโดยทั่วไป เรามักถือว่าการทดลองแต่ละบลอก เป็นตัวแทนของประชากร ของซ้ำทั้งหมด ไม่ใช่เป็นค่าเฉพาะของการทดลองครั้งใดครั้งหนึ่ง ดังนั้นผลของบลอกจึงเป็นปัจจัยสุ่มนั่นเอง จึงเห็นได้ว่าบางครั้งจะพบแบบจำลองหรือโมเดลชนิดผสม[6] หรือโมเดลที่ 3 คือทรีตเมนต์เป็นปัจจัยคงที่แต่บลอกเป็นปัจจัยสุ่ม

10.6 ค่าสูญหาย

ในการทดลองบางครั้งอาจมีค่าสูญหาย เช่น พืชบางแปลงถูกทำลาย สัตว์ทดลองตายเสียก่อนที่จะสิ้นสุดการทดลอง หรือมีการบันทึกข้อมูลผิดพลาด ซึ่งจำเป็นต้องตัดทิ้งไป ในการทดลองแบบ CRD นั้น เราอาจวิเคราะห์โดยไม่ใช้ค่านี้ก็ได้ แต่ในการทดลองแบบ RCB เราต้องมีข้อมูลของทุกทรีตเมนต์ครบทุกค่า ซึ่งอาจประมาณค่าที่สูญหายหรือใช้การไม่ได้นี้โดยวิธีคำนวณ สมการในการคำนวณค่าสูญหายคือ

$$\hat{X} = \frac{kT_i + nB_j - G}{(n-1)(k-1)} \qquad \ldots(10\text{-}4)$$

เมื่อ $\hat{X}$ = ค่าสูญหาย

n = จำนวนบลอก

k = จำนวนทรีตเมนต์

T_i = ผลบวกของทรีตเมนต์ที่มีค่าสูญหาย

B_j = ผลบวกของบลอกที่มีค่าสูญหาย

ตาราง 10.6.1 น้ำหนักของเมล็ดถั่วเขียวพันธุ์ต่าง ๆ ซึ่งมีค่าสูญหาย 1 ค่า (กรัม/100 เมล็ด)

พันธุ์	บลอก				รวม
	I	II	III	IV	
กพส.1	7.2	7.4	7.3	7.2	29.1
กพส.2	7.4	7.4	7.5	7.3	29.6
อู่ทอง 1	6.9	7.1	7.2	6.8	28.0
424-61	6.5	6.4	-	6.6	19.5
8-50-16	5.8	6.2	6.2	6.3	24.5
รวม	33.8	34.5	28.2	34.2	130.7

การคำนวณค่าสูญหาย 1 ค่า

เมื่อมีค่าสูญหายเพียง 1 ค่า ดังแสดงในตาราง 10.6.1 อาจคำนวณหาค่าสูญหายได้ดังนี้

$$k = 5,\ n = 4,\ T_i = 19.5,\ B_j = 28.2,\ G = 130.7$$

$$\hat{X} = \frac{(5)(19.5) + (4)(28.2) - 130.7}{(4-1)(5-1)}$$

$$= 6.6 \text{ กรัม/100 เมล็ด}$$

แล้วนำค่า 6.6 ที่คำนวณได้ใส่ลงในแปลงสูญหาย เปลี่ยนผลรวมของทรีตเมนต์ ของบลอก และผลรวมทั้งหมด ซึ่งได้เท่ากับ 26.1, 34.8 และ 137.3 ตามลำดับ แล้วทำการวิเคราะห์วาเรียนซ์โดยวิธีปกติที่กล่าวมาแล้ว โดยใช้ผลบวกที่ได้ใหม่ ซึ่งได้ผลดังนี้

$$TSS = 4.87,\ SSB = 0.11,\ SSTr = 4.54 \text{ และ } SSE = 0.22$$

แล้วนำค่าที่คำนวณได้ลงตารางวิเคราะห์วาเรียนซ์ดังตาราง 10.6.2 ในตารางนี้ df ของ total และ error จะลดลงไปอย่างละ 1 และคำนวณค่าปรับ[7] สำหรับ mean square เนื่องจากทรีตเมนต์ (C) ดังนี้

$$C = \frac{[B - (k-1)\hat{X}]^2}{k(k-1)^2} = \frac{[28.2 - (5-1)\,6\cdot 6]^2}{5(5-1)^2} = 0.04$$

ทั้งนี้ B = ผลรวมของบลอกที่มีค่าสูญหาย และ $\hat{X}$ คือค่าสูญหายที่คำนวณได้ แล้วนำค่า C นี้ไปลบออกจาก MSTr ได้

$$MSTr(adjusted) = 1.14 - 0.04 = 1.10$$

แล้วนำค่าที่ได้ทั้งหมดนี้ลงตารางวิเคราะห์วาเรียนซ์ดังตาราง 10.6.2

ตาราง 10.6.2 ผลการวิเคราะห์วาเรียนซ์ของข้อมูลที่มีค่าสูญหาย 1 ค่าจากตาราง 10.6.1

Sources	df	SS	MS	F	F(table) 5%	F(table) 1%
Blocks	3	0.11	0.04	2.00	3.59	6.23
Treatments	4	4.54	1.10	55.00**	3.36	5.65
Error	12 − 1 = 11	0.22	0.02			
Total	19 − 1 = 18	4.87				

** = แตกต่างทางสถิติที่ระดับ 1 เปอร์เซ็นต์

เมื่อต้องการเปรียบเทียบระหว่างค่าเฉลี่ยของทรีตเมนต์ที่มีค่าสูญหาย กับทรีตเมนต์อื่น ๆ โดยใช้ LSD ก็อาจหา $s_{\bar{d}}$ โดยใช้สมการดังนี้

$$s_{\bar{d}} = \sqrt{MSE\left[\frac{2}{n} + \frac{k}{n(n-1)(k-1)}\right]} \qquad \ldots(10\text{-}5)$$

$$= \sqrt{0.02\left[\frac{2}{4} + \frac{5}{(4)(3)(4)}\right]}$$

$$= 0.11 \text{ กรัม/100 เมล็ด}$$

การคำนวณค่าสูญหายเกิน 1 ค่า

เมื่อมีค่าสูญหายเกิน 1 ค่า ก็ใช้สมการ (10 - 4) เช่นเดิม แต่เราต้องประมาณค่าสูญหายบางค่าแบบหยาบ ๆ เสียก่อน โดยใช้สมการ $(X_{i\cdot} + X_{\cdot j})/2$ เมื่อ $\overline{X}_{i\cdot}$ และ $X_{\cdot j}$ คือค่าเฉลี่ยของทรีตเมนต์และบลอกที่มีค่าสูญหาย ประมาณอย่างนี้ทุกค่าแต่เว้นไว้ 1 ค่า นำค่าที่ประมาณอย่างหยาบ ๆ นี้ไปบวกกับผลรวมทั้งหมด แล้วคำนวณค่าที่เว้นไว้ 1 ค่า โดยใช้สมการ (10-4) ดังตัวอย่างในตาราง 10.6.3

ตาราง 10.6.3 ข้อมูลที่มีค่าสูญหาย 2 ค่า

ทรีตเมนต์	บลอก I	บลอก II	บลอก III	บลอก IV	รวม
A	8	7	6	9	30
B	12	8	X_{23}	6	26
C	X_{31}	8	9	12	29
D	13	12	10	12	47
รวม	33	35	25	39	132

ในตารางดังกล่าวนั้นมีค่าสูญหาย 2 ค่า คือในตำแหน่ง X_{23} และ X_{31} เราอาจคำนวณ X_{31} แบบหยาบ ๆ เสียก่อน โดยใช้สมการ

$$X_{31} = (\overline{X}_{3.} + \overline{X}_{.1}) / 2 = [29/3 + 33/3] / 2 = 10.33$$

แล้วนำค่า 10.33 นี้ไปรวมกับผลรวมทั้งหมดเสียก่อนซึ่งได้ 142.33 แล้วคำนวณ X_{23} โดยใช้สมการ (10-4) การคำนวณจะหมุนเวียนจนกว่าค่าประมาณจะไม่เปลี่ยนแปลงมากนัก ดังนี้

(1) $X_{23} = [(4)(26) + (4)(25) - 142.33] / (3)(3) = 6.85$

(2) $X_{31} = [(4)(29) + (4)(33) - 138.85] / 9 = 12.12$
จงสังเกตว่า 138.85 = 132 + 6.85 แล้วเวียนกลับมาคำนวณ X_{23} ใหม่

(3) $X_{23} = [(4)(26) + (4)(25) - 144.12] / 9 = 6.65$
ทั้งนี้ 144.12 = 132 + 12.12 แล้วคำนวณ X_{31} ใหม่

(4) $X_{31} = [(4)(29) + (4)(33) - 138.65] / 9 = 12.15$

(5) $X_{23} = [(4)(26) + (4)(25) - 144.15] / 9 = 6.65$

(6) สิ้นสุดการคำนวณ

ซึ่งจะเห็นได้ว่า X_{31} และ X_{23} ที่คำนวณได้มีค่าคงที่หรือค่อนข้างคงที่แล้ว คือคำนวณต่อไปก็ไม่เปลี่ยนแปลงมากนัก จึงอาจใช้เป็นค่าสูญหายได้ นำค่านี้ใส่ลงในช่องสูญหาย เปลี่ยนแปลงผลรวมของบลอกของทรีตเมนต์ และผลรวมทั้งหมด แล้วทำการวิเคราะห์วาเรียนซ์โดยวิธีปกติ แต่ df ของ total และ error ลดลงอย่างละ 2 ค่า ทั้งนี้ MSTr ที่หาได้มักสูงกว่าที่ควรจะเป็น แต่ F-test มีความคลาดเคลื่อนเพียงเล็กน้อย เว้นแต่ว่ามีค่าสูญหายมากค่าเกินไป อย่างไรก็ดีเราอาจปรับค่า MSTr ตามวิธีการซึ่งจะอธิบายต่อไป คือก็นำข้อมูลลงตารางให้สมบูรณ์ เช่นตาราง 10.6.4 แล้วทำการวิเคราะห์วาเรียนซ์ตามปกติ เช่น TSS = 1,511.85 − 1,421.29 = 90.56 แล้วนำผลการวิเคราะห์ลงตาราง 10.6.5

ตาราง 10.6.4 ข้อมูลสูญหายและผลรวมจากตาราง 10.6.3

ทรีตเมนต์	บลอก				รวม
	I	II	III	IV	
A					
B			6.65		32.65
C	12.15				41.15
D					
รวม	45.15		31.65		150.80

ตาราง 10.6.5 ผลการวิเคราะห์วาเรียนซ์เมื่อข้อมูลสูญหาย 2 ค่า

Sources	df	SS	MS
Blocks	3	25.27	-
Treatments	3	45.79	15.26
Error	7	19.50	2.79
Total	13	90.56	

ต่อไปก็ปรับค่า SSTr โดยใช้วิธีคำนวณดังนี้ คือหา Total SS (ข้อมูลเดิมตาราง 10.6.3) = 1,320 - 1,244.57 = 75.43 ต่อจากนั้นก็นำ SSerror มาลบออก และคำนวณค่า SSTr (adjusted) ต่อไปดังนี้

Sources	df	SS	MS
Total (คำนวณจากตารางเดิม 10.6.3) (บรรทัด 1)	13	75.43	
Error (จากตารางใหม่ 10.6.5) (บรรทัด 2)	7	19.50	2.79
Blocks + Treatments (บรรทัด 1 ลบ 2)	6	55.93	
Block (Unadjusted)	3	13.26	
Treatments (adjusted)	3	45.79	15.26

ทั้งนี้ Block (Unadjusted) คำนวณได้ดังนี้

$$= \frac{33^2}{3} + \frac{35^2}{4} + \frac{25^2}{3} + \frac{39^2}{4} - 1{,}244.57 \quad \text{(ข้อมูลจากตารางเดิม } = 13.26)$$

แล้วนำ SSTr ที่ปรับค่าแล้วไปทดสอบความแตกต่างต่อไป

การเปรียบเทียบระหว่างค่าเฉลี่ยของทรีตเมนต์ที่มีค่าสูญหายด้วยกัน เช่น B กับ C หรือระหว่างทรีตเมนต์ที่มีค่าสูญหายกับทรีตเมนต์ปกติ ต้องประมาณจำนวนบลอกสัมฤทธิ์ (n_e)[8] ดังนี้

(1) ในบลอกที่มีค่าสังเกตอยู่ครบทั้ง 2 ทรีตเมนต์ที่เปรียบเทียบกัน เช่น บลอกที่ 2 หรือ 4 ให้มีค่าเท่ากับ 1

(2) ในบลอกที่มีค่าสังเกตของทรีตเมนต์หนึ่งอยู่ และอีกทรีตเมนต์ไม่มี ให้มีค่าเท่ากับ (k - 2)/(k - 1) เช่น ในบลอกที่ 1 หรือ 3 มีค่าบลอกสัมฤทธิ์ (4 - 2)/(4 - 1) = 2/3

(3) ในบลอกที่ไม่มีค่าสังเกตของทรีตเมนต์นั้นให้มีค่าเป็น 0

ตัวอย่างเช่นเมื่อเปรียบเทียบระหว่างทรีตเมนต์ B และ C ค่าบลอกสัมฤทธิ์ของบลอก B และ C ในแต่ละซ้ำหาได้ดังนี้

	บลอกที่ 1	2	3	4	รวม
ค่าบลอกสัมฤทธิ์ของ B	2/3	1	0	1	2.67
ค่าบลอกสัมฤทธิ์ของ C	0	1	2/3	1	2.67

ดังนั้นหาได้ว่า

$$s_{\bar{d}} = \sqrt{MSE\left(\frac{1}{n_{eb}} + \frac{1}{n_{ec}}\right)}$$

$$= \sqrt{2.79\left(\frac{1}{2.67} + \frac{1}{2.67}\right)}$$

$$= 1.44$$

เมื่อเปรียบเทียบระหว่างทรีตเมนต์ที่มีค่าสูญหายและไม่มีค่าสูญหาย ก็คำนวณค่าบลอกสัมฤทธิ์โดยวิธีที่กล่าวมาแล้ว เช่น ต้องการเปรียบเทียบระหว่าง A และ B ค่าบลอกสัมฤทธิ์ของ A = 1 + 1 + 2/3 + 1 = 3.67 และ B = 1 + 1 + 0 + 1 = 3

10.7 ประสิทธิภาพสัมพันธ์

เมื่อได้ทำการทดลองโดยใช้แผนการทดลองแบบ RCB แล้ว บางครั้งก็สงสัยว่าการใช้บลอกจะเพิ่มความเที่ยงตรงในการทดลองได้ดีกว่า CRD ซึ่งไม่ใช้บลอกหรือไม่ ข้อเปรียบเทียบเช่นนี้เราเรียกว่า "ประสิทธิภาพสัมพันธ์" ซึ่งเขียนย่อ ๆ ว่า RE หาได้จากสมการ

$$\text{RE (ของ RCB ต่อ CRD)} = \frac{\mathbf{E_e(CRD)}}{\mathbf{E_e(RCB)}} \times \frac{(\mathbf{n}_1+1)(\mathbf{n}_2+3)}{(\mathbf{n}_1+3)(\mathbf{n}_2+1)} \times 100 \qquad \text{...(10-6)}$$

ทั้งนี้ให้

$E_e(CRD)$ = MS error ของ CRD ซึ่งหาโดยวิธีการประมาณ

$E_e(RCB)$ = MS error ของ RCB ซึ่งได้จากตารางวิเคราะห์วาเรียนซ์

n_1, n_2 = df ของ error ใน RCB และ CRD ตามลำดับ

ในสมการ (10-6) นั้น ส่วนแรกของสมการเป็นการคำนวณ RE ส่วนที่สองเป็นตัวปรับค่า df ในกรณี df ของ error น้อยกว่า 20

ในสมการ (10-6) อาจคำนวณ E_e (CRD) โดยใช้สมการ

$$E_e(CRD) = \frac{n_b E_b + (n_t + n_e) E_e}{n_b + n_t + n_e} \qquad \text{...(10-7)}$$

หรือใช้สมการ

$$E_e(\text{CRD}) = \frac{(n-1)\text{MSB}+n(k-1)\text{MSE}}{nk-1} \qquad \ldots(10\text{-}8)$$

เมื่อ k = จำนวนทรีตเมนต์, n = จำนวนบลอก, MSB = mean square ของบลอก และ MSE = mean square ของ error เมื่อใช้สมการ (10-6) และ (10-7) ก็สามารถหาประสิทธิภาพสัมพันธ์ของ RCB ต่อ CRD จากผลการวิเคราะห์ในตาราง 10.4.2 ดังนี้

$$E_e(\text{CRD}) = \frac{(3)(4.28)+4(5)(2.41)}{23} = \frac{61.04}{23} = 2.65$$

$$\text{RE}(\text{RCB}/\text{CRD}) = \frac{2.65}{2.41}\text{x}\frac{(15+1)(18+3)}{(15+3)(18+1)}\times 100 = 108.02\%$$

แสดงว่าการทดลองโดยใช้ RCB ให้ผลดีกว่าแบบ CRD หมายถึงว่า ถ้าใช้ RCB จำนวน 100 บลอก ให้ผลดีเท่ากับใช้ CRD 108.02 ซ้ำ ซึ่งแสดงว่า RCB มีประสิทธิภาพสูงกว่า CRD เพียงเล็กน้อย

10.8 ข้อกำหนดในการวิเคราะห์วาเรียนซ์

ในการใช้แผนการทดลองชนิดต่างๆ เพื่อทำการเปรียบเทียบผลของทรีตเมนต์ ดังที่กล่าวมาแล้วนั้น ใช่ว่าจะกระทำได้อย่างอิสระ วิธีการทางสถิติดังกล่าวที่กล่าวมาแล้วมักมีข้อกำหนดมากมาย การที่มิได้มีการกล่าวถึงข้อกำหนดเหล่านั้นในบทต้น ๆ ก็เพราะต้องรอให้ผู้ศึกษาวิชานี้ ได้เข้าใจถึงเทคนิคการวางแผนการทดลองเสียก่อน เพื่อจะได้เข้าใจในเรื่องข้อกำหนดต่าง ๆ ได้ดีขึ้น ข้อกำหนดในการวิเคราะห์วาเรียนซ์มีดังนี้: (1) ค่าที่สังเกตหรือวัดได้มีการกระจายที่เป็นอิสระแก่กัน[10], (2) ค่าที่สังเกตหรือค่าที่วัดได้กระจายแบบปกติ[11], (3) วาเรียนซ์ของแต่ละทรีตเมนต์เท่ากัน[12], และ (4) อิทธิพลของทรีตเมนต์ บลอก และความคลาดเคลื่อนเป็นแบบบวก[13] ถ้าหากข้อมูลที่วัดได้ไม่มีลักษณะหรือคุณสมบัติเช่นนี้ จะทำให้มีผลกระทบต่อความไวของการทดสอบโดยใช้สถิติ F และ t ซึ่งจะทำให้การสรุปผลเกี่ยวกับการทดลองคลาดเคลื่อนจากความเป็นจริง ต่อไปนี้จะกล่าวถึงรายละเอียดของข้อกำหนดเหล่านี้ ดังนี้

(1) ค่าสังเกตหรือค่าที่วัดได้มีการกระจายที่เป็นอิสระแก่กัน ข้อกำหนดนี้มีความหมายเท่ากับว่า ความคลาดเคลื่อนของข้อมูลเป็นอิสระแก่กัน หมายถึงว่าทรีตเมนต์หนึ่งไม่มีอิทธิพลต่ออีกทรีตเมนต์หนึ่ง หรืออาจกล่าวได้อีกนัยหนึ่งว่า ความคลาดเคลื่อนที่เกิดกับทรีตเมนต์หนึ่งไม่สัมพันธ์กับความคลาดเคลื่อนที่เกิดกับอีกทรีตเมนต์หนึ่ง ตัวอย่างเช่น ในการทดลองที่มี 4 ทรีตเมนต์ คือ A, B, C และ D สมมุติว่า ในการทดลองเราจัดวางทรีตเมนต์เป็นลำดับที่แน่นอน ดังรูป 10.8.1 จากรูปดังกล่าว จะเห็นได้ว่า ทรีตเมนต์ A และ B อยู่ชิดกันเสมอในทุก ๆ ซ้ำ ทรีตเมนต์ A และ C ห่างกัน 1 แปลงเสมอในทุก ๆ ซ้ำ เช่นกัน ซึ่งโดยปกติแล้วทรีตเมนต์ที่อยู่ใกล้เคียงกันจะมีความสัมพันธ์กันเสมอ เช่น พืช 2 พันธุ์ ที่ปลูก

ใกล้เคียงกัน จะเนื่องจากธรรมชาติของพื้นที่ทดลอง หรือโดยสาเหตุอื่น ๆ ก็ตาม ผลผลิตของพืชมักจะสูงหรือต่ำไปในทางเดียวกัน ซึ่งเรียกได้ว่าทรีตเมนต์มีความสัมพันธ์กันนั่นเอง ความสัมพันธ์นี้กระจายออกไปยังทรีตเมนต์ที่อยู่ในลำดับถัดออกไปทุกทรีตเมนต์ ถ้าหากว่าการจัดวางนั้นมีลำดับที่แน่นอน ดังนั้นทรีตเมนต์ที่อยู่ใกล้กัน จะมีความคลาดเคลื่อนที่มีค่าใกล้เคียงกันมากกว่าทรีตเมนต์ที่อยู่ห่างกัน ถ้าจัดวางทรีตเมนต์ระบบเช่นนี้ก็จะมีความสัมพันธ์กันเป็นชุด ๆ เสมอไป

ซ้ำที่ 1	A	B	C	D
ซ้ำที่ 2	D	A	B	C
ซ้ำที่ 3	C	D	A	B
ซ้ำที่ 4	B	C	D	A

รูป 10.8.1 การจัดทรีตเมนต์ในการเปรียบเทียบทรีตเมนต์ A, B, C และ D ในการทดลอง 4 ซ้ำ

ในกรณีดังกล่าวแล้วข้างบนนั้นแสดงให้เห็นว่า การจัดทรีตเมนต์ไว้อย่างมีระบบหรือลำดับนั้นจะทำให้การกระจายของความคลาดเคลื่อนไม่เป็นอิสระแก่กัน วิธีการแก้ไขข้อขัดแย้งต่อข้อกำหนดดังกล่าวนี้ สามารถกระทำได้โดยการสุ่ม การสุ่มทำให้สามารถวางทรีตเมนต์ตามหลักการวางแผนการทดลอง คือให้ทุกทรีตเมนต์มีโอกาสอยู่ใกล้ชิดกันหรือห่างกันเท่ากัน

(2) ค่าที่สังเกตหรือค่าที่วัดได้มีการกระจายแบบปกติ ข้อกำหนดนี้มีความหมายว่า ความคลาดเคลื่อนมีการกระจายแบบปกติ การกระจายแบบนี้หมายถึงว่า ข้อมูลมีการกระจายรูประฆัง คือเป็นรูปโค้ง เมื่อแบ่งครึ่งที่ฐานแล้วลากเส้นตั้งฉากก็จะแบ่งพื้นที่สองด้านออกเป็น 2 ส่วนเท่ากัน ความคลาดเคลื่อนที่กระจายแบบนี้อาจเขียนว่า $\varepsilon_{ij} \sim NID(0, \sigma^2)$ คือความคลาดเคลื่อนเป็นอิสระแก่กัน กระจายแบบปกติโดยมีค่าเฉลี่ยเท่ากับศูนย์ และวาเรียนซ์ σ^2 โดยปกติแล้วข้อมูลทางชีววิทยา เช่น ข้อมูลเกี่ยวกับลักษณะของพืชหรือสัตว์มีการกระจายแบบปกติ หรือใกล้เคียงปกติ ข้อมูลบางอย่างไม่กระจายแบบปกติ เช่น ข้อมูลจากการนับ เช่น จำนวนไข่ของไก่ จำนวนแผลของโรคที่ใบพืช ฯลฯ ข้อมูลแบบนี้มีการกระจายไบโนเมียล ข้อมูลเกี่ยวกับอัตราการเป็นโรคของพืช อัตราการตายของหมูหลังคลอด ฯลฯ ข้อมูลแบบนี้มักมีการกระจายแบบปัวซอง

ข้อมูลไม่กระจายแบบปกติ จะก่อให้เกิดผลกระทบต่อการทดสอบโดยใช้สถิติ F คือทำให้เพิ่มโอกาสที่ทรีตเมนต์แตกต่างกันมากขึ้น อย่างไรก็ดีถ้าการกระจายคลาดเคลื่อนจากปกติเพียงเล็กน้อย ก็สามารถใช้สถิติ F ได้ ทั้งนี้เพราะสถิติ F ไม่ค่อยได้รับความกระทบจากการที่ข้อมูลไม่กระจายแบบปกติได้โดยง่าย คุณสมบัติเช่นนี้เป็นข้อดี เพราะข้อมูลทางชีววิทยามักมีการกระจายที่คลาดเคลื่อนจากการกระจายแบบปกติได้เสมอ โดยเฉพาะอย่างยิ่งข้อมูลที่มีค่าสังเกตเป็นจำนวนน้อย ๆ หรือข้อมูลที่มีค่าบางค่าอยู่นอกกลุ่ม คือสูงหรือต่ำกว่าปกติ

การทดสอบความคลาดเคลื่อนจากการกระจายแบบปกติ อาจกระทำโดยใช้วิธีไค-สแควร์ คือทดสอบความสอดคล้อง[4] โดยให้จำนวนที่เกิดจากการกระจายในทางทฤษฎีเป็นค่าคาดหมาย และค่าที่สำรวจได้เป็นค่าสังเกต อย่างไรก็ดี วิธีนี้ควรใช้เมื่อตัวอย่างมีขนาดใหญ่พอสมควร ในกรณีที่สังเกตด้วยสายตาหรือทดสอบโดยวิธีต่าง ๆ แล้วเห็นว่า ข้อมูลที่สังเกตไม่กระจายแบบปกติ วิธีการปรับปรุงกระทำได้โดยการแปลงข้อมูลโดยวิธีต่าง ๆ ซึ่งจะกล่าวในตอน 10.9 ต่อไป

การทดสอบการกระจายแบบปกติ

การทดสอบว่าข้อมูลมีการกระจายแบบปกติหรือไม่ วิธีการที่สะดวกที่สุดคือ นำค่าเบี่ยงเบนจากค่าเฉลี่ยมาเขียนกราฟ ถ้าหากว่าข้อมูลมีการกระจายแบบปกติแล้ว ค่าเบี่ยงแบนนี้จะมีการกระจายแบบปกติและมีค่าเฉลี่ยเป็นศูนย์หรือใกล้ศูนย์ อย่างไรก็ดี ถ้าเป็นตัวอย่างขนาดเล็กค่าที่ได้มาอาจมีค่าสูง ๆ หรือต่ำ ๆ มากกว่าปกติ ทำให้การกระจายคลาดเคลื่อนออกไป ข้อมูลตัวใดที่สูงหรือต่ำมากจะให้ค่าเบี่ยงเบนสูง แม้มีเพียงไม่กี่ค่าก็ทำให้ผลของการวิเคราะห์วาเรียนซ์เปลี่ยนไป ค่าเหล่านี้อาจเกิดจากผลของการทดลองหรือความผิดพลาดในการจัดการข้อมูล วิธีการตรวจสอบอย่างคร่าว ๆ คือการใช้คะแนนมาตรฐาน เช่น ข้อมูลในตาราง 10.8.3 หาได้ว่า

$$d_{23} \quad = \quad \frac{25.33}{\sqrt{269.92}} = \frac{25.33}{16.42} = 1.54$$

ซึ่งพบว่าไม่มีความสำคัญทางสถิติ

การตรวจสอบการกระจายแบบปกติอีกวิธีหนึ่งคือ ใช้วิธีการทดสอบโดยใช้ไค-สแควร์ โดยทดสอบการสอดคล้องระหว่างค่าสังเกตและจากการคาดหมายโดยใช้สมการ

(3) วาเรียนซ์ของทรีตเมนต์เท่ากัน ข้อกำหนดข้อนี้หมายถึงว่า แต่ละทรีตเมนต์มีวาเรียนซ์ร่วมกัน ทั้งนี้ถึงแม้ว่าค่าเฉลี่ยของทรีตเมนต์จะแตกต่างกันก็ตาม ข้อมูลบางชนิด วาเรียนซ์อาจมีความสัมพันธ์กับค่าเฉลี่ยคือ ถ้าค่าเฉลี่ยสูงวาเรียนซ์ก็สูงตามไปด้วย ดังนั้นถ้าค่าเฉลี่ยไม่แตกต่างกันมากหรือค่าเฉลี่ยสูงสุดไม่แตกต่างจากค่าต่ำสุดเกินไป วาเรียนซ์ของทรีตเมนต์ต่าง ๆ ก็ไม่แตกต่างกันมากเช่นกัน ในทางตรงกันข้าม ถ้าค่าเฉลี่ยแตกต่างกันมาก เช่น ในการทดลองเกี่ยวกับพืช ทรีตเมนต์ที่ใช้ปุ๋ยอัตราสูงมักแตกต่างจากทรีตเมนต์ที่ไม่ใส่ปุ๋ยเลย ผลผลิตอาจแตกต่างกันถึง 2 เท่า เช่นนี้อาจทำให้วาเรียนซ์ต่างกันก็ได้ ในการทดลองเกี่ยวกับแมลง พืชพันธุ์ที่ต้านทาน อาจมีแมลงกัดกินต้นละ 4-5 ตัว แต่พันธุ์ที่ไม่ต้านทาน อาจมีแมลงตั้งแต่ต้นละ 50 ถึง 100 ตัว ดังนี้จะเห็นว่าพันธุ์ต้านทานให้ค่าเฉลี่ยและวาเรียนซ์ต่ำ แต่พันธุ์ไม่ต้านทานกลับมีค่าเฉลี่ยและวาเรียนซ์สูง

เหตุที่ต้องมีข้อกำหนดว่า วาเรียนซ์ของทรีตเมนต์ต่าง ๆ มีค่าเท่ากันก็เพราะว่าวาเรียนซ์ของความคลาดเคลื่อนในตารางวิเคราะห์วาเรียนซ์จากการใช้แผนการทดลองก็คือ ผลรวมของวาเรียนซ์นั่นเอง วาเรียนซ์รวมนี้เกิดจากความคลาดเคลื่อนของแต่ละทรีตเมนต์ ดังที่กล่าวมาแล้วในตอน 4.3 ว่า

ก่อนที่จะรวมวาเรียนซ์เข้าด้วยกัน (เพื่อหาวาเรียนซ์เฉลี่ย) ได้ ต้องมีการทดสอบเสียก่อนว่า วาเรียนซ์เหล่านั้นต้องเท่ากัน ในกรณีของการทดลองที่มีจำนวนมากกว่า 2 ทรีตเมนต์ การทดสอบวาเรียนซ์ เท่ากัน อาจใช้ Bartlett test ซึ่งแสดงไว้ในตอน 7.4 ถ้าทดสอบแล้วพบว่า วาเรียนซ์แตกต่างกันอย่างมากก็ไม่ควรดำเนินการวิเคราะห์วาเรียนซ์ โดยทันที ทั้งนี้เพราะทำให้การทดสอบโดยใช้ค่า F และ t ถึงระดับความแตกต่างได้ง่ายกว่าความเป็นจริง

วิธีการที่จะหลีกเลี่ยงไม่ให้วาเรียนซ์ต่างกัน สามารถกระทำได้โดยใช้ความระมัดระวังในการทดลอง เช่น ให้การดูแลทุก ๆ ทรีตเมนต์เหมือน ๆ กัน ก่อนวิเคราะห์วาเรียนซ์ ถ้าพบว่าทรีตเมนต์ใดมีค่าสูงหรือต่ำผิดปกติเราอาจไม่นำมาวิเคราะห์ก็ได้

การทดสอบดุลยภาคของวาเรียนซ์

การทดสอบดุลภาคของวาเรียนซ์ คือการทดสอบว่า วาเรียนซ์ของทรีตเมนต์ต่าง ๆ เท่ากัน หรือทดสอบสมมุติฐาน $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = ... = \sigma_k^2$ ซึ่งมีวิธีการทดสอบหลายวิธี จะแสดงพร้อมกับตัวอย่างโดยใช้ข้อมูลในตาราง 9.3.1 ซึ่งมีวาเรียนซ์ ดังนี้

ทรีตเมนต์	A	B	C	D	E
วาเรียนซ์	4.25	6.67	32.92	6.00	9.50

(1) วิธีของฮาร์ตเลย์ (Hartley's test)

วิธีของฮาร์ตเลย์ใช้สถิติดังนี้

$$F_{max} = \frac{s^2_{max}}{s^2_{min}}$$

ซึ่งมี df ตัวตั้ง = k (k = จำนวนทรีตเมนต์) และตัวหาร = n – 1 (n = จำนวนซ้ำ) ถ้าจำนวนซ้ำไม่เท่ากัน df ตัวหารใช้ $n_i - 1$ (เมื่อ n_i เป็นจำนวนซ้ำสูงสุด)

จากวาเรียนซ์ของตาราง 9.3.1 หาได้ว่า

$$F_{max} = 32.92 / 4.25 = 7.75$$

เมื่อเปิดตาราง ผ.13 ที่ df 5, 3 พบว่า ค่า F_{max} ที่ระดับความแตกต่าง 0.05 = 50.7 ดังนั้น จึงสรุปว่าการทดลองนี้ให้วาเรียนซ์ที่มีดุลยภาค

(2) วิธีของคอกคราน (Cochran's test)

วิธีของคอกครานใช้สถิติดังนี้

$$C = \frac{s^2_{max}}{\sum s_i^2}$$

ซึ่งมี df ตัวตั้ง = k, ตัวหาร = n – 1 ถ้าจำนวนซ้ำไม่เท่ากัน df ตัวหารใช้ n_i- 1 (เมื่อ n_i เป็นจำนวนซ้ำสูงสุด)

จากวาเรียนซ์ของตาราง 9.3.1 หาได้ว่า

$$C = 32.92/(4.25 + 6.67 + 32.92 + 6.00 + 9.50)$$
$$= 0.555$$

เมื่อเปิดตาราง ผ.14 พบว่า $C_{(0.05)}$ df 5, 3 = 0.598 ดังนั้น จึงสรุปได้ว่า การทดลองนี้ให้วาเรียนซ์ที่มีดุลยภาค

(3) วิธีของบาล์ตเลตต์ (Bartlett's test)

วิธีของบาร์ตเลตต์ ได้อธิบายไว้แล้วในตอน 7.4 เมื่อจะใช้กับข้อมูลในตาราง 9.3.1 ก็เปลี่ยนแปลงจำนวนทรีตเมนต์, df และวาเรียนซ์เท่านั้น

(4) วิธีของเลอวีน (Levene's test)

วิธีนี้นำค่าสัมบูรณ์ของค่าเบี่ยงจากค่าเฉลี่ย ($|X_i - \bar{X}|$) ของข้อมูลทุกค่า มาวิเคราะห์วาเรียนซ์ตามแผนการทดลองแบบ CRD จัดทำตารางวิเคราะห์วาเรียนซ์ แล้วคำนวณและทดสอบค่า F ตามปกติ ซึ่งจัดได้ว่าเป็นวิธีที่ง่าย ไม่ต้องใช้ตารางทดสอบเพิ่มเติม

(4) อิทธิพลของทรีตเมนต์ บลอก และความคลาดเคลื่อนเป็นแบบบวก ในการทดลองเปรียบเทียบพันธุ์พืชโดยทำการปลูกแต่ละพันธุ์ในหลาย ๆ แปลง ถึงแม้ว่าพืชพันธุ์เดียวกันในแต่ละแปลงให้ผลผลิตต่างกัน แต่เราก็เข้าใจว่า แต่ละพันธุ์ควรมีค่าคงที่ ส่วนที่แตกต่างนั้นอาจเกิดจากสาเหตุอื่น เช่น อาจเป็นความคลาดเคลื่อน เช่นนี้หมายความว่า แต่ละค่าที่เราสังเกตหรือวัดได้คือผลของทรีตเมนต์รวมกับความคลาดเคลื่อนนั่นเอง ซึ่งเป็นการรวมในแบบบวก ดังที่เราได้เห็นสำหรับแผนการทดลองแบบ CRD ในสมการ (9-3) และถ้าใช้แผนการทดลองแบบ RCB ก็ได้ ดังสมการ (10-3) ในสมการเหล่านี้เห็นได้ว่าผลของทรีตเมนต์ บลอก และความคลาดเคลื่อนที่เข้าไปทบในค่าเฉลี่ย (μ) เป็นแบบบวก

จากตัวอย่างข้างบน อาจกล่าวได้ว่า เมื่อใดค่าต่าง ๆ ที่ประกอบกันขึ้นมีค่าสังเกตเป็นแบบบวกแล้ว เราอาจพบว่าผลของทรีตเมนต์หนึ่ง ๆ จะมีค่าเท่ากันในทุกบลอก หรืออาจกล่าวได้ว่าในบลอกหนึ่ง ๆ ผลของบลอกที่มีต่อทุกทรีตเมนต์เท่ากัน หรืออาจพูดอีกนัยหนึ่งว่า ถ้าผลของทรีตเมนต์และบลอกเป็นแบบบวก ก็ไม่มีปฏิกิริยาระหว่างทรีตเมนต์และบลอกนั่นเอง การแสดงผลแบบบวกอาจอธิบายจากค่าสังเกตดังตารางดังนี้

	บลอก 1	บลอก 2	ความแตกต่าง (ผลของบลอก)
ทรีตเมนต์ A	10	20	10
ทรีตเมนต์ B	25	35	10
ความแตกต่าง (ผลของทรีตเมนต์)	15	15	

ซึ่งเห็นได้ว่า ความแตกต่างระหว่างทรีตเมนต์ A ในบลอกที่ 1 และ 2 มีค่าเท่ากับความแตกต่างระหว่างทรีตเมนต์ B ในบลอกที่ 1 และ 2 แสดงว่าผลของทรีตเมนต์เป็นแบบบวก ในการทำคล้ายคลึงกัน เราก็สามารถอธิบาย ได้ว่า ผลของบลอกเป็นแบบบวกได้เช่นกัน ทั้งนี้เพราะความแตกต่างระหว่าง ทรีตเมนต์ A และ B เท่ากันในทุกบลอก

การทดลองบางชนิด ผลของทรีตเมนต์เป็นแบบคูณ ซึ่งมักเกิดกับข้อมูลที่เกิดจากการนับ เช่น ทดสอบการต้านทานแมลงของข้าว 2 พันธุ์ พันธุ์หนึ่งไม่ต้านทานต่อแมลง อีกพันธุ์หนึ่งต้านทาน ถ้าเราปล่อยแมลงในบลอกที่หนึ่งกอละ 10 ตัว ในบลอกที่สองกอละ 20 ตัว แมลงแต่ละตัวกัดพันธุ์ต้านทานตัวละ 1 แผล พันธุ์ไม่ต้านทานตัวละ 2 แผล เมื่อนับจำนวนแผลก็เป็นดังนี้

	บลอก		ความแตกต่าง	เปอร์เซ็นต์
	1(10 ตัว)	2(20 ตัว)		
พันธุ์ต้านทาน	10	20	10	100
พันธุ์ไม่ต้านทาน	20	40	20	100
ความแตกต่าง	10	20		
เปอร์เซ็นต์	100	100		

จากตารางข้างบนจะเห็นได้ว่า ค่าในตารางคำนวณจากจำนวนการทำลายของแมลง ความแตกต่างระหว่างทรีตเมนต์เดียวกันที่อยู่ต่างบลอกไม่เท่ากัน และความแตกต่างระหว่างบลอกที่รับต่าง ทรีตเมนต์ไม่เท่ากัน แต่เมื่อคำนวณเป็นเปอร์เซ็นต์แล้วจะเท่ากัน ผลของทรีตเมนต์และบลอกแบบนี้เรียกว่า แบบคูณ[16] ซึ่งสามารถทำให้เป็นแบบบวกได้โดยการใช้ค่าล็อก ดังนี้ $\log(XY) = \log X + \log Y$ การกระทำเช่นนี้ต่อข้อมูลเรียกว่า การแปลงข้อมูล ซึ่งจะกล่าวถึงต่อไป

ถ้าหากว่ามีการฝ่าฝืนข้อสันนิษฐานข้อนี้ ค่าวาเรียนซ์ที่คำนวณได้จะสูงขึ้น ทำให้จับความแตกต่างได้น้อย โดยปกติยิ่งจำนวนทรีตเมนต์และบลอกสูงขึ้นเท่าใด วาเรียนซ์ก็จะสูงขึ้นเท่านั้น Tukey (1949) ได้เสนอวิธีการทดสอบข้อมูลว่า องค์ประกอบต่าง ๆ เป็นแบบบวกหรือไม่ โดยการแยกวาเรียนซ์ของความคลาดเคลื่อนออกมาเป็นส่วนของความไม่เป็นบวก ซึ่งมี 1 df วิธีการ ค่อนข้างสลับซับซ้อนดังแสดงในตาราง 10.8.3

การทดลองบางชนิด ผลของทรีตเมนต์จะเพิ่มในแบบคูณ เช่นนี้ย่อมเป็นการฝ่าฝืนข้อกำหนดข้อนี้ และไม่สามารถนำข้อมูลไปวิเคราะห์วาเรียนซ์ได้ทันที ตัวอย่างในตาราง 10.8.1 แสดงผลของทรีตเมนต์ใน 2 แบบ คือแบบบวกและแบบคูณ จะเห็นได้ว่า ในแบบบวกนั้นผลของทรีตเมนต์เท่ากันในทุก บลอก คือ ทรีตเมนต์ A และ B แตกต่างกัน 20 เท่ากัน ทั้งในบลอก 1 และบลอก 2 ในทำนองเดียวกันผลของบลอกเท่ากันทุกทรีตเมนต์คือแตกต่างกัน 60 แต่ในแบบคูณนั้น ผลของทรีตเมนต์ในบลอกที่ 1 และ

2 เพิ่มขึ้นเท่ากับผลคูณของทรีตเมนต์ A กับ 1.2 และผลของ บลอกเพิ่มขึ้นเท่ากับบลอกที่ 1 คูณ 1.5 หรือถ้าคิดเป็นเปอร์เซ็นต์ก็เพิ่มขึ้น 17 และ 50 เปอร์เซ็นต์ตามลำดับ เช่นนี้แสดงว่าผลของทรีตเมนต์และบลอกเป็นแบบคูณ

ตาราง 10.8.1 ผลของทรีตเมนต์และบลอกเป็นแบบบวกและคูณ

ทรีตเมนต์	แบบบวก		ผลต่าง	แบบคูณ		ผลต่าง
	บลอก 1	บลอก 2		บลอก 1	บลอก 2	
A	100	160	60	100	150	50 (50%)
B	120	180	60	120	180	60 (50%)
ผลต่าง	20	20		20 (20%)	30 (20%)	

ผลของทรีตเมนต์แบบคูณอาจเกิดขึ้นในการทดลองเกี่ยวกับแมลง ทั้งนี้เพราะการทำลายจะสัมพันธ์กับจำนวนแมลง ยกตัวอย่างง่าย ๆ เช่น เราทดลองเปรียบเทียบการกัดกินใบข้าวของแมลง 10 และ 20 ตัว ที่เกิดกับข้าวพันธุ์ต้านทานและไม่ต้านทานดังที่กล่าวมาแล้ว โดยที่แมลงกัดพันธุ์ต้านทานตัวละ 1 แผล และไม่ต้านทาน 2 แผล ดังนั้นแมลง 10 ตัวกัดข้าวพันธุ์ต้านทาน 10 แผลไม่ต้านทาน 20 แผล ดังนั้นแมลง 20 ตัวกัดพันธุ์ต้านทาน 20 แผล และพันธุ์ไม่ต้านทาน 40 แผล ซึ่งเป็นอัตราส่วนที่สมเหตุผลดังแสดงในตาราง 10.8.2 จากตารางดังกล่าวนี้เห็นได้ว่า ผลของทรีตเมนต์และบลอกเป็นแบบคูณ คือเพิ่มขึ้นในอัตราคูณ อย่างไรก็ดีการแสดงผลเช่นนี้สามารถทำให้เป็นอัตราบวกได้โดยใช้ค่าล๊อกคือ $\log(XY) = \log X + \log Y$ ซึ่งจะเห็นได้ว่าการแปลงวิธีนี้ทำให้ผลของทรีตเมนต์และบลอกที่อยู่ในอัตราคูณเปลี่ยนมาอยู่ในอัตราบวก เมื่อทำเป็นอัตราบวกแล้วก็สามารถวิเคราะห์ได้ตามปกติ

ตาราง 10.8.2 จำนวนแผลจากการทำลายของแมลง ซึ่งแสดงผลแบบคูณ แต่สามารถกระทำให้เป็นแบบบวกโดยใช้ค่าล๊อก

ทรีตเมนต์	ต้านทาน (บลอก 1)	ไม่ต้านทาน (บลอก 2)	ผลต่าง	ค่าล็อก		ผลต่าง
				บลอก 1	บลอก 2	
แมลง 10 ตัว	10	20	10 (100%)	1.00	1.30	0.30
แมลง 20 ตัว	20	40	20 (100%)	1.30	1.60	0.30
ผลต่าง	10 (100%)	20 (100%)		0.30	0.30	

การทดสอบความไม่เป็นบวก

ในการวิเคราะห์วาเรียนซ์นั้น มีข้อกำหนดที่ว่าผลของทรีตเมนต์และบลอกเป็นแบบบวก คือ ไม่มีปฏิกิริยาระหว่างทรีตเมนต์และบลอก ข้อกำหนดนี้จะเป็นจริงหรือไม่สำหรับข้อมูลที่เราวิเคราะห์ก็อาจทดสอบโดยใช้วิธีของ Tukey (1949) โดยการแยก SSE ออกเป็น 2 ส่วน ส่วนที่ 1 คือ SSE non-addivity ซึ่งมี 1 df และส่วนที่เหลือเป็นพวก additive โดยใช้วิธีการดังนี้

(1) $d_{i}. = \overline{X}_{i}. - \overline{X}..$

$d._{j} = X._{j} - X..$

ทั้งนี้

$\overline{X}_{i}.$ = ค่าเฉลี่ยของทรีตเมนต์ที่ i

$\overline{X}._{j}$ = ค่าเฉลี่ยของบลอกที่ j

$\overline{X}..$ = ค่าเฉลี่ยทั้งหมด

(2) คำนวณ $W_i = \sum X_{ij} d_j$ และ $N = \sum W_i d_i$

(3) คำนวณ $D = (\sum d_i^2)(\sum d_j^2)$

(4) $SS\ (\text{non-additivity}) = \dfrac{N^2}{D}$

จากข้อมูลในตาราง 10.8.3

ตาราง 10.8.3 ข้อมูลใช้ทดสอบความไม่เป็นบวกของผลของทรีตเมนต์และบลอกจากการทดสอบปุ๋ย

ปุ๋ย	ครั้งที่ใส่ 1	2	3	ผลรวม $X_{i}.$	เฉลี่ย $\overline{X}_{i}.$	d_i	$W_i = \sum X_{ij} d_j$
1	2	5	10	17	5.67	- 9.00	113.86
2	2	15	40	57	19.00	4.33	544.56
3	3	20	40	63	21.00	6.33	517.54
4	2	7	30	39	13.00	- 1.67	414.62
รวม $X._{j}$	9	47	120				1590.58
เฉลี่ย $\overline{X}_{ij}$	2.25	11.75	30.00		14.67		
d_j	- 12.42	- 2.92	15.33				

(1) $d_{1.} = 5.67 - 14.67 = -9.00, \ldots, d_{4} = 13.00 - 14.67 = -1.67$

$d_{.1} = 2.25 - 14.67 = -12.42, \ldots, d_{.3} = 30.00 - 14.67 = 15.33$

(2) $W_1 = (2)(-12.42) + \ldots + (10)(15.33) = 113.86$

$W_4 = (2)(-12.42) + \ldots + (30)(15.33) = 414.62$

check $= 1590.58 = (9)(-12.42) + (47)(-2.92) + (120)(15.33)$

(3) $N = \sum W_i d_i = (113.86)(-9.00) + \ldots + (414.62)(-1.67) = 3{,}916.82$

(4) $\sum d_i^2 = (-9.00)^2 + \ldots + (-1.67)^2 = 142.61$

$\sum d_j^2 = (-12.42)^2 + \ldots + (15.33)^2 = 397.79$

$D = (\sum d_i^2)(\sum d_j^2) = 56{,}728.83$

(5) $\text{SS(non} - \text{additivity)} = \dfrac{N^2}{D} = \dfrac{(3{,}916.82)^2}{56{,}728.83} = 270.44$

เมื่อทำการวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 10.8.3 และได้ผลดังตาราง 10.8.4 จากการทดสอบพบว่า ข้อมูลนี้มีความไม่เป็นแบบบวก ซึ่งต้องดำเนินการแปลงข้อมูลต่อไป

ตาราง 10.8.4 ผลการวิเคราะห์วาเรียนซ์ของข้อมูลในตาราง 10.8.3

Sources	df	SS	MS	F
Fertilizers	3	428.00		
Applications	2	1591.17		
Error	6	319.50		
Non-additivity	1	270.44	270.44	27.26**
Additivity	5	49.58	9.92	

10.9 การแปลงข้อมูล

ดังที่ได้กล่าวมาแล้วในตอน 10.8 ว่า ข้อมูลจากการทดลองต้องมีคุณสมบัติสอดคล้องกับข้อกำหนดต่าง ๆ การวิเคราะห์ทางสถิติจึงให้ผลเชื่อถือได้ เช่นมีข้อกำหนดว่าความคลาดเคลื่อนเป็นอิสระแก่กัน มีการกระจายแบบปกติ และมีวาเรียนซ์เท่ากัน เมื่อได้ข้อมูลมาแล้ว การตรวจดูอย่างหยาบ ๆ อาจบอกได้ว่า ข้อมูลมีอาการผิดปกติอย่างไรบ้าง เช่น มีการกระจายแบบปกติ หรืออาจโย้ไปทางใดทางหนึ่ง ถ้าโย้เพียงเล็กน้อยก็ไม่กระทบต่อผลการวิเคราะห์มากนัก แต่ถ้าโย้มาก ๆ การแปลผลการทดลองอาจคลาดเคลื่อนจากความเป็นจริง ในการทดลองเกี่ยวกับปรากฏการณ์บางชนิดพบว่า ค่าเฉลี่ยและวาเรียนซ์มีความสัมพันธ์กัน เพราะฉะนั้นทรีตเมนต์ใดที่ค่าเฉลี่ยต่ำ วาเรียนซ์ก็

ต่ำไปด้วย ตรงกันข้ามกับพวกที่มีค่าเฉลี่ยสูง ซึ่งทำให้วาเรียนซ์สูงได้เช่นกัน ดังนั้นเราไม่อาจกล่าว ได้ว่า ทรีตเมนต์เหล่านี้มีวาเรียนซ์ร่วมกันหรือเท่ากัน เมื่อข้อมูลมีความผิดปกติหรือไม่สอดคล้องกับ ข้อกำหนด เช่นนี้ก็ใช่ว่าจะต้องทิ้งไปเลย แต่ในทางตรงกันข้าม ความผิดปกติอันนั้นคือธรรมชาติของการเกิดข้อมูลชนิดนั้น ทั้งนี้เพราะไม่จำเป็นเสมอไปที่ข้อมูลจะมีการกระจายแบบปกติ แต่อาจกระจายแบบปัวซองที่ค่าเฉลี่ยและวาเรียนซ์เท่ากันก็ได้ แต่เมื่อเรานำข้อมูลเหล่านี้มาวิเคราะห์จะต้องแก้ไข ปรับปรุง ดัดแปลง ฯลฯ โดยวิธีการต่าง ๆ ให้การ กระจายอยู่ในรูปของการกระจายแบบปกติเสียก่อน เพื่อให้สอดคล้องกับข้อกำหนดของวิธีการทางสถิติที่ใช้ ทั้งนี้เพื่อให้ผลการทดลองเป็นที่เชื่อถือได้ วิธีการดัดแปลง ข้อมูลนี้เราเรียกว่า การแปลงข้อมูล[18] คือเปลี่ยนแปลงมาตราวัดจากแบบหนึ่ง (ที่เราวัดเอาไว้) ไปเป็นแบบอื่น ๆ ที่เหมาะสม การแปลงข้อมูลโดยวิธีใดก็ตามจะไม่ทำให้ลำดับความสำคัญ หรือคุณค่าของข้อมูลเปลี่ยนแปลงไปแต่อย่างใด เช่น ข้อมูลมีค่า 25 และ 16 เมื่อแปลงโดยถอดรากสองก็จะได้ 5 และ 4 ซึ่งใกล้กันมากกว่าเดิม แต่ลำดับความสำคัญไม่เปลี่ยนแปลง คือผลของทรีตเมนต์ยังสำคัญมากน้อยเหมือนเดิม จึงเห็นได้ว่าการแปลงข้อมูลไม่ได้ทำให้ลำดับสูงต่ำหรือมากน้อยของข้อมูลเปลี่ยนไป อย่างไรก็ดี ข้อมูลแต่ละแบบมีวิธีการแปลงแตกต่างกันไปตามความเหมาะสม วิธีการแปลงมีดังนี้

ก. วิธีถอดรากที่สอง[19] การแปลงวิธีนี้ใช้กับข้อมูลจากการนับ ข้อมูลที่มีค่าลงตัวต่ำ ๆ เป็นค่าสังเกตที่ไม่ค่อยปรากฏ เช่น จำนวนต้นพืชที่เป็นโรคในแต่ละแปลง จำนวนแมลงในพื้นที่ขนาดหนึ่ง จำนวนการเกิดอุบัติเหตุในถนนแต่ละสาย จำนวนสินค้าที่ชำรุดในการขนส่งเที่ยวหนึ่ง ๆ ข้อมูลเหล่านี้มักกระจายแบบปัวซอง คือมีความเบ้ และค่าเฉลี่ยและวาเรียนซ์เท่ากัน หรือเพิ่มลดไปด้วยกัน คือค่าเฉลี่ยและวาเรียนซ์สัมพันธ์กันอย่างแน่นแฟ้น

การแปลงข้อมูลประเภทนี้ จะทำให้ความห่างของค่าสังเกตแคบลง ซึ่งกระทำโดยวิธีถอดรากที่สองทุก ๆ ค่าก่อนการวิเคราะห์วาเรียนซ์ การถอดรากที่สองเป็นการลดขนาดของวาเรียนซ์และทำให้การกระจายแบบปัวซองเป็นการกระจายแบบปกติ ถ้าให้ค่าสังเกตเป็น X ค่าที่แปลงเป็น X′ ก็หาได้ว่า $X' = \sqrt{X}$ โดยมีข้อสังเกตดังนี้

(1) ถ้าค่าที่สังเกต (X) ต่าง ๆ มีค่าน้อยกว่า 10 และค่าสังเกตบางค่าเป็นศูนย์ Bartlett (1936) แนะนำให้แปลงโดยใช้ $X' = \sqrt{X+0.5}$ คือนำค่าบางค่าไปบวกกับค่าสังเกตที่เป็นศูนย์ อย่างไรก็ดี มีการบวกด้วยค่าอื่นก็มี เช่น $\sqrt{(X+1)}, \sqrt{(X+3/8)}, \sqrt{X}+\sqrt{(X+1)}$ อย่างนี้เป็นต้น

(2) ค่าสังเกตสูงกว่า 10 และไม่มีค่าสังเกตเป็นศูนย์ ก็แปลงทุกค่าโดยใช้ $X' = \sqrt{X}$ ได้ทันที

(3) ถ้าค่าสังเกตมีค่าเกิน 50 ก็ไม่จำเป็นต้องแปลงข้อมูลดังกล่าวนี้

ตัวอย่างผลการแปลงข้อมูลแสดงไว้ในตาราง 10.9.1 และ 10.9.2 ซึ่งเป็นการแปลงโดย ใช้สมการ $X' = \sqrt{(X+0.5)}$ หลังการแปลงค่าทำให้วาเรียนซ์ของทรีตเมนต์ต่าง ๆ ใกล้เคียงกัน โดยเฉพาะอย่างยิ่งทรีตเมนต์ B ซึ่งวาเรียนซ์สูงกว่าทรีตเมนต์ C ถึง 4 เท่าลงมาเหลือเพียง 2 เท่า

ภายหลังการแปลงค่าก็ดำเนินการวิเคราะห์วาเรียนซ์จากข้อมูลแปลงตามปกติ เมื่อทดสอบทางสถิติเพื่อแสดงว่า ค่าเฉลี่ยแตกต่างหรือไม่แตกต่างกันแล้ว ก็แปลงข้อมูล (ค่าเฉลี่ย) กลับมาเสนอข้อมูลในมาตราวัดเดิม คือคืนขนาดของค่าสังเกต โดยใช้วิธียกกำลังสอง เช่น ทรีต เมนต์ที่ 1 ตาราง 10.9.2 มีค่าเฉลี่ย 8.74/4 = 2.185 เมื่อแปลงสู่มาตราเดิมก็หาได้จาก $X = (2.185)^2 - 0.5 = 4.27$ เมื่อทำการวิเคราะห์วาเรียนซ์ก็ได้ผลดังตาราง 10.9.3 พบว่าข้อมูลแปลงให้สัมประสิทธิ์ของความแปรปรวน (CV) ต่ำลง และการเรียงลำดับค่าเฉลี่ยยังเป็นดังเดิม (ตาราง 10.9.5)

ตาราง 10.9.1 จำนวนไส้เดือนฝอยหลังจากการใช้สารเคมี 5 ชนิด

ทรีตเมนต์	บลอก I	บลอก II	บลอก III	บลอก IV	ค่าเฉลี่ย	พิสัย	s_i^2
A	8	4	4	2	4.50	6	6.33
B	9	8	17	8	10.50	9	19.00
C	4	9	5	6	6.00	5	4.67
D	6	6	4	4	5.00	2	1.33
E	2	4	5	3	3.50	3	1.67

ตาราง 10.9.2 ผลจากการแปลงค่าในตาราง 10.9.1 โดย ใช้ $X' = \sqrt{X + 0.5}$

ทรีตเมนต์	บลอก I	บลอก II	บลอก III	บลอก IV	ค่าเฉลี่ย	พิสัย	s_i^2
A	2.92	2.12	2.12	1.58	2.1850	1.34	0.3049
B	3.08	2.92	4.18	2.92	3.2750	1.26	0.3697
C	2.12	3.08	2.35	2.55	2.5250	0.96	0.1678
D	2.55	2.55	2.12	2.12	2.3350	0.43	0.0616
E	1.58	2.12	2.35	1.87	1.9800	0.77	0.1095

การแปลงข้อมูลโดยวิธีถอดรากที่สอง ควรใช้กับข้อมูลที่บันทึกเป็นตัวเลขลงตัวทั่ว ๆ ไป โดยเฉพาะอย่างยิ่งมีค่าตั้งแต่ 10 ถึง 100 แต่ถ้าข้อมูลมีค่าเกิน 50 ขึ้นไป ก็ไม่มีความจำเป็นต้องแปลง อาจใช้กับข้อมูลที่บันทึกเป็นเปอร์เซ็นต์ก็ได้ ซึ่งใช้กับข้อมูลที่มีค่าระหว่าง 0 ถึง 30 เปอร์เซ็นต์ สำหรับเปอร์เซ็นต์ในช่วงอื่น ๆ ให้ใช้การแปลงวิธีอื่นต่อไป

ตาราง 10.9.3 ผลการวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 10.9.1 และ 10.9.2

Sources	df	ข้อมูลดิบ			ข้อมูลแปลง		
		SS	MS	F	SS	MS	F
Blocks	3	15.00	5.00		0.5005	0.1668	
Treatments	4	118.80	29.70	4.24*	3.9604	0.9901	4.68*
Error	12	84.00	7.00		2.5401	0.2117	
Total	19	217.80			7.0010		
		CV(%) = 44.80%			CV = 29.34%		

2. วิธีใช้ค่าล็อก[20] ข้อมูลบางชนิดสามารถแปลงโดยใช้ค่าล็อก การแปลงวิธีนี้เหมาะสำหรับข้อมูลที่ได้จากการนับ ข้อมูลที่มีขนาดแตกต่างกันมากหรือมีพิสัยกว้าง ข้อมูลที่ ค่าเฉลี่ยแตกต่างกันมากและมีวาเรียนซ์เพิ่ม - ลดตามค่าเฉลี่ย เช่น มีค่าตั้งแต่ 100 ถึง 10,000 เป็นต้น นอกจากนั้นเหมาะสำหรับการทดลองที่ผลของทรีตเมนต์และบลอกเป็นแบบคูณ หรือผลของทรีตเมนต์และบลอกเพิ่มขึ้นในเปอร์เซ็นต์ที่เท่ากันดังตัวอย่างในตาราง 10.8.1 ใช้สำหรับข้อมูลที่ขนาดของค่าเฉลี่ยและวาเรียนซ์ มีความสัมพันธ์ไปในทางเดียวกัน หรือข้อมูลที่มีสัมประสิทธิ์ของความแปรปรวน (CV) ค่อนข้างคงที่ การแปลงโดยใช้ค่าล็อกทำให้อัตราคูณเปลี่ยนเป็นอัตราบวก เช่น ตาราง 10.8.2 การแปลงทำให้ช่วงกว้างระหว่างข้อมูลแคบลง ทำให้วาเรียนซ์ของทรีตเมนต์ต่าง ๆ คงที่ หรือเท่ากัน และการกระจายเป็นแบบปกติ

ในการแปลงข้อมูลโดยใช้ค่าล็อก ให้นำค่าที่สังเกตมาแปลงโดยใช้สมการ $X' = \log X_i$ ถ้ามีค่าสังเกตบางค่าเป็นศูนย์ก็อาจแปลงโดยใช้ $X'=\log(X_i + 0.5)$ หรือ $X'=\log(X_i + 1)$ คือนำค่าบางค่าไปบวก เมื่อแปลงเรียบร้อยแล้วก็ทำการวิเคราะห์ข้อมูลโดยใช้ค่าแปลงนั้น ทำการทดสอบทางสถิติ เช่น เมื่อจะเปรียบเทียบค่าเฉลี่ยก็กระทำโดยใช้ทั้งค่าเฉลี่ยที่เป็นค่าล็อก และค่า lsd หรือ DMRT ที่คำนวณมาจาก MS error ที่วิเคราะห์ได้มาโดยใช้ค่าล็อก แต่เมื่อจะเสนอข้อมูลในรายงานก็แปลงค่าเฉลี่ยกลับโดยใช้ตารางคืนค่าล็อก[21]

ตัวอย่างการแปลงข้อมูลแสดงไว้ในตาราง 10.9.4 ซึ่งจะเห็นได้ว่า พิสัยของทรีตเมนต์ต่าง ๆ ลดลง เมื่อนำมาวิเคราะห์วาเรียนซ์ก็ได้ผลดังตาราง 10.9.5 ซึ่งเห็นได้ว่าสัมประสิทธิ์ของความปรวนแปรลดลง และตัวอย่างการเสนอผลการเปรียบเทียบค่าเฉลี่ยแสดงไว้ในตาราง 10.9.6

3. วิธีอาร์คไซน์หรือแองกูลาร์ [22] การแปลงโดยวิธีอาร์คไซน์ใช้กับข้อมูลที่ได้จากการนับที่เสนอเป็นเปอร์เซ็นต์หรืออัตราส่วน เช่น เปอร์เซ็นต์ความงอก เปอร์เซ็นต์ของต้นพืชที่เป็นโรค ฯลฯ ตัวอย่างเช่น ในการสำรวจพืชพันธุ์หนึ่ง 100 ต้น พบต้นเป็นโรค 38 ต้น ดังนั้นอัตราส่วนต้นเป็นโรคเท่ากับ 0.38 หรือ 38 เปอร์เซ็นต์

ตาราง 10.9.4 ผลจากการแปลงค่าในตาราง 10.9.1 โดยใช้ $X'=\log(X+1)$

ทรีตเมนต์	บลอก				ค่าเฉลี่ย	พิสัย	วาเรียนซ์
	I	II	III	IV			
A	0.95	0.70	0.70	0.48	0.7075	0.47	0.0369
B	1.00	0.95	1.26	0.95	1.0400	0.31	0.0221
C	0.70	1.00	0.78	0.85	0.8325	0.30	0.0162
D	0.85	0.85	0.70	0.70	0.7750	0.15	0.0075
E	0.48	0.70	0.78	0.60	0.6400	0.30	0.0168

ตารางที่ 10.9.5 ผลจากการวิเคราะห์วาเรียนซ์ของข้อมูลในตาราง 10.9.4

Sources	df	SS	MS	F
Blocks	3	0.0510	0.0170	<1
Treatments	4	0.3718	0.0929	4.51*
Error	12	0.2474	0.0206	
Total	19	0.6702		

CV = 16.06%

ตารางที่ 10.9.6 การเปรียบเทียบค่าเฉลี่ยโดยวิธี DMRT จากข้อมูลดิบในตาราง 10.9.1 และข้อมูลแปลงโดยการถอดรากสอง (ตาราง 10.9.2) และใช้ค่าลอก (ตาราง 10.9.4)

ทรีตเมนต์	ลำดับ	ข้อมูลดิบ	แปลง $X'=\sqrt{X+0.5}$		แปลง $X'=\log(X+1)$	
			X'	$X=[(X')^2-0.5]$	X	$X=(\text{Anti}\log X')-1$
A	4	4.5b	2.19	4.30b	0.7075	4.09b
B	1	10.5a	3.28	10.26a	1.0400	9.11a
C	2	6.0b	2.53	5.90b	0.8325	5.81b
D	3	5.0b	2.34	4.98b	0.7750	4.96b
E	5	3.5b	1.98	3.42b	0.6400	3.37b

ข้อมูลที่บันทึกเป็นอัตราส่วนหรือเปอร์เซ็นต์นี้ กระจายแบบไบโนเมียล หากเมื่อ p ไม่เท่ากับ q จะทำให้การกระจายมีความเบ้ แต่ถ้า p และ q เท่ากันหรือมีค่าใกล้เคียงกัน การกระจายก็ใกล้เคียงการกระจายแบบปกติ การกระจายแบบนี้วาเรียนซ์มีความสัมพันธ์กับค่าเฉลี่ย ทั้งนี้วาเรียนซ์เท่ากับ p(1 - p)/n ซึ่งเห็นว่า วาเรียนซ์จะมีค่าต่ำตรงปลายโค้งกระจาย แต่จะค่อย ๆ คงที่เมื่อ p มีค่าสูงขึ้น ในการแปลงนั้น ข้อมูลที่มีอัตราส่วนอยู่ระหว่าง 0.30 และ 0.70 อัตราส่วนนอกเหนือจากช่วงนี้เมื่อแปลงโดยใช้วิธี อาร์คไซน์ ดังแสดงในตาราง ผ.11 ในการแปลงนั้น ถ้าอัตราส่วนมีค่า 0 ให้แทนด้วย 100 - (1/4n) ถ้าอัตราส่วนมีค่า 100 เปอร์เซ็นต์ ให้แทนด้วย n - (1/4n) ก่อนแปลงค่า ภายหลังแปลงค่าเรียบร้อยแล้ว ก็ดำเนินการวิเคราะห์วาเรียนซ์ และเปรียบเทียบค่าเฉลี่ยได้ตามปกติ เช่น อาจใช้วิธี DMRT แล้วเสนอผล (ค่าเฉลี่ย) ขั้นสุดท้ายโดยใช้มาตราวัดเดิม ดังวิธีการแปลงโดยถอดรากสองและใช้ค่าล็อกดังที่กล่าวมาแล้ว

พึงเข้าใจด้วยว่า ข้อมูลบางชนิด ถึงแม้จะวัดเป็นเปอร์เซ็นต์ ก็ไม่จำเป็นต้องแปลงค่า เช่น เปอร์เซ็นต์โปรตีน เปอร์เซ็นต์ไขมันในเมล็ดของพืช หรือการเสนอผลการทดลองแบบเปอร์เซ็นต์ที่เกิดจากค่าสังเกตขนาดใหญ่ (n > 100) เช่นอัตราส่วนของต้นพืชเป็นโรคที่สำรวจจากจำนวนต้นมากกว่า 100 ต้นต่อแปลง

10.10 แบบฝึกหัด

1. ในการทดลองแบบ RCB จงพิสูจน์ให้เห็นว่าเราอาจแยกความแปรปรวนทั้งหมด (TSS) ออกได้เป็น 3 ส่วน คือความแปรปรวนเนื่องจากทรีตเมนต์ (SSTr), บลอก (SSB) และจากความคลาดเคลื่อน (SSE)

2. การจัดแบ่งหน่วยรองรับการทดลองออกเป็นกลุ่ม ๆ (แต่ละกลุ่มเรียกว่าบลอก) มีความจำเป็นอย่างไร เมื่อจัดเช่นนี้แล้วให้ข้อดีแก่การทดลองอย่างไรบ้าง

3. ในการทดลองเปรียบเทียบปุ๋ย 5 สูตร กับถั่วลิสงพันธุ์ไทนาน 9 ได้ผลผลิตเมล็ด (กก./ไร่) ดังนี้

ปุ๋ย	บลอก			
	I	II	III	IV
ไม่ใส่ปุ๋ย	180	172	168	201
15-15-15 (50 กก./ไร่)	205	192	184	240
15-15-15 (50 กก./ไร่) + ปูนขาว	215	221	192	232
3-9-6	220	198	212	205
3-9-6 + ปูนขาว	216	218	220	197

จากข้อมูลนี้

ก. จงวิเคราะห์วาเรียนซ์และแสดง F-test

ข. จงคำนวณหาความคลาดเคลื่อนของค่าเฉลี่ย ($s_{\bar{x}}$) และความคลาดเคลื่อนของความแตกต่าง ($s_{\bar{d}}$)

4. ในการทดลองเปรียบเทียบสารเคมี 4 ชนิด ที่ช่วยเพิ่มความงอกให้แก่ถั่วเหลือง โดยคลุกแก่เมล็ดถั่วเหลืองก่อนปลูกชนิดละ 5 ตัวอย่าง ตัวอย่างละ 100 เมล็ด แล้วสำรวจจำนวนเมล็ดที่ไม่งอกได้ผลดังนี้

ทรีตเมนต์	บลอก					รวม
	I	II	III	IV	V	
Check	8	10	12	13	11	54
Arasan	2	6	7	11	5	31
Spergon	4	10	9	8	10	41
Semesan, Jr.	3	5	9	10	6	33
Fermate	9	7	5	5	3	29
รวม	26	38	42	47	35	188

จงแปลงข้อมูลโดยใช้วิธีการที่เหมาะสมแสดงการวิเคราะห์วาเรียนซ์และตารางวิเคราะห์วาเรียนซ์

5. ในการใช้ยากำจัดแมลงฉีดพืชชนิดหนึ่ง ใช้แผนการทดลองแบบ RCB จำนวน 4 บลอก โดยนับจำนวนตัวหนอนของแมลงหลังจากฉีดยา ดังนี้

ทรีตเมนต์	บลอก			
	I	II	III	IV
ยา ก.	9	12	0	1
ยา ข.	2	8	5	1
ไม่ฉีดยา	35	28	2	15
ยา ค.	1	0	0	0

จงแปลงข้อมูลดังกล่าวโดยใช้วิธีที่เหมาะสม แล้วทำการวิเคราะห์วาเรียนซ์

คำในบท

(1) experimental unit, (2) block, (3) randomized complete block design, (4) replication, (5) additive, (6) mixed model (model III), (7) correction factor, (8) effective number of replication, (9) relative efficiency, (10) independence of experimental error, (11) normal distribution of experimental error, (12) homogeneity of experimental error, (13) additivity of experimental error, (14) goodness-of-fit test, (15) pooled variance, (16) multiplicative, (17) logarithm, (18) transformation, (19) square root transformation, (20) logarithm transformation, (21) anti-logarithm, (22) arc sine or angular transformation.

บทที่ 11
แผนการทดลองแบบลาตินสแควร์

11.1 คำนำ

ในการทดลองเกี่ยวกับพืช อาจพบว่าความอุดมสมบูรณ์ของดินที่ใช้ทดลองมีแนวตั้งฉากกัน คือมีทั้งแนวเหนือ-ใต้ ตะวันออก-ตะวันตก หรือในพื้นที่บุกเบิกใหม่ ๆ เราไม่ทราบแน่ชัดว่าความอุดมสมบูรณ์ของดินอยู่ในแนวใด เมื่อพื้นที่ทดลองมีลักษณะเช่นนี้ ก็ควรใช้แผนการทดลองที่เหมาะสมที่สามารถแยกความคลาดเคลื่อนเนื่องจากความไม่สม่ำเสมอของพื้นที่ในแนวตั้งและแนวนอนออกมาเสียต่างหากจากปัญหาอื่น ๆ ในการทดลอง

ในพื้นที่ซึ่งเคยปลูกพืชหลายชนิดเป็นลำดับเรียงกัน เช่น แปลงแรกปลูกข้าวไร่ แปลงถัดไปปลูกถั่ว และแปลงต่อไปปลูกข้าวโพดหวาน ตามลำดับ สมมุติว่าลำดับนี้อยู่ในแนวนอน (จากซ้ายไปขวา) แต่ความอุดมสมบูรณ์ของดินแปลงนี้แตกต่างกันในแนวตั้ง เมื่อจะทดลองพืชอื่นในพื้นที่นี้เป็นแปลงใหญ่ครอบคลุมทั้งสามแปลง แผนการทดลองที่ควรใช้คือ ให้มีบลอกทั้งแนวตั้งและแนวนอน

การทดลองโดยใช้กระถางในเรือนเพาะชำ การวางกระถางให้ชิดผนังทั้งสองด้าน (แนวมุมฉาก) อาจทำให้กระถางต่าง ๆ ได้รับแสง หรือสภาพแวดล้อมบางอย่างผิดกับกระถางที่วางอยู่ด้านในการทดลองเช่นนี้แต่ละแถวในแนวตั้งและแนวนอนควรมีครบทุกทรีตเมนต์ จึงควรจัดบลอกทั้งแนวตั้งและแนวนอนนั่นเอง

แผนการทดลองที่มีการจัดบลอกทั้งแนวตั้งและแนวนอนนี้ เรียกว่าแผนการทดลองแบบลาตินสแควร์[1]

11.2 แผนการทดลองแบบลาตินสแควร์

แผนการทดลองแบบลาตินสแควร์ คือแผนการทดลองที่มีการจัดบลอกไว้ทั้งสองทาง ให้อยู่ในแนวตั้งฉากกัน คือจัดบลอกทั้งในแนวตั้งและแนวนอน ซึ่งเรียกว่าสดมภ์[2] และแถว[3] ตามลำดับ บลอกในสองทางนี้มีครบทุกทรีตเมนต์ จึงเรียกว่าเป็นซ้ำ[4] หรือเป็นบลอกที่สมบูรณ์[5] การจัดบลอกแบบนี้ทำให้สามารถแยกความคลาดเคลื่อนในการทดลองที่เกิดในทั้งสองแนว ออกมาจากความคลาดเคลื่อนที่ไม่ทราบสาเหตุ ทำให้ผลการทดลองถูกต้องยิ่งขึ้น

จากรายละเอียดที่กล่าวมาแล้วนี้ทำให้เราทราบว่า การทดลองแบบลาตินสแควร์มีการจัดทรีตเมนต์ในแนวตั้งฉากกัน เช่น มี 4 ทรีตเมนต์ (A, B, C และ D) อาจมีรูปร่างของแปลงทดลองดังนี้

A	D	B	C
D	C	A	B
C	B	D	A
B	A	C	D

ซึ่งเห็นว่าแต่ละแถวและสดมภ์มีครบทุกทรีตเมนต์ จึงเรียกว่าเป็นบลอก แต่ถ้าความอุดมสมบูรณ์วิ่งไปในทางเดียวกัน แล้วค่อยลดลงเป็นขั้น ๆ อาจวางทรีตเมนต์ตามรูปดังนี้ก็ได้

A D B C	D C A B	C B D A	B A C D

ขนาดของการทดลองแบบลาตินสแควร์ขึ้นอยู่กับจำนวนทรีตเมนต์ คืออาจมีขนาดเล็กที่สุดตั้งแต่ 2 x 2[6] ขนาดใหญ่ที่สุดไม่ควรเกิน 12 x 12[7] ที่นิยมกันคือขนาดตั้งแต่ 5 x 5 ถึง 10 x 10 เพราะถ้ามีขนาดเล็กกว่า 5 ทรีตเมนต์ จำนวน df ของ error ก็จะน้อยเกินไป

วิธีการสุ่ม

การสุ่มในลาตินสแควร์สามารถเริ่มต้นการสุ่มจากแผนการทดลองมาตรฐาน คือสุ่มจาก standard latin square แผนการทดลองมาตรฐาน เป็นแผนที่จัดวางแถวแรกไว้ตามลำดับตัวอักษรหรือตัวเลข เช่น A, B, C, D หรือ 1, 2, 3, 4 แถวที่สองมีลำดับเหมือนแถวแรก แต่เลื่อนตัวอักษรหรือตัวเลขไปทางขวา 1 ลำดับ คือเริ่มต้นจาก B หรือ 2 แล้วเรียงลำดับต่อๆ ไป หรือลำดับที่เรียงต่อไปเป็นอย่างอื่นที่คุมโดยแถวแรก เช่น แผนการทดลองมาตรฐานของการทดลอง 3 ทรีตเมนต์มี 1 ชุด คือ

A	B	C
B	C	A
C	A	B

การดำเนินการแบบนี้เป็นการวางทรีตเมนต์ตามลำดับ หรือตามเข็มนาฬิกา แต่เมื่อจะนำมาใช้ต้องมีการสุ่มแถวและสดมภ์เสียก่อน ซึ่งสามารถสุ่มได้ทั้งสิ้น 12 แบบ คือ

(1)	(2)	(3)	(4)	(5)	(6)
A B C	A C B	B A C	C A B	B C A	C B A
B C A	B A C	C B A	A B C	C A B	A C B
C A B	C B A	A C B	B C A	A B C	B A C

(7)	(8)	(9)	(10)	(11)	(12)
A B C	A C B	B A C	C A B	B C A	C B A
C A B	C B A	A C B	B C A	A B C	B A C
B C A	B A C	C B A	A B C	C A B	A C B

ทั้งนี้จำนวนแบบหาได้จากสมการ $p!(p-1)! \times$ standard square (p = จำนวนทรีตเมนต์) ใน 12 แบบนี้จะเลือกใช้แบบใดก็ได้

ในการทดลองที่มีขนาดใหญ่ขึ้น การสุ่มจะมีความซับซ้อนขึ้นจากเดิม ทั้งนี้เพราะแผนมาตราฐานจะเพิ่มขึ้น การสุ่มก็สุ่มหาแผนมาตรฐานเสียก่อน ต่อจากนั้นสุ่มวางทรีตเมนต์ต่อไป เช่น ในการทดลองที่มี 4 ทรีตเมนต์มีแผนมาตรฐาน 4 แบบ คือ

(1)	(2)	(3)	(4)
A B C D	A B C D	A B C D	A B C D
B A D C	B C D A	B D A C	B A D C
C D B A	C D A B	C A D B	C D A B
D C A B	D A B C	D C B A	D C B A

เมื่อจะวางแผนการทดลอง อาจจะสุ่มใช้แบบใดแบบหนึ่ง ใน 4 แบบนี้ แต่เมื่อสุ่มได้แบบใดแล้วก็สุ่มอีก 2 ขั้นตอน คือ สุ่มวางสดมภ์และวางแถวดังนี้

สมมุติว่าเราสุ่มได้ standard square แบบที่ 2 ซึ่งมีลำดับของทรีตเมนต์เป็น

แถวที่ 1	A	B	C	D
แถวที่ 2	B	C	D	A
แถวที่ 3	C	D	A	B
แถวที่ 4	D	A	B	

ในขั้นตอนต่อไปก็ทำการสุ่มเพื่อวางแถวและวางสดมภ์ ดังนี้

1. วิธีสุ่มแถว ให้แต่ละแถวดังข้างบนเป็นเหมือนตัวเลขหน่วยเดียวกัน แล้วใช้วิธีการสุ่มเหมือนดังอธิบายมาแล้ว เมื่อสุ่มได้แถวใดแล้วก็วางทรีตเมนต์ในแถวนั้นทั้งชุดลงในบลอก เช่น สุ่มได้แถว 4, 1, 3 และ 2 เป็นลำดับจากบนลงล่าง ก็วางลงในแนวนอนตั้งแถวดังนี้

	สดมภ์ที่			
แถวที่	1	2	3	4
แถวที่ 4	D	A	B	C
แถวที่ 1	A	B	C	D
แถวที่ 3	C	D	A	B
แถวที่ 2	B	C	D	A

2. **วิธีสุ่มสดมภ์** ทำเช่นเดียวกับการสุ่มแถว ขั้นนี้ต้องทำต่อเนื่องจากการสุ่มแถว ก่อนอื่นให้หมายเลขสดมภ์เสียก่อน คือให้ลำดับเป็น 1, 2, 3, 4 แต่ละสดมภ์เท่ากับเลขชุดเดียว เมื่อสุ่มได้สดมภ์ใดก็เขียนลงทั้งสดมภ์ เช่น ได้สดมภ์ 2, 1, 4 และ 3 ก็ได้ผลดังนี้

	สดมภ์ที่			
แถวที่	2	1	4	3
แถวที่ 4	A	D	C	B
แถวที่ 1	B	A	D	C
แถวที่ 3	D	C	B	A
แถวที่ 2	C	B	A	D

ซึ่งเป็นการสิ้นสุดการสุ่ม สดมภ์และแถวกลายเป็นบลอกในแนวตั้งและแนวนอนตามลำดับ ซึ่งอาจให้หมายเลขเสียใหม่ว่าสดมภ์ที่ 1, 2, 3 ,4 และแถวที่ 1, 2, 3, และ 4 นำลำดับแปลงที่จัดได้นี้ไปใช้ในการทดลองต่อไป

สำหรับการทดลองที่มี 5 ทรีตเมนต์ มีแผนมาตรฐานทั้งสิ้น 56 แบบ แต่ Yates (1933) แนะนำให้ใช้ 2 แบบ คือ

(1)						(2)				
A	B	C	D	E		A	B	C	D	E
B	A	D	E	C		B	C	D	E	A
C	E	A	B	D		C	D	E	A	B
D	C	E	A	B		D	E	A	B	C
E	D	B	C	A		E	A	B	C	D

ในการสุ่มนั้นเขียนเลข 1 ถึง 56 ใส่ลงในภาชนะ ถ้าสุ่มได้เลข 1-50 ให้ใช้แผนมาตรฐานที่ 1 ถ้าสุ่มได้ 51-56 ให้ใช้แผนที่ 2 เพื่อทำการสุ่มวางสดมภ์และแถวดังที่กล่าวมาแล้วต่อไป

ถ้าการทดลองมีจำนวนทรีตเมนต์มากขึ้น เราอาจสร้างแบบมาตรฐานขึ้นมาเอง โดยวิธีการวางทรีตเมนต์ตามลำดับหรือตามเข็มนาฬิกาดังที่กล่าวมาแล้วนั่นเอง หรืออาจสุ่มวิธีอื่นใดก็ได้ที่ทำให้ทรีตเมนต์หนึ่ง ๆ ปรากฏในแต่ละแถวและสดมภ์เพียงครั้งเดียว

11.3 การวิเคราะห์วาเรียนซ์

ในแผนการทดลองแบบลาตินสแควร์นั้น เราแยกบลอกออกเป็น 2 ทาง คือแถวและสดมภ์ จึงมีแหล่งของความปรวนแปรแตกต่างจาก RCB คือมีความปรวนแปรเนื่องจากแหล่งต่าง ๆ ดังนี้ : (1) แถว (บลอกในแนวนอน), (2) สดมภ์ (บลอกในแนวตั้ง), (3) ทรีตเมนต์ และ (4) ความคลาดเคลื่อนในการทดลอง วิธีการคำนวณก็คล้ายกับวิธีที่เคยกล่าวมาแล้วในแผนการทดลองแบบ CRD และ RCB จึงอาจสรุปรายละเอียดการวิเคราะห์วาเรียนซ์ ดังตาราง 11.3.1 ซึ่งให้ t เป็นจำนวนทรีตเมนต์ ทั้งนี้ sum of squares ต่าง ๆ ให้คำย่อดังนี้

TSS = total sum of squares

SSR = row sum of squares

SSC = column sum of squares

SSTr = treatment sum of squares

SSE = error sum of squares

ตาราง 11.3.1 แสดงวิธีการวิเคราะห์วาเรียนซ์ในลาตินสแควร์

Sources	df	SS	MS	EMS
Rows	$t-1$	$\sum R_i^2 / t - CF (= SSR)$	SSR/t – 1	$\sigma^2 + t\sigma_r^2$
Columns	$t-1$	$\sum C_j^2 / t - CF (= SSC)$	SSC/t – 1	$\sigma^2 + t\sigma_c^2$
Treatments	$t-1$	$\sum T_r^2 / t - CF (= SSTr)$	SSTr/t – 1	$\sigma^2 + t\sigma_t^2$
Error	$(t-1)(t-2)$	TSS-SSR-SSC-SSTr (=SSE)	SSE/(t – 1) (t – 2)	σ^2
Total	t^2-1	$\sum X_{ij}^2 - CF (= TSS)$		

ในการทดลองเปรียบเทียบน้ำหนักต้นแห้งของถั่วเขียวพันธุ์ต่าง ๆ จำนวน 5 พันธุ์ (A, B, C, และ E) โดยใช้แผนการทดลองแบบลาตินสแควร์ ดังแสดงในตาราง 11.3.2

สมการทางคณิตศาสตร์ของการทดลองแบบลาตินสแควร์ คือ

$$X_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \varepsilon_{ijk} \qquad \ldots(11\text{-}1)$$

เมื่อให้

α = ผลของแถว ; i มีค่าตั้งแต่ 1, 2, ..., t

β = ผลของสดมภ์ ; j มีค่าตั้งแต่ 1, 2, ... , t

τ = ผลของทรีตเมนต์ ; k มีค่าตั้งแต่ 1, 2, ... , t

ทั้งนี้เพราะจำนวนแถวจำนวนสดมภ์ และจำนวนทรีตเมนต์เท่ากัน คือ $\tau = c = t$

การวิเคราะห์วาเรียนซ์

ให้ R, C และ Tr เป็นผลรวมของแถว สดมภ์ และทรีตเมนต์ตามลำดับ แล้วดำเนินการวิเคราะห์วาเรียนซ์ ดังนี้

ตาราง 11.3.2 น้ำหนักต้นแห้งของถั่วเขียวอายุ 40 วัน (กรัม/ต้น)

แถว	สดมภ์ 1	2	3	4	5	รวม
1	B : 5	E : 3	A : 4	C : 7	D : 4	23
2	D : 3	A : 5	E : 2	B : 5	C : 5	20
3	E : 3	B : 6	C : 6	D : 3	A : 6	24
4	A : 4	C : 5	D : 3	E : 3	B : 4	19
5	C : 6	D : 4	B : 5	A : 5	E : 4	24
รวม	21	23	20	23	23	110

ผลรวมของพันธุ์ (ทรีตเมนต์)

A	B	C	D	E
24	25	29	17	15

$$\begin{aligned} CF &= (\sum X_{ij})^2 / t^2 \\ &= (110)^2 / 25 = 484.0 \\ TSS &= \sum X_{ij}^2 - CF \\ &= 5^2 + 3^2 + \ldots + 4^2 - 484.0 = 38.0 \\ SSR &= (\sum R^2)/t - CF \\ &= (23^2 + 20^2 + \ldots + 24^2)/5 - 484.0 = 4.4 \end{aligned}$$

$$SSC = (\sum C^2)/t - CF$$
$$= (21^2 + 23^2 + \ldots + 23^2)/5 - 484.0 = 1.6$$
$$SSTr = (\sum Tr^2)/t - CF$$
$$= (24^2 + 25^2 + \ldots + 15^2)/5 - 484.0 = 27.2$$
$$SSE = TSS - SSR - SSC - SSTr$$
$$= 38.0 - 4.4 - 1.6 - 27.2 = 4.8$$

แล้วนำผลการวิเคราะห์ลงตารางวาเรียนซ์ดังตาราง 11.3.3 ต่อไป

ตาราง 11.3.3 ผลการวิเคราะห์วาเรียนซ์ของข้อมูลในตาราง 10.3.2

Sources	df	SS	MS	F	F(table) 5%	F(table) 1%
Rows	4	4.4	1.1	2.75^{ns}	3.26	5.41
Columns	4	1.6	0.4	1.00^{ns}	3.26	5.41
Treatments	4	27.2	6.8	17.00**	3.26	5.41
Error	12	4.8	0.4			
Total	24	38.0				

** แตกต่างทางสถิติในระดับ 1 เปอร์เซ็นต์, ns = not significant ; CV (%) = 14.37

ถ้าหากจะเปรียบเทียบค่าเฉลี่ยของทรีตเมนต์ ก็อาจหาได้ว่า $s_{\bar{d}} = \sqrt{(2\ MSE/t)}$ และ $s_{\bar{X}} = \sqrt{(MSE/t)}$ ซึ่งนำไปใช้ประโยชน์ดังแสดงในบทที่ 12 ต่อไป

ค่าสูญหาย ในกรณีที่มีค่าสูญหายในลาตินสแควร์ ก็คำนวณโดยใช้สมการ

$$\hat{X} = \frac{t(R + C + T) - 2G}{(t-1)(t-2)} \quad \ldots(11\text{-}2)$$

ทั้งนี้ให้

$\hat{X}$ = ค่าสูญหาย

R = ผลรวมของแถวที่มีค่าสูญหาย

C = ผลรวมของสดมภ์ที่มีค่าสูญหาย

T = ผลรวมของทรีตเมนต์ที่มีค่าสูญหาย

G = ผลรวมทั้งหมด

t = จำนวนทรีตเมนต์

เมื่อใช้ค่าสูญหายในการวิเคราะห์วาเรียนซ์ จำนวน df ของ TSS และ SSE จะลดลงเท่ากับจำนวนค่าสูญหายที่เกิดขึ้น ในขณะเดียวกันเมื่อมีการใช้ค่าสูญหายต้องปรับ SSTr โดยใช้ค่าที่คำนวณจากสมการ

$$[G - R - C - (t-1)\,T]^2 / [(t-1)\,(t-2)]^2 \qquad \ldots(11\text{-}3)$$

เมื่อมีค่าสูญหายเพียงค่าเดียว และต้องการเปรียบเทียบค่าเฉลี่ยของทรีตเมนต์ที่มีค่าสูญหายกับทรีตเมนต์อื่น ๆ อาจคำนวณ $s_{\bar{d}}$ ได้ดังสมการ

$$s_{\bar{d}} = \sqrt{MSE\left[\frac{2}{t} + \frac{1}{(t-1)\,(t-2)}\right]} \qquad \ldots(11\text{-}4)$$

ในกรณีที่มีค่าสูญหายเกิน 1 ค่า การคำนวณความคลาดเคลื่อนมาตรฐานของความแตกต่าง ($s_{\bar{d}}$) นั้น ใช้วิธีการเดียวกันดังที่กล่าวมาแล้วในตอน 10.6 เพื่อหาค่าซ้ำสัมฤทธิ์ คือใช้วิธีการของ Yates (1933) ในการเปรียบเทียบระหว่างทรีตเมนต์ ถ้าทรีตเมนต์ที่เปรียบเทียบกันปรากฏทั้งในแถวและสดมภ์ให้มีค่า 1, ถ้าอีกค่าหนึ่งไม่มีในแถวหรือสดมภ์ให้เท่ากับ 2/3 ถ้าอีกค่าหนึ่งไม่มีทั้งในแถวและสดมภ์ให้เท่ากับ 1/3 และค่าตัวเองไม่มีก็ให้มีค่าเป็น 0 เช่น ในการทดลองแบบ 6 x 6 ซึ่ง B สูญหายไป 2 ค่า และ E สูญหายไป 1 ค่า ดังแสดงในวงเล็บดังนี้

B	E	C	F	D	A
F	D	(B)	E	A	C
C	A	E	D	B	F
(E)	C	A	(B)	F	D
A	F	D	C	E	B
D	B	F	A	C	E

เมื่อตรวจดูผลผลิตของ B และ E ในแถวและคอลัมน์ที่เกี่ยวข้อง จะได้จำนวนซ้ำดังนี้

$$B = \frac{2}{3}+0+1+0+1+1=\frac{11}{3}$$

คือในแถวที่ 1 ในสดมภ์ที่ตรงกับค่า B ไม่มี E จึงให้ค่า 2/3 ในแถวที่ 2 ตรงค่า B ปรากฏว่า B หายไป จึงให้ค่า 0 ในแถวที่ 3 ตรงค่า B มีครบทุกค่าจึงให้ 1 ในแถวที่ 4 ไม่มีค่า B จึงให้ค่า 0 แถวที่ 5 และ 6 มีครบทุกค่าจึงให้ 1 ในทำนองเดียวกัน

$$E = 1+\frac{1}{3}+\frac{2}{3}+0+1+1=4$$

ดังนั้น หาได้ว่าความคลาดเคลื่อนมาตรฐานของความแตกต่าง เมื่อเปรียบเทียบระหว่างค่าเฉลี่ย B และ E

$$s_{\bar{d}} = \sqrt{MSE\left(\frac{3}{11}+\frac{1}{4}\right)}$$

เมื่อต้องการเปรียบเทียบระหว่าง A และ B

$$s_{\bar{d}} = \sqrt{MSE\left(\frac{3}{14}+\frac{1}{4}\right)}$$

ซึ่งเห็นได้ว่า ค่าซ้ำสัมฤทธิ์เปลี่ยนแปลงไปตามคู่ที่เปรียบเทียบ

11.4 ประสิทธิภาพของการจัดแถวและสดมภ์

ความเที่ยงตรงในการใช้แผนการทดลองแบบลาตินสแควร์ เมื่อเปรียบเทียบกับ RCB อาจประมาณได้สองแง่คือ เมื่อใช้แถวเป็นบลอกและเมื่อใช้สดมภ์เป็นบลอก หากใช้สดมภ์เป็นบลอกก็ประมาณ E_e (RCB) ดังนี้

$$E_e(RCB) = [n_cE_c + (n_t + n_e)E_e]/(n_c + n_t + n_e) \qquad \ldots(11\text{-}5)$$

เมื่อให้

E_c, E_e = MSC และ MSE ตามลำดับ

n_c, n_t, n_e = df ของสดมภ์ ทรีตเมนต์ และ error ตามลำดับ

หากใช้แถวเป็นบลอก ก็ใช้สมการคล้ายคลึงกัน เพียงแต่ใช้ E_r และ n_r แทน E_c และ n_c ในสมการนี้

ถ้าหากว่า df ในช่อง error น้อยกว่า 20 ก็อาจปรับประสิทธิภาพสัมพันธ์โดยคูณด้วย $(n_1 + 1)(n_2 + 3)/(n_1 + 3)(n_2 + 1)$ เช่นจากข้อมูลในตาราง 11.3.2 ตารางวิเคราะห์วาเรียนซ์ 11.3.3 หากใช้สดมภ์เป็นบลอกก็หาได้ว่า

$$RE = \frac{E_e(RCB)}{E_e(LS)} \times \frac{(n_1 + 1)(n_2 + 3)}{(n_1 + 3)(n_2 + 1)} \times 100 \qquad \ldots(11\text{-}6)$$

ทั้งนี้ n_1, n_2 คือ df ของ error ในลาตินสแควร์และ RCB ตามลำดับ เมื่อแทนค่าในสมการก็ได้

$$\begin{aligned} RE &= \frac{(4)(0.4) + (4 + 12)(0.4)}{(4 + 4 + 12)(0.4)} \times \frac{(12 + 1)(16 + 3)}{(12 + 3)(16 + 1)} \times 100 \\ &= 1.00 \times 0.97 \times 100 \\ &= 97\% \end{aligned}$$

ถ้าใช้สดมภ์เป็นบลอกก็เพิ่มความเที่ยงตรงในการทดลองประมาณ 3 เปอร์เซ็นต์ คือถ้าเราใช้แผนการทดลอง RCB แล้วใช้สดมภ์ของลาตินสแควร์นี้เป็นบลอก ซึ่งจัดว่าไม่เหมาะสม แต่ถ้าใช้แถวเป็นบลอกก็จะได้ RE 131 เปอร์เซ็นต์

11.5 แผนการทดลองแบบครอสโอเวอร์

การทดลองที่มี 2 ทรีตเมนต์ เช่น A และ B และมีหน่วยทดลองจำนวนน้อย การทดลองก็จะเล็กเกินไป อาจปรับปรุงโดยใช้แผนการทดลองแบบครอสโอเวอร์[8] คือสลับการใช้หน่วยทดลองซึ่งคล้าย ๆ กับแผนการทดลองแบบลาตินสแควร์ คือ ในแต่ละบลอกหรือสดมภ์มีครบทุกทรีตเมนต์และในแต่ละแถวมีจำนวนทรีตเมนต์เท่ากัน เช่น ในการเปรียบเทียบน้ำมันเชื้อเพลิงจากพืช (A) และน้ำมันเบนซิน (B) โดยใช้รถยนต์ 8 คัน แต่ละคันใช้น้ำมันเชื้อเพลิง A หรือ B ก่อนในจำนวนคัน คืออย่างละ 4 คัน เท่ากัน เมื่อใช้น้ำมันเชื้อเพลิง A แล้วเปลี่ยนเป็น B และเมื่อใช้ B แล้วเปลี่ยนเป็น A ดังนี้

สดมภ์ที่ (รถยนต์)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
แถวที่ 1 (ใช้ก่อน)	A	B	B	A	A	B	B	A
แถวที่ 2 (ใช้หลัง)	B	A	A	B	B	A	A	B

จะเห็นว่า รถแต่ละคันเป็นบลอก เราอาจแยกการทดลองนี้ออกเป็นลาตินสแควร์ 4 การทดลอง ดังนี้

1		2		3		4	
A	B	B	A	A	B	B	A
B	A	A	B	B	A	A	B

แล้วอาจวิเคราะห์ได้ 2 แบบ คือแบบครอสโอเวอร์ และแบบลาตินสแควร์ที่มีหลายการทดลอง ดังนี้

ครอสโอเวอร์

Sources		df
Rows	$(r-1)$ =	1
Columns	$(c-1)$ =	7
Treatments	$(t-1)$ =	1
Error	$(t-1)(c-2)$ =	6
Total	$rct-1$ =	15

ลาตินสแควร์

Sources		df
Squares	$(s-1)$	= 3
Rows	$s(t-1)$	= 4
Columns	$s(t-1)$	= 4
Treatments	$(t-1)$	= 1
Error	$(s-1)(t-1)$	= 3
Total	st^2-1	= 15

เห็นได้ว่าแผนการทดลองแบบครอสโอเวอร์ มี df ของ error สูงกว่าลาตินสแควร์ และเหมาะสมกับการทดลองที่หน่วยทดลองมีจำนวนน้อย จึงจัดให้ทรีตเมนต์เกิดเป็นคู่หรือเป็นชุดเพื่อใช้หน่วยทดลองเดียวกันตัวอย่างเช่น

(1) ทดลองอาหาร A และ B กับโค 4 ตัว (ก, ข, ค, ง) โดยสุ่มให้ 2 ตัวได้รับอาหาร A ก่อน ส่วนอีก 2 ตัวให้รับอาหาร B ก่อน เมื่อเก็บข้อมูล เช่น ชั่งน้ำหนักแล้วให้มีช่วงพักโดยให้กินอาหารทั่วๆ ไป สักระยะหนึ่งก่อนให้อาหารสลับ คือ 2 ตัวที่เคยได้รับอาหาร A ให้รับอาหาร B, อีก 2 ตัวที่เคยรับอาหาร B ให้รับอาหาร A ดังนี้

โค	ก	ข	ค	ง
ช่วงแรก	A	B	B	A
ช่วงที่สอง	B	A	A	B

(2) ทดลองเปรียบเทียบยาลดความดันโลหิต A, B กับผู้ป่วย 6 คน โดยให้มีช่วงพักและสลับชนิดของยา

คนไข้	ก	ข	ค	ง	จ	ฉ
ช่วงแรก	A	B	B	A	A	B
ช่วงที่สอง	B	A	A	B	B	A

ตัวอย่าง ในการเปรียบเทียบเครื่องพิมพ์ดีด 2 ชนิด (A, B) โดยใช้พนักงานพิมพ์ดีด 6 คน ให้แต่ละคนได้พิมพ์ทั้ง 2 เครื่อง แต่มีการทิ้งระยะห่าง 1 วัน ผลที่ได้เป็นคำ/นาที ดังนี้

คนที่	1	2	3	4	5	6	รวม
วันแรก	A:32	B:44	A:40	B:40	A:40	B:35	231
วันที่สอง	B:40	A:38	B:43	A:42	B:42	A:35	240
รวม	72	82	83	82	82	70	471

ผลรวมของเครื่องพิมพ์ดีด A = 227, B = 244

วิเคราะห์วาเรียนซ์

เมื่อให้ R, C, Tr เป็นผลรวมของแถว สดมภ์ และทรีตเมนต์ตามลำดับ ให้ r เป็นจำนวนแถว ให้ c เป็นจำนวนสดมภ์ ดังนี้

$$CF = (\sum X_{ij})^2 / rc = (471)^2 / 12 = 18{,}486.75$$

$$TSS = \sum X_{ij}^2 - CF$$

$$= 32^2 + 40^2 + \ldots + 35^2 - 18{,}486.75 = 144.25$$

$$SSR = (\sum R^2)/c - CF$$

$$= (231^2 + 240^2)/6 - 18{,}486.75 = 6.75$$

$$SSC = (\sum C^2)/r - CF$$

$$= (72^2 + 82^2 + \ldots + 70^2)/2 - 18{,}486.75 = 85.75$$

$$SSTr = (\sum Tr^2)/c - CF$$

$$= (227^2 + 244^2)/6 - 18{,}486.75 = 24.08$$

$$SSE = TSS - SSR - SSC - SSTr$$

$$= 144.25 - 6.75 - 85.75 - 24.08 = 27.67$$

แล้วนำผลการวิเคราะห์ตารางวิเคราะห์วาเรียนซ์ดังตาราง 11.5.1

ถ้าต้องการจะเปรียบเทียบระหว่างค่าเฉลี่ยของทรีตเมนต์ก็อาจคำนวณ $s_{\bar{x}}$ และ $s_{\bar{d}}$ ดังนี้

$$s_{\bar{X}} = \sqrt{\frac{MSE}{c}} = \sqrt{\frac{6.92}{6}} = 1.07$$

$$s_{\bar{d}} = \sqrt{\frac{2MSE}{c}} = \sqrt{\frac{2(6.92)}{6}} = 1.52$$

ทั้งนี้ c = จำนวนสดมภ์

ตาราง 11.5.1 ตารางวิเคราะห์วาเรียนซ์แผนการทดลองแบบครอสโอเวอร์

Sources	df	SS	MS	F	F(table)	
					5%	1%
Rows	1	6.75	6.75	0.98^{ns}	7.71	21.20
Columns	5	85.75	17.15	2.48^{ns}	6.25	15.98
Treatments	1	24.08	24.08	3.48^{ns}	7.21	21.20
Error	4	27.67	6.92			
Total	11	144.25				

ns = not significant ; CV(%) = 17.63

ถ้ามีค่าสูญหายก็คำนวณโดยใช้สมการ

$$X = \frac{cC + t(R+T) - 2G}{(t-1)(c-2)} \quad \ldots(11\text{-}7)$$

ทั้งนี้ C, R และ T = ผลรวมของสดมภ์ แถว และทรีตเมนต์ที่มีค่าสูญหายตามลำดับ G = ผลรวมทั้งหมด, t = จำนวนทรีตเมนต์ และ c = จำนวนสดมภ์

ในกรณีที่มีค่าสูญหายก็คำนวณได้ว่า

$$s_{\bar{d}} = \sqrt{MSE\left[\frac{2}{c} + \frac{t}{c(t-1)(c-2)}\right]}$$

11.6 กรีโค-ลาตินสแควร์

ในแผนการทดลองแบบลาตินสแควร์ แต่ละทรีตเมนต์ได้รับการทดลองเท่ากับจำนวนซ้ำ ดังนั้นเปิดโอกาสให้สามารถนำทรีตเมนต์ใหม่จำนวนเท่ากับทรีตเมนต์เดิมเข้าไปฝาก เราอาจเรียกว่าการจัดทรีตเมนต์ดั้งเดิมเป็นการจัดแบบลาติน ส่วนการฝากทรีตเมนต์ใหม่เข้าไปเป็นการจัดแบบกรีก แผนการทดลองที่เกิดจากการจัดแบบนี้เรียกว่า "กรีโค-ลาตินสแควร์[9]" เช่น จากแผนการทดลองลาตินสแควร์มาตรฐานก็อาจจัดแผนการทดลองกรีโค-ลาตินสแควร์มาตรฐานได้ดังนี้

ลาตินสแควร์				กรีโค-ลาตินสแควร์			
A	B	C	D	$A\alpha$	$B\beta$	$C\gamma$	$D\delta$
B	A	D	C	$B\gamma$	$A\delta$	$D\alpha$	$C\beta$
C	D	A	B	$C\delta$	$D\gamma$	$A\beta$	$B\alpha$
D	C	B	A	$D\beta$	$C\alpha$	$B\delta$	$A\gamma$

ซึ่งเห็นได้ว่าแต่ละทรีตเมนต์ที่เป็นลาตินนั้นมีทรีตเมนต์ที่เป็นกรีกปรากฏอยู่ครบ ในการจัดนี้เห็นว่า อักษรลาตินและกรีกแต่ละชนิดจะจับคู่กันเพียงการทดลองละ 1 ครั้งเท่านั้น เรียกว่าเป็นการจัดไปคนละแนวที่ตั้งฉากกัน

ตัวอย่างเช่น A, B, C, D เป็นพืช 4 พันธุ์ ส่วน $\alpha, \beta, \gamma, \delta$ เป็นปุ๋ย 4 ชนิด วิธีการวิเคราะห์วาเรียนซ์คล้ายคลึงกับแผนการทดลองแบบลาตินสแควร์ เพียงแต่เพิ่มทรีตเมนต์ที่สอง (อักษรกรีกเข้าไป) ซึ่งอาจแยก df ได้ดังนี้ 1 – 2

Sources	df
Rows	$r-1 = 3$
Columns	$c-1 = 3$
Treatments (Latin)	$t-1 = 3$
Treatments (Greek)	$g-1 = 3$
Error	$(t-1)(t-3) = 3$
Total	$t^2-1 = 15$

วิธีการวิเคราะห์วาเรียนซ์คล้ายกับแผนการทดลองแบบลาตินสแควร์นั่นเอง คือ Sum of Squares จากแถว คอลัมน์ลาตินและกรีกให้คำนวณจากส่วนที่เกี่ยวข้อง

ตัวอย่าง สมมุติว่าฟาร์มแห่งหนึ่ง ได้ทำการทดลองเปรียบเทียบโปรตีน 4 ระดับ (A, B, C, D) โดยใช้สุกร 4 พันธุ์ (1, 2, 3, 4) และอายุต่าง ๆ กัน 4 ชุด (1, 2, 3, 4) ในอาหารแต่ละชุดได้ทำการผสมไลซีน 4 ระดับ $(\alpha, \beta, \delta, \gamma)$ โดยจัดทรีตเมนต์แบบ Graeco-Latin Squares ผลการทดลองเป็นน้ำหนักเพิ่มของสุกรในช่วงเวลาหนึ่งแสดงในตาราง 11.6.1

ตาราง 11.6.1 น้ำหนักเพิ่มของสุกรที่เลี้ยงโดยอาหารที่ใช้โปรตีนและไลซีนระดับต่าง ๆ กัน

พันธุ์สุกร	อายุของสุกร				รวม
	1	2	3	4	
1	$C\beta = 6$	$B\gamma = 5$	$D\delta = 7$	$A\alpha = 4$	22
2	$B\alpha = 4$	$C\delta = 6$	$A\gamma = 5$	$D\beta = 6$	21
3	$A\delta = 5$	$D\alpha = 6$	$B\beta = 3$	$C\gamma = 8$	22
4	$D\gamma = 5$	$A\beta = 2$	$C\alpha = 9$	$B\delta = 3$	19
รวม	20	19	24	21	84

ทั้งนี้การวิเคราะห์วาเรียนซ์คล้ายคลึงกับการวิเคราะห์ในแผนการทดลองแบบลาตินสแควร์ดังตาราง 11.3.2 ดังนี้

$$CF = \frac{(\sum X_{ij})^2}{t^2} = \frac{84^2}{16} = 441.00$$

$$TSS = \sum X_{ij}^2 - CF$$

$$= 6^2 + 4^2 + 5^2 +.\ + 3^2 - 441.00 = 51.00$$

$$SSR = \frac{\sum R_i^2}{t} - CF$$

$$= (22^2 + 21^2 + 22^2 + 19^2)/4 - 441.00 = 1.50$$

$$SSC = \frac{\sum C_j^2}{t} - CF$$

$$= (20^2 + 19^2 + 24^2 + 21^2)4 - 441.00 = 3.50$$

ผลรวมของโปรตีนและไลซีนเป็นดังนี้

โปรตีน (ลาติน) A = 16, B = 15, C = 29, D = 24 และไลซีน (กรีก) $\alpha = 23$, $\beta = 17$, $\delta = 21$ และ $\gamma = 23$ ดังนั้นค่า SS เหล่านี้ได้แก่

$$SS(Protein) = \frac{\sum Latin^2}{t} - CF$$

$$= (16^2 + 15^2 + 29^2 + 24^2)/4 - 441.00 = 33.50$$

$$SS(Lysine) = \frac{\sum Greek^2}{t} - CF$$

$$= (23^2 + 17^2 + 21^2 + 23^2)/4 - 441.00 = 6.00$$

ส่วน Error SS หาได้โดยใช้การหักลบดังนี้

$$SSE = TSS - SSR - SSC - SS(Latin) - SS(Greek)$$

$$= 51.00 - 1.50 - 3.50 - 33.50 - 6.00 = 6.50$$

แล้วแสดงผลการวิเคราะห์วาเรียนซ์ดังตาราง 15.6.2

ตาราง 15.6.2 การวิเคราะห์วาเรียนซ์ของน้ำหนักเพิ่มของสุกรที่ได้รับโปรตีนและไลซีนระดับต่าง ๆ กัน

Sources	df	SS	MS	F
Rows	3	1.50	0.50	0.23
Columns	3	3.50	1.17	0.54
Latins (Protein)	3	33.50	11.17	5.14
Greeks (Lysine)	3	6.00	2.00	0.92
Error	3	6.50	2.17	
Total	15	51.00		

จากการวิเคราะห์วาเรียนซ์พบว่า ไม่มีรายการใดแตกต่างในทางสถิติ ทั้งนี้สาเหตุประการหนึ่งเนื่องมาจาก df ของ SSE ค่อนข้างต่ำ เนื่องจากได้แยกส่วนหนึ่งไปเป็นของความปรวนแปรเนื่องจากไลซีน

ในกรีโค-ลาตินสแควร์อย่างน้อยต้องมีขนาด 4 x 4 จึงมี df เพื่อประมาณ MSE ถ้าเพิ่มขนาดการทดลองเป็น 5 x 5, 6 x 6, 7 x 7, ฯลฯ เราก็สามารถเพิ่มชนิดของทรีตเมนต์เข้าไปอีกก็ได้ ดังนั้นจึงสรุปได้ว่าจำนวนชนิดของความแปรปรวน (Rows, Columns,...) ที่แยกได้จะเท่ากับจำนวน ทรีตเมนต์ที่ใช้ในการทดลองนั้น การทดลองที่เรานำทรีตเมนต์เข้าไปฝากเพิ่มจากอักษรกรีกเรียกว่าไฮเปอร์สแควร์[10] อย่างไรก็ดี ในแง่ปฏิบัติไม่ค่อยมีการใช้แผนการทดลองนี้

11.7 แบบฝึกหัด

1. จากข้อมูลดังต่อไปนี้

แถว	สดมภ์ 1	2	3	4	5	รวม	รวมทรีตเมนต์	
1	4.9	6.4	3.3	9.5	11.8	35.9	A	34.2
2	9.3	4.0	6.2	5.1	5.4	30.0	B	32.3
3	7.6	15.4	6.5	6.0	4.6	40.1	C	65.6
4	5.3	7.6	13.2	8.6	4.9	39.6	D	39.8
5	9.3	6.3	11.8	15.9	7.6	50.9	E	24.6
รวม	36.4	39.7	41.0	45.1	34.3	196.5		196.5

ก. จงแสดงแบบจำลองทางคณิตศาสตร์

ข. แสดงการวิเคราะห์วาเรียนซ์ และแสดงตารางวิเคราะห์วาเรียนซ์

ค. คำนวณหา $s_{\bar{x}}$ และ $s_{\bar{d}}$

ง. จงคำนวณประสิทธิภาพสัมพันธ์กับ RCB เมื่อใช้แถวเป็นบลอก

2. ในการทดลองเปรียบเทียบน้ำหนักต้นแห้งของถั่วเขียว 2 พันธุ์ (A, B) แต่ละพันธุ์ปลูกในกระถางเดียวกัน โดยมีลำดับก่อนหลัง ผลการทดลองเป็นกรัม/ต้น ดังนี้

กระถางที่	1	2	3	4	5	6	7	8
ปลูกก่อน	A : 7	B : 6	B : 5	A : 6	B : 4	A : 5	A : 6	B : 7
ปลูกหลัง	B : 5	A : 6	A : 5	B : 5	A : 7	B : 5	B : 6	A : 5

จงทำการวิเคราะห์วาเรียนซ์ และทดสอบว่าพันธุ์และการปลูกก่อนหลังแตกต่างกันหรือไม่

3. ในการทดลองเปรียบเทียบผลของอาหาร 3 ชนิด (A, B และ C) ต่อการให้นมของโค การทดลองใช้โค 3 ตัว และทดลอง 3 ช่วงการให้นม ได้น้ำหนักนม (กก./6 เดือน) ดังนี้

ช่วงการให้นม	โคตัวที่ 1	2	3
1	A : 276	B : 402	C : 427
2	B : 235	C : 494	A : 348
3	C : 384	A : 323	B : 378

จงวิเคราะห์วาเรียนซ์โดยไม่จำเป็นต้องคำนึงถึงผลตกค้างของอาหารแต่ละชนิด

4. ในการเปรียบเทียบผลผลิตของถั่วเขียว 10 พันธุ์ ได้ผลผลิต (กรัม/ตารางเมตร)

แถว	สดมภ์						รวม
	1	2	3	4	5	6	
1	B : 220	F : 98	D : 149	A : 92	E : 282	C : 169	1,010
2	A : 74	E : 238	B : 158	C : 228	F : 48	D : 188	934
3	D : 118	C : 279	F : 118	E : 278	B : 176	A : 65	1,034
4	E : 295	B : 222	A : 54	D : 104	C : 213	F : 163	1,051
5	C : 187	D : 90	E : 242	F : 96	A : 66	B : 122	803
6	F : 90	A : 124	C : 195	B : 109	D : 79	E : 211	808
รวม	984	1,051	916	907	864	918	5,640

จงแสดงการวิเคราะห์วาเรียนซ์ และแสดงตารางวิเคราะห์วาเรียนซ์

คำในบท

(1) Latin square design, (2) column, (3) row, (4) replication, (5) complete block, (6) 2 x 2 Latin square, (7) 12 x 12 latin square, (8) cross or change-over design, (9) Graeco-Latin (triple square grouping) design. (10) hyper-square.

บทที่ 12
การเปรียบเทียบค่าเฉลี่ย

12.1 คำนำ

การวิเคราะห์วาเรียนซ์เป็นเพียงขั้นหนึ่งของการวิเคราะห์ข้อมูล เป็นการตรวจสอบดูว่าทรีตเมนต์มีความแตกต่างกันหรือไม่ แต่การทดสอบมิได้สิ้นสุดเพียงแค่นั้น ถึงแม้ F - test แสดงว่าทรีตเมนต์มีความแตกต่างกันในทางสถิติ เราก็ไม่ทราบว่าทรีตเมนต์ใดแตกต่างกันบ้าง เมื่อการวิเคราะห์วาเรียนซ์แสดงนัยสำคัญก็อาจบอกได้เพียงหยาบ ๆ ว่าทรีตเมนต์ที่มีค่าเฉลี่ยสูงสุดและต่ำสุดควรแตกต่างกัน แต่ในการทดลองส่วนมากเราต้องการเปรียบเทียบระหว่างทรีตเมนต์ใดทรีตเมนต์หนึ่งโดยเฉพาะ ดังนั้นรายละเอียดในการเปรียบเทียบค่าเฉลี่ยในรูปแบบต่าง ๆ ก็มีความสำคัญไม่ยิ่งหย่อนกว่าการวิเคราะห์วาเรียนซ์ วิธีการเปรียบเทียบค่าเฉลี่ยมีอยู่หลายวิธี แต่ละวิธีมีวัตถุประสงค์ในการใช้แตกต่างกัน ซึ่งเราจะได้เรียนรู้ในบทนี้

12.2 วิธีแอลเอสดี

การเปรียบเทียบค่าเฉลี่ยโดยวิธี LSD [1] เป็นวิธีที่ง่าย และมีการใช้กันแพร่หลายที่สุด เป็นวิธีที่ใช้เปรียบเทียบเพื่อหาความแตกต่างระหว่างทรีตเมนต์ที่จัดเป็นคู่ ๆ โดยเฉพาะอย่างยิ่งเหมาะสมกับการเปรียบเทียบระหว่างค่าเฉลี่ยบางคู่ที่เราตั้งใจจะหาความแตกต่าง เช่น ระหว่างทรีตเมนต์ต่าง ๆ กับทรีตเมนต์เปรียบเทียบหรือทรีตเมนต์มาตรฐาน[2] เป็นต้น

การเปรียบเทียบโดยใช้ LSD คือการที่นำเอาความแตกต่างระหว่างทรีตเมนต์ไปเปรียบเทียบกับค่า LSD ที่คำนวณได้ ถ้าความแตกต่างนั้นสูงกว่าค่า LSD ก็ถือว่าค่าเฉลี่ยคู่นั้นแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ทั้งนี้สมการใช้หาค่า LSD คือ

$$LSD_\alpha = (t_\alpha)(s_{\bar{d}}) \qquad \ldots(12\text{-}1)$$

ทั้งนี้ t_α คือค่า t จากตาราง เป็นค่า two-tailed เปิดตาราง t โดยใช้ df ของ error ส่วน $s_{\bar{d}}$ คือความคลาดเคลื่อนมาตรฐานของความแตกต่าง ซึ่งหาได้โดยใช้สมการ

$$s_{\bar{d}} = \sqrt{\frac{2s^2}{n}}$$

ทั้งนี้ s^2 คือ MSE หรือ MS error และ n คือจำนวนซ้ำ (จำนวนบลอก) ซึ่งเป็นจำนวนครั้งที่ทดลองแต่ละทรีตเมนต์ ในกรณีที่การทดลองมีค่าสังเกตไม่เท่ากันก็ใช้สมการ

$$s_{\bar{d}} = \sqrt{s^2\left[\frac{1}{n_i}+\frac{1}{n_j}\right]} \qquad \ldots(12\text{-}2)$$

เมื่อ n_i และ n_j คือจำนวนซ้ำหรือจำนวนครั้งที่ทดลองทรีตเมนต์ที่ i และที่ j ซึ่งเปรียบเทียบกันตามลำดับ

ความจริงแล้วการใช้ค่า LSD มิใช่เรื่องใหม่แต่ประการใด ทั้งนี้เพราะเราได้คุ้นเคยกับสมการ $t = (\bar{X}_1 - \bar{X}_2)/s_{\bar{d}}$ มาแล้วในบทที่ 4 เมื่อทำการคูณไขว้ก็จะได้ $(t)(s_{\bar{d}}) = \bar{X}_1 - \bar{X}_2$ ดังนั้นจึงเห็นได้ว่า $(t)(s_{\bar{d}})$ นี่เองที่เรียกว่า LSD

ตัวอย่าง เช่น ในการทดลองเปรียบเทียบวิธีการปราบวัชพืชของถั่วเหลืองในตอน 9.3 (ตาราง 9.3.1) ผลจากการวิเคราะห์วาเรียนซ์ดังตาราง 9.3.2 ปรากฏว่า ทรีตเมนต์แตกต่างกันอย่างมีนัยสำคัญยิ่ง ในการใช้ LSD นั้น เราควรตั้งใจมาก่อนว่าต้องการเปรียบเทียบระหว่างทรีตเมนต์ใดบ้าง ในตัวอย่างนี้ถ้าเปรียบเทียบทุกทรีตเมนต์กับทรีตเมนต์ที่ไม่มีการปราบวัชพืช ซึ่งจัดเป็นทรีตเมนต์มาตรฐาน ก็มีขั้นตอนในการเปรียบเทียบดังนี้

(1) หาความแตกต่างระหว่างค่าเฉลี่ยของทรีตเมนต์ที่ต้องการเปรียบเทียบ

(2) คำนวณหาค่า LSD ระดับที่ต้องการ คือระดับ 0.05 หรือ 0.01 ในตัวอย่างนี้มีค่าต่าง ๆ ดังนี้

$$n = 4, t_{0.05}\ df15 = 2.131,\ t_{0.01}\ df15 = 2.947,\ MSE = 11.88$$

ดังนั้น

$$LSD_{0.05} = 2.131\sqrt{\frac{2(11.88)}{4}} = 5.18 \text{ กก.}$$

$$LSD_{0.05} = 2.947\sqrt{\frac{2(11.88)}{4}} = 7.16 \text{ กก.}$$

(3) นำค่า LSD นี้ไปเปรียบเทียบกับความแตกต่างระหว่างทรีตเมนต์ต่าง ๆ เป็นคู่ ๆ ที่ต้องการหรือเปรียบเทียบกับทรีตเมนต์มาตรฐาน (ไม่กำจัดวัชพืช)

เช่น ระหว่างทรีตเมนต์มาตรฐาน (A) กับทรีตเมนต์ที่กำจัดวัชพืชเมื่ออายุ 30 วัน (B) มีความแตกต่างดังนี้

$$B - A = 23.00 - 14.25 = 8.75^{**}$$

เมื่อได้เปรียบเทียบกับทรีตเมนต์มาตรฐาน ครบทุกชุดแล้ว ก็นำผลลงตารางดังตารางที่ 12.2.1 พบว่ามีความแตกต่างในระดับ 0.01 ทุกชุด ในตารางนี้เราอาจเรียงค่าเฉลี่ยจากสูงไปหาต่ำ แล้วนำเอา

ทรีตเมนต์มาตรฐานไว้ในบรรทัดแรกหรือบรรทัดสุดท้าย แต่ก็ไม่จำเป็นต้องเรียงลำดับเช่นนี้เสมอไป ในตารางดังกล่าวนี้เราอาจนำเอาเครื่องหมายดอกจันมาไว้ที่ค่าเฉลี่ยก็ได้ ค่าเฉลี่ยใดมีดอกจันก็แสดงว่าแตกต่างจากค่า LSD ส่วนจะแตกต่างระดับใดนั้นก็ให้ดูที่จำนวนดอกจัน

ตาราง 12.2.1 การเปรียบเทียบค่าเฉลี่ยระหว่างทรีตเมนต์ต่าง ๆ กับทรีตเมนต์เปรียบเทียบ (A) จากข้อมูลในตาราง 9.3.1

การทดลอง	ค่าเฉลี่ย (กก.)	ความแตกต่างจาก control (กก.)
A. ไม่กำจัดวัชพืช	14.25	
B. กำจัดเมื่ออายุ 30 วัน	23.00	8.75**
C. กำจัดเมื่ออายุ 30, 60 วัน	28.75	14.50**
D. กำจัดเมื่ออายุ 25, 50, 75 วัน	32.00	17.75**
E. ใช้สารเคมี	30.25	16.00**

** = แตกต่างทางสถิติในระดับ 1 เปอร์เซ็นต์ จากทรีตเมนต์เปรียบเทียบ (ไม่กำจัดวัชพืช)

การใช้ LSD เมื่อค่าสังเกตของแต่ละทรีตเมนต์ไม่เท่ากัน

ในแผนการทดลองแบบ CRD ที่ค่าสังเกตของแต่ละทรีตเมนต์มีจำนวนไม่เท่ากัน ดังตัวอย่างในเรื่องของอาหารไก่ในตอน 9.4 ซึ่งเห็นได้ว่าอาหารโปรตีน A, B, C และ D มีค่าสังเกต (n) เท่ากับ 9, 10, 11 และ 7 ค่าตามลำดับ สมมุติว่าทรีตเมนต์ D เป็นทรีตเมนต์มาตรฐาน ในการเปรียบเทียบระหว่างทรีตเมนต์ต่าง ๆ กับทรีตเมนต์มาตรฐานกระทำดังนี้

(1) หาความแตกต่างระหว่างค่าเฉลี่ยของทรีตเมนต์ต่าง ๆ กับทรีตเมนต์มาตรฐาน

(2) คำนวณหา LSD โดยใช้สมการ (12-1) เปิดตาราง t ระดับ 0.05 และ 0.01 โดยใช้ df 33 (ตาราง 9.4.2) เนื่องจากตาราง t ไม่มีค่าที่ df ดังกล่าว จึงเปิดตารางที่ df 30 ได้ค่าเท่ากับ 2.042 และ 2.750 ตามลำดับ

(1) เมื่อเปรียบเทียบ A กับ D

$$LSD = (t_\alpha)(s_{\bar{d}})$$

ทั้งนี้

$$s_{\bar{d}} = \sqrt{s^2\left[\frac{1}{n_i} + \frac{1}{n_j}\right]}$$

MSE (ตาราง 9.4.2) = 85.80 , $n_i = 9$ และ $n_j = 7$

ดังนั้น

$$s_{\bar{d}} = \sqrt{85.80\left[\frac{1}{9}+\frac{1}{7}\right]} = 4.67$$

ดังนั้น

$LSD_{0.05} = (2.042)\,(4.67) = 9.54$

$LSD_{0.01} = (2.750)\,(4.67) = 12.84$

(2) เปรียบเทียบ B กับ D

$$s_{\bar{d}} = \sqrt{85.80\left[\frac{1}{10}+\frac{1}{7}\right]} = 4.56$$

$LSD_{0.05} = (2.042)\,(4.56) = 9.31$

$LSD_{0.01} = (2.750)\,(4.56) = 12.54$

(3) เปรียบเทียบ C กับ D

$$s_{\bar{d}} = \sqrt{85.80\left[\frac{1}{11}+\frac{1}{7}\right]} = 4.48$$

$LSD_{0.05} = (2.042)\,(4.48) = 9.15$ กรัม

$LSD_{0.01} = (2.750)\,(4.48) = 12.32$ กรัม

นำ LSD ที่คำนวณได้เหล่านี้ไปเปรียบเทียบกับความแตกต่างระหว่างค่าเฉลี่ยของแต่ละทรีตเมนต์กับทรีตเมนต์มาตรฐาน ซึ่งได้ผลดังแสดงในตาราง 12.2.2

ตาราง 12.2.2 การเปรียบเทียบระหว่างทรีตเมนต์ต่าง ๆ กับทรีตเมนต์มาตรฐานจากการทดลองเกี่ยวกับอาหารไก่ผสมโปรตีน 4 ชนิด

ชนิดของโปรตีน	จำนวนค่าสังเกต	น.น. เพิ่มเฉลี่ย (กรัม)	ค่าแตกต่างจาก check
A	9	68.33	10.76 *
B	10	52.20	5.37 ns
C	11	54.18	3.39 ns
D (check)	7	57.57	-

ns = not significant, * = แตกต่างทางสถิติในระดับ 5 เปอร์เซ็นต์

โดยสรุปแล้ว LSD มีการใช้กันมากที่สุด วิธีนี้แนะนำสำหรับการเปรียบเทียบแบบจับคู่ และควรเป็นการเปรียบเทียบในคู่ที่เราตั้งใจเอาไว้ เช่นเปรียบเทียบระหว่างทรีตเมนต์ต่าง ๆ ที่ทดลองกับทรีตเมนต์มาตรฐาน วิธี LSD ไม่เหมาะสำหรับการเปรียบเทียบแบบทุกค่าหรือแบบกลุ่ม โดยเฉพาะอย่างยิ่ง เมื่อมีทรีตเมนต์มาก ๆ ทั้งนี้เพราะวิธีนี้อาจพบว่าความแตกต่างระหว่างทรีตเมนต์บางคู่สูงกว่า LSD ทั้ง ๆ ที่ทรีตเมนต์ไม่แตกต่างกันจริง โดยสรุปแล้วเราใช้ LSD เมื่อการทดสอบโดยใช้ F - test พบว่าทรีตเมนต์มีความแตกต่างกันในทางสถิติ และเมื่อจะใช้วิธีแบบเปรียบเทียบเป็นกลุ่ม ก็อาจใช้เมื่อมีทรีตเมนต์จำนวนน้อย ๆ เช่น น้อยกว่า 6

12.3 วิธีดีเอ็มอาร์ที

วิธีดีเอมอาร์ที (DMRT) มาจากคำว่าวิธี Duncan' s New Multiple Range Test เป็นวิธีการเปรียบเทียบค่าเฉลี่ยแบบกลุ่ม [3] ใช้เปรียบเทียบได้ทุกคู่ที่ต้องการ เช่นมีค่าเฉลี่ย 10 ค่า ก็จัดได้ $(10 \times 9)/2 = 45$ คู่เปรียบเทียบ การใช้วิธีนี้ไม่จำเป็นต้องพบความแตกต่างโดยวิธี F - test ก็ได้ สมมุติว่าใช้ข้อมูลจากการทดลองในตาราง 9.3.1 และผลการวิเคราะห์วาเรียนซ์ในตาราง 9.3.2 ก็อาจแสดงวิธีการเปรียบเทียบค่าเฉลี่ยโดยวิธี DMRT ดังนี้

(1) คำนวณหา

$$s_{\bar{x}} = \sqrt{\frac{MSE}{n}}$$

$$= \sqrt{\frac{11.88}{4}} \quad \text{ก.ก./ไร่}$$

(2) เปิดตาราง SSR [4] ค่านี้เปิดจากตาราง ผ.10 โดยเปิดที่ระดับ 0.05 และ 0.01 วิธีเปิดตารางให้ดูค่า df ของ error ซึ่งเท่ากับ 15 ค่านี้อยู่ในสดมภ์แรก (ซ้ายมือ) และ ให้ดูค่า p ด้านบนของตาราง ในกรณีนี้เปิดค่า p จาก 2 ถึง 5 (เริ่มจาก 2 เสมอ แล้วนับไปเท่ากับจำนวนทรีตเมนต์) เมื่อได้ค่า SSR แล้วก็หาค่าเปรียบเทียบ (LSR)[5] โดยใช้สมการ

$$LSR = (SSR)(s_{\bar{x}}) \qquad \ldots(12\text{-}3)$$

เช่นที่ระดับความแตกต่าง 0.05 , p = 2

$$LSR = (3.01)(1.72) = 5.18$$

เมื่อคำนวณทุกค่าแล้วได้ผลดังนี้

	ค่า p			
	2	3	4	5
SSR (0.05)	3.01	3.16	3.25	3.31
SSR (0.01)	4.17	4.37	4.50	5.58
LSR (0.05)	5.18	5.43	5.59	5.69
LSR (0.01)	7.17	7.52	7.74	9.59

(3) จัดลำดับค่าเฉลี่ย จากตาราง 9.3.1 อาจเรียงค่าเฉลี่ยจากมากไปหาน้อยหรือจากน้อยไปหามากก็ได้ ทั้งนี้แล้วแต่แนวโน้มของชุดข้อมูลที่ใช้

ลำดับค่าเฉลี่ย	1	2	3	4	5
ทรีตเมนต์	A	B	C	E	D
ผลผลิต (กก.)	14.25	23.00	28.75	30.25	32.00

(4) ทดสอบความแตกต่าง ทดสอบเป็นขั้น ๆ โดยเริ่มจากค่าสูงสุดมาตั้ง และลบด้วยค่าต่ำสุด (D – A) ลบด้วยค่าต่ำรอง (D – B) และขยับไปเรื่อย ๆ เป็นลำดับจนถึงค่าชิดสูงสุด (D – E) ต่อจากนั้นก็กระทำเช่นเดียวกับค่าสูงรอง และรอง ๆ ลงไปจนเหลือคู่สุดท้าย (B – A) แล้วนำความแตกต่างนั้นไปเทียบกับค่า LSR ที่เหมาะสม ถ้าคู่ชิดใช้ p = 2 แล้วห่างออกไปเป็นลำดับใช้ p = 3, 4 และ 5 เช่น ใช้ระดับความแตกต่าง 0.01 ได้ผลดังนี้

$$D - A = 32.00 - 14.25 = 17.75 > 7.87^{**} \quad (p = 5)$$
$$D - B = 32.00 - 23.00 = 9.00 > 7.74^{**} \quad (p = 4)$$
$$D - C = 32.00 - 28.75 = 3.25 < 7.52^{ns} \quad (p = 3)$$
$$E - A = 30.25 - 14.25 = 16.00 > 7.74^{**} \quad (p = 4)$$
$$E - B = 30.25 - 23.00 = 7.25 < 7.52^{ns} \quad (p = 3)$$
$$C - A = 28.75 - 14.25 = 14.50 > 7.52^{**} \quad (p = 3)$$
$$C - B = 28.75 - 23.00 = 5.75 < 7.17^{ns} \quad (p = 2)$$
$$B - A = 23.00 - 14.25 = 8.75 > 7.17^{**} \quad (p = 2)$$

เสนอผลของการเปรียบเทียบให้ดูง่าย ๆ โดยใช้เส้นร่วมดังนี้

A	B	C	E	D
14.25	23.00	28.75	30.25	32.00

ค่าเฉลี่ยที่ขีดเส้นใต้ติดต่อกันไม่แตกต่างกันในทางสถิติ ในระดับความแตกต่าง 0.01 หรืออาจใช้ตัวอักษรดังนี้

A	B	C	E	D
14.25c	23.00b	28.75ab	30.25ab	32.00a

ค่าเฉลี่ยที่ตามด้วยอักษรคนละชนิด แตกต่างกันในทางสถิติในระดับ 0.01

การใช้ DMRT เมื่อทรีตเมนต์มีค่าสังเกตไม่เท่ากัน

ในกรณีของการทดลองแบบสุ่มตลอด (CRD) แต่ละทรีตเมนต์อาจมีค่าสังเกตไม่เท่ากัน เช่น จากการทดลองเรื่องอาหารไก่ในตอน 9.4 ซึ่ง A, B, C และ D มีค่าสังเกตเท่ากับ 9, 10, 11 และ 7 ตามลำดับ ดังแสดงในตาราง 12.2.2 การทดลองนี้มี df เท่ากับ 33 และ MSE เท่ากับ 85.80 ในการเปรียบเทียบระหว่างค่าเฉลี่ยของการทดลองนี้ $s_{\bar{x}}$ ของการเปรียบเทียบแต่ละคู่ไม่เท่ากัน ดังนั้นต้องมีการดัดแปลงสมการ ในการคำนวณหา LSR ดังนี้ คือ

1. นำค่า SSR ของแต่ละชุดเปรียบเทียบ (ที่หาได้จากตาราง SSR ที่ df เท่ากับ 33) ไปคูณกับ $\sqrt{\text{MSE}}$
2. นำค่าจากข้อ 1 ไปคูณกับ $\sqrt{\frac{1}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$ ก็จะได้ LSR สำหรับการเปรียบเทียบระหว่างค่าเฉลี่ยของทรีตเมนต์ i และ j

ตาราง 12.2.3 แสดงค่า SSR (significant studentized range) ที่ df = 33, ที่ระดับ 0.05

df	α	p		
		2	3	4
33	0.05	2.89	3.04	3.12

นำค่าเฉลี่ยในตาราง 12.2.2 มาเรียงลำดับ

ลำดับค่าเฉลี่ย	1	2	3	4
ทรีตเมนต์	B	C	D	A
น.น. เพิ่ม (กรัม)	52.20	54.18	57.57	68.33

ถ้าเปรียบเทียบระหว่าง A และ B ก็คำนวณได้ว่า

$$\text{LSR} = (\text{SSR})(\sqrt{\text{MSE}})\sqrt{\frac{1}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

SSR (0.05, df 33) = 3.12; MSE = $\sqrt{85.80}$; $n_i = 9$, $n_j = 10$ ดังนั้น

$$\text{LSR} = (3.12)(\sqrt{85.80})\sqrt{\frac{1}{2}\left(\frac{1}{9}+\frac{1}{10}\right)}$$

$$= 9.38$$

เมื่อเปรียบเทียบกับความแตกต่าง

$$A - B = 68.33 - 52.20 = 16.13 > 9.38$$

ซึ่งถือว่ามีความแตกต่างกันในระดับ 0.05

12.4 วิธี Tukey'w หรือ HSD (Honesty Significant Difference)

วิธีนี้ไม่ค่อยเห็นใช้กันบ่อยนัก ใช้ค่าเดียวเปรียบเทียบกับความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่จนครบทุกคู่ เหมือนค่า LSD คำนวณค่าเปรียบเทียบโดยใช้สมการ

$$w = \left[q(\alpha, p)\right]\left[s_{\bar{x}}\right] \qquad \ldots(12\text{-}4)$$

ทั้งนี้

q คือค่าจากตาราง percentage points of the studentize range (ตาราง ผ.12 ในภาคผนวก) เปิดโดยใช้ df ของ error จากตาราง วิเคราะห์วาเรียนซ์

α คือระดับความแตกต่างที่ต้องการใช้

p คือจำนวนค่าเฉลี่ย หรือจำนวนทรีตเมนต์

เช่นจากข้อมูลในตาราง 9.3.1 และ 12.2.1 ซึ่งมีค่า p = 5, df = 15, MSE = 11.88, q = 4.367 ดังนั้น ถ้าคำนวณหา w ที่ระดับความแตกต่าง 0.05 ก็ได้

$$w = 4.367\sqrt{\frac{11.88}{4}}$$

$$= 7.53$$

เมื่อนำค่าที่ได้ไปเปรียบเทียบกับความแตกต่างระหว่างค่าเฉลี่ยก็จะได้ผลดังนี้

A	B	C	E	D
14.25	23.00	28.75	30.25	32.00

ซึ่งให้ความแตกต่างเหมือนกับวิธี DMRT

12.5 วิธี Student - Newman Keul's Test

วิธีนี้คล้ายกับวิธี DMRT โดยมีค่าเปรียบเทียบเป็นคู่ ๆ โดยเฉพาะ แล้วแต่ลำดับคู่ คำนวณค่าเปรียบเทียบโดยใช้สมการ

$$w_p = \left[q(\alpha, p)\right]\left[s_{\bar{x}}\right] \qquad \ldots(12\text{-}5)$$

ทั้งนี้ q คือ ค่าจากตาราง percentage points of the studentized range เปิดโดยใช้ df ของ error จากตาราง วิเคราะห์วาเรียนซ์

α คือระดับความแตกต่างที่ต้องการใช้

p คือ2, 3, 4, ..., ระยะระหว่างค่าเฉลี่ยที่เปรียบเทียบกัน

เมื่อ $\alpha = 0.05$; df = 15, p = 2, 3, 4, 5 ค่า q จากตารางก็เท่ากับ 3.014, 3.674, 4.076 และ 4.367 ตามลำดับและ $s_{\bar{x}} = 1.72$ ก็อาจหา w_p ได้ดังนี้

p	2	3	4	5
$q(\alpha, p)$	3.014	3.674	4.076	4.367
W_p	5.18	6.32	7.01	7.51

แล้วนำค่าเหล่านี้ ไปเปรียบเทียบกับความแตกต่างแต่ละคู่ทรีตเมนต์ อาจเป็นลำดับดังเช่นวิธี DMRT ก็จะได้ผลดังนี้

A	B	C	E	D
14.25	23.00	28.75	30.25	32.00

ซึ่งคล้ายกับวิธี DMRT เช่นกัน

การเสนอผลการทดลองเพื่อแสดงความแตกต่างโดยวิธี DMRT, HSD หรือ Student Newman Keul's Test

ในการหาความแตกต่างโดยใช้วิธี DMRT หรือวิธีอื่น ๆ ในการเปรียบเทียบแบบกลุ่มนั้น เรามีการเรียงลำดับค่าเฉลี่ยจากน้อยไปหามาก หรือมากไปหาน้อยแล้วแต่สะดวก เมื่อมีความแตกต่าง เราก็จัดเป็นกลุ่ม ๆ โดยขีดเส้นใต้ แต่ในการเสนอผลการทดลองในงานวิจัยต่าง ๆ มักใช้เป็นตัวอักษรตามหลังค่าเฉลี่ย และการเสนอด้วยตัวอักษรไม่จำเป็นต้องเรียงลำดับค่าเฉลี่ย แต่ควรเรียงตามลำดับความสำคัญความสัมพันธ์หรือกลุ่มของทรีตเมนต์ เช่น

ทรีตเมนต์ (ถั่วเขียว)	จำนวนฝัก/ต้น (DMRT , 0.05)
1. สายพันธุ์ 102-21	18 ab
2. สายพันธุ์ 103-10	19 a
3. สายพันธุ์ 102.20	17 b
4. มอ. 1	15 c
5. อู่ทอง 1	15 c

ซึ่งค่าเฉลี่ยอยู่ในตำแหน่งของลำดับทรีตเมนต์ แต่ยังชี้ลำดับค่าเฉลี่ยได้โดยดูลำดับตัวอักษรตามหลังค่าเฉลี่ยนั้น ๆ

12.6 การเปรียบเทียบระหว่างกลุ่ม

ในการเปรียบเทียบระหว่างทรีตเมนต์ หรือระหว่างชุดของทรีตเมนต์ อาจกระทำโดยการแยก sum of squares ของทรีตเมนต์ออกเป็นส่วน ๆ ตามชุดเปรียบเทียบ แล้วศึกษาว่ามีความแตกต่างระหว่างชุดหรือไม่

ในการทดลองที่ใช้ทรีตเมต์ต่าง ๆ ชุดหนึ่ง เช่น มี 5 ทรีตเมนต์ ดังตาราง 9.3.1 อาจแยกเปรียบเทียบระหว่างทรีตเมนต์ต่าง ๆ ที่เราสนใจ เช่น A เป็นทรีตเมนต์มาตรฐาน คือไม่มีการปราบวัชพืชเลย อาจเปรียบเทียบกับทรีตเมนต์อื่น ๆ (คือ A vs BCDE) หรือ C กับ E เป็นเทคนิคการปราบวัชพืชคนละวิธีก็อาจเปรียบเทียบกันได้ (คือ C vs E) และเราอาจเปรียบเทียบ ระหว่างทรีตเมนต์หรือชุดของทรีตเมนต์อื่น ๆ อีก แต่ละชุดเปรียบเทียบมี df = 1 ดังนั้นจำนวนชุดเปรียบเทียบมีสูงสุดไม่เกินจำนวน df ของทรีตเมนต์ (คือไม่เกิน t - 1 ชุด) การใช้วิธีนี้ดำเนินดังนี้ :

(1) แยกกลุ่มเปรียบเทียบที่เราสนใจ กลุ่มนี้เรียกว่าชุดเปรียบเทียบ (L_j) เช่น ในกรณีของข้อมูลในตาราง 9.3.1 ก็ได้ L_1, L_2 , L_3 และ L_4

(2) กำหนดสัมประสิทธิ์ของการเปรียบเทียบ (c_i) สัมประสิทธิ์ของแต่ละ L_j รวมกันได้ศูนย์ (คือ $\sum c_i = 0$) และผลบวกของผลคูณของสัมประสิทธิ์ของแต่ละชุดในสดมภ์เดียวกันก็มีค่าเท่ากับศูนย์ ดังแสดงในตาราง 12.6.1

ตาราง 12.6.1 สัมประสิทธิ์ของชุดเปรียบเทียบ

ชุดเปรียบเทียบ	ทรีตเมนต์				
	A	B	C	D	i
ชุดเปรียบเทียบที่ 1 (L_1)	c_{11}	c_{21}	c_{31}	c_{41}	......c_{i1}
ชุดเปรียบเทียบที่ 2 (L_2)	c_{12}	c_{22}	c_{32}	c_{42}	......c_{i2}
⋮	⋮	⋮	⋮	⋮	⋮
ชุดเปรียบเทียบที่ j (L_j)	c_{1j}	c_{2j}	c_{3j}	c_{4j}	......c_{ij}

จากสัมประสิทธิ์ในตาราง 12.6.1 อาจเขียนได้

$$\sum c_i = 0$$

คือผลรวมของสัมประสิทธิ์แต่ละชุดเปรียบเทียบ (หรือแต่ละแถว) เท่ากับศูนย์ นอกจากนั้น การกำหนดสัมประสิทธิ์ ต้องอำนวยให้ผลบวกของผลคูณระหว่างสัมประสิทธิ์ของชุดเปรียบเทียบในสดมภ์เดียวกันเท่ากับศูนย์ หรืออาจเขียนได้ว่า

$$(c_{11})(c_{12}) + (c_{21})(c_{22}) + ... + (c_{i1})(c_{i2}) = 0$$

หรือเขียนว่า

$$\sum L_1 L_2 = 0$$

และเมื่อคูณระหว่างทุกชุดก็เขียนได้ว่า

$$\sum L_1 L_2 = \sum L_1 L_3 = \sum L_1 L_j = ... = \sum L_{j-1} L_j = 0$$

(3) วิธีการกำหนดสัมประสิทธิ์ของชุดเปรียบเทียบ จากข้อ 2 กำหนดว่าผลบวกของสัมประสิทธิ์ในแต่ละชุดเปรียบเทียบเท่ากับศูนย์ ($\sum c_i = 0$) เช่นนั้นชุดเปรียบเทียบต้องมีทั้งฝ่ายบวกและฝ่ายลบ ผลบวกของแต่ละฝ่ายต้องเท่ากัน ตัวอย่างเช่นมีทรีตเมนต์ต่าง ๆ อยู่ 5 ทรีตเมนต์ คือ A, B, C, D และ E ต้องการเปรียบเทียบระหว่าง 2 กลุ่ม คือ ABC vs DE คือกลุ่มแรกมี 3 ทรีตเมนต์ กลุ่มหลังมี 2 ทรีตเมนต์ ก็อาจหาสัมประสิทธิ์การเปรียบเทียบ โดยใช้ ค.ร.น. ดังนี้

จำนวนสมาชิกกลุ่มแรก	3
จำนวนสมาชิกกลุ่มหลัง	2
ดังนั้น ค.ร.น. เท่ากับ	6

สัมประสิทธิ์ของกลุ่มใด เท่ากับ ค.ร.น. หารด้วยจำนวนสมาชิกของกลุ่มนั้น เมื่อหารแล้วใส่เครื่องหมายบวก - ลบ เพื่อให้ได้ผลรวมเท่ากับศูนย์

สัมประสิทธิ์ของกลุ่มแรก $\frac{6}{3} = 2$

สัมประสิทธิ์ของกลุ่มหลัง $\frac{6}{2} = 3$

ดังนั้นอาจแสดงชุดเปรียบเทียบและสัมประสิทธิ์ได้ดังนี้

ทรีตเมนต์	A	B	C	D	E
สัมประสิทธิ์ (ABC vs DE)	2	2	2	−3	−3

ในกรณีที่ชุดเปรียบเทียบมีไม่ครบทุกทรีตเมนต์ ทรีตเมนต์ใดที่ไม่เกี่ยวข้อง ก็ให้สัมประสิทธิ์เท่ากับศูนย์ เช่น A vs BC

ทรีตเมนต์	A	B	C	D	E
ABC vs DE (L_1)	2	2	2	−3	−3
A vs BC (L_2)	2	−1	−1	0	0

สังเกตุว่านอกจาก $\sum c_i$ ของแต่ละบรรทัดเท่ากับศูนย์แล้วผลบวกของผลคูณของสัมประสิทธิ์ในบรรทัดที่หนึ่งและสองต้องเท่ากับศูนย์ด้วย คือ

$$\sum L_1 L_2 = (2)(2) + (2)(-1) + (2)(-1) + (-3)(0) + (-3)(0) = 0$$

หากผลบวกของผลคูณมีผลเช่นนี้ เรียกว่าการเปรียบเทียบนั้น มีความสมดุลในสองทิศทางหรือเรียกว่า มีออโธโกนอล [8] ค่าต่าง ๆ ดังกล่าวนี้แสดงเป็นตัวอย่างได้ดังตาราง 12.6.2

ตาราง 12.6.2 ชุดเปรียบเทียบระหว่างทรีตเมนต์ จากการทดลองปราบวัชพืชในถั่วเหลือง (ตาราง 9.3.1)

ชุดเปรียบเทียบ [7]		**ผลรวมของทรีตเมนต์**			
	A	**B**	**C**	**D**	**E**
	57	**92**	**115**	**128**	**121**
A vs BCDE (L_1)	+4	−1	−1	−1	−1
D vs CE (L_2)	0	0	+1	−2	+1
C vs E (L_3)	0	0	+1	0	−1
B vs CDE (L_4)	0	+3	−1	−1	−1

จากตาราง 12.6.2 เห็นได้ว่า $\sum c_i$ ของแต่ละ L_j มีค่าเท่ากับ 0 เช่น ใน L_1 หาได้ว่า $\sum c_i = 4 + (-1) + (-1) + (-1) + (-1) = 0$ และผลบวกของผลคูณระหว่างสัมประสิทธิ์ของทรีตเมนต์เดียวกันในชุดต่าง ๆ เป็น คู่ ๆ ไม่ว่าคู่ใดจะมีค่าเป็นศูนย์ คือ $\sum L_1L_2 = \sum L_1L_3 = \sum L_1L_4 = ... = \sum L_3L_4 = 0$ ถ้าชุดเปรียบเทียบมีคุณสมบัติเช่นนี้เรียกว่า มีออโทโกนอลกัน[8] ทุกชุด ซึ่งแสดงตัวอย่างได้ดังนี้

$$\begin{aligned}\sum L_1L_2 &= \sum c_{i1}c_{i2} \\ &= (+4)(0)+(-1)(0)+(-1)(+1)+(-1)(-2)+(-1)(+1) \\ &= 0\end{aligned}$$

$$\begin{aligned}\sum L_1L_3 &= \sum c_{i1}c_{i3} \\ &= (+4)(0)+(-1)(0)+(-1)(+1)+(-1)(0)+(-1)(-1) \\ &= 0\end{aligned}$$

คู่อื่น ๆ ก็หาได้เช่นเดียวกัน

ถ้าคุณสมบัติที่กล่าวแล้วนี้เป็นความจริง เราก็สามารถแยก treatment SS ออกเป็นชุด ๆ แต่ละชุดมี 1 df และแสดงได้ว่า

$$SS(L_1) + SS(L_2) + SS(L_3) + SS(L_4) = \text{treatment SS}$$

สมการทั่วไปสำหรับการคำนวณ $SS(L_j)$ คือ

$$SS(L_j) = (\sum c_iT_i)^2 / n\sum c_i^2 \qquad ...(12\text{-}6)$$

ทั้งนี้ให้

c_i = สัมประสิทธิ์ของชุดเปรียบเทียบ

T_i = ผลบวกของแต่ละทรีตเมนต์

n = จำนวนค่าสังเกตของแต่ละทรีตเมนต์

$\sum c_i^2$ = ผลรวมของสัมประสิทธิ์ยกกำลังสอง เช่นใน $L_1, \sum c_i^2 = 4^2 + (-1)^2 + \ldots + (-1)^2 = 20$

ตัวอย่าง จากข้อมูลในตาราง 12.6.2 เมื่อคำนวณหา sum of squares ของชุดเปรียบเทียบต่าง ๆ ได้ผลดังนี้

$$\begin{aligned}
SS(L_1) &= \frac{1}{(4)(20)}\left[(4)(57)+(-1)(92)+(-1)(115)+(-1)(128)+(-1)(121)\right]^2 \\
&= 649.80 \\
SS(L_2) &= \frac{1}{(4)(6)}\left[(1)(115)+(-2)(128)+(1)(121)\right]^2 \\
&= 16.67 \\
SS(L_3) &= \frac{1}{(4)(2)}\left[(1)(115)+(-1)(121)\right]^2 \\
&= 4.50 \\
SS(L_4) &= \frac{1}{(4)(12)}\left[(3)(92)+(-1)(115)+(-1)(128)+(-1)(121)\right]^2 \\
&= 161.33
\end{aligned}$$

จะเห็นได้ว่า ผลบวกของส่วนย่อย ๆ เหล่านี้ เท่ากับ SSTr ในตาราง 9.3.2 คือ

$$SS(L_1) + SS(L_2) + SS(L_3) + SS(L_4) = 832.30$$

ซึ่งจะเป็นเช่นนี้เมื่อชุดเปรียบเทียบมีออโธโกนอลเท่านั้น นำค่าดังกล่าวลงตารางวิเคราะห์วาเรียนซ์ ดังตาราง 12.6.3 แล้วคำนวณหา MS, F แล้วทดสอบโดยเปรียบเทียบกับค่า F ในตาราง F ต่อไป ในการเปิดตาราง F นั้น df ของชุดเปรียบเทียบเท่ากับ 1 เสมอ

ตาราง 12.6.3 ผลการวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 9.3.1

Sources	df	SS	MS	F	F (table) 5%	F (table) 1%
Treatments	4	832.30				
1. A vs BCDE	1	649.80	649.80	54.69**	4.54	8.68
2. CE vs D	1	16.67	16.67	1.40^{ns}		
3. C vs E	1	4.50	4.50	0.38^{ns}		
4. B vs CDE	1	161.33	161.33	13.58**		
Error	15	178.25	11.88			
Total	19	1,010.55				

ในกรณีที่เราต้องการแสดงชุดเปรียบเทียบที่เราสนใจเพียงบางชุด เช่น ต้องการทราบเฉพาะ A vs BCDE ว่ามีความสำคัญหรือไม่ โดยไม่สนใจชุดอื่น ๆ เมื่อเรามีค่า SSTr อยู่แล้ว ก็คำนวณเฉพาะ SS (A vs BCDE) เท่านั้น SS ที่เหลือได้จากการหักลบดังนี้

$$\text{Remainder SS} = \text{SSTr} - \text{SS(AvsBCDE)}$$

$$= 832.30 - 649.80 = 182.50$$

แล้วนำลงตารางดังแสดงในตาราง 12.6.2 ต่อไป

ตาราง 12.6.4 ผลการวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 9.3.1

Sources	df	SS	MS	F	F (table) 5%	F (table) 1%
Treatments	4	832.30				
A vs BCDE	1	649.80	649.80	54.69**	4.54	8.64
Remainder	3	182.50	60.83	5.12*	2.39	5.42
Error	15	178.25	11.88			
Total	19	1,010.55				

12.5 แบบฝึกหัด

1. ในการศึกษาถึงการใช้ปุ๋ย 3 ชนิด คือปุ๋ยชีวภาพ (เช่น ปุ๋ยหมัก) ปุ๋ยเคมี และส่วนผสมระหว่างปุ๋ยชีวภาพและปุ๋ยเคมี ปุ๋ยแต่ละชนิดมี การทดลอง 4 อัตราหรือทรีตเมนต์ และทรีตเมนต์ที่ไม่ใช้ปุ๋ยชนิดใด ๆ เลย (อาจเรียกว่าทรีตเมนต์มาตรฐาน) ร่วมอยู่ในการทดลองด้วย เราจะใช้วิธีใดในการเปรียบเทียบค่าเฉลี่ย

ก. เมื่อเปรียบเทียบระหว่างทรีตเมนต์มาตรฐานกับทรีตเมนต์อื่น ๆ

ข. เมื่อเปรียบเทียบระหว่างทรีตเมนต์ต่างๆ ว่าทรีตเมนต์ใดมีความแตกต่างกันบ้าง

ค. เมื่อเปรียบเทียบระหว่างทรีตเมนต์ต่างๆ ภายในกลุ่มปุ๋ยชนิดเดียวกัน

ง. เมื่อเปรียบเทียบระหว่างปุ๋ยชีวภาพและปุ๋ยเคมี

2. จากผลการวิเคราะห์วาเรียนซ์ในข้อ 3 แบบฝึกหัดที่ 10.10 จงเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยโดยใช้วิธีการที่เหมาะสม ถ้าหากต้องการที่จะเปรียบเทียบระหว่างการใส่ปุ๋ยและไม่ใส่ และในขณะเดียวกันเปรียบเทียบระหว่างปุ๋ยเคมีคนละสูตร จงแสดงวิธีการเปรียบเทียบดังกล่าว

3. จากผลของการวิเคราะห์วาเรียนซ์ในข้อ 3 แบบฝึกหัดที่ 10.10 จงทำการเปรียบเทียบค่าเฉลี่ยโดยวิธี DMRT

4. จงแสดงสัมประสิทธิ์ของชุดเปรียบเทียบที่มีออโธโกนอลระหว่างชุดต่าง ๆ และแต่ละชุดมี 1 df สำหรับการทดลองที่มี 3, 4, 5 ทรีตเมนต์

5. จงอธิบายถึงข้อดีข้อเสียของวิธีการเปรียบเทียบค่าเฉลี่ยค่าต่าง ๆ คือ LSD, DMRT และการเปรียบเทียบแบบออโทโกนอล และคิดว่าเมื่อใดจึงควรจะใช้วิธีการแต่ละชนิด

คำในบท

(1) least significant difference, (2) control, check, (3) multiple comparision, (4) significant studentized range, (5) Tukey's w-procedure, least significance range (6) Student-Newman-Keuls' test, (7) contrast, (8) orthogonal comparision.

บทที่ 13
การทดลองแบบแฟกตอเรียล

13.1 คำนำ

แผนการทดลองที่กล่าวถึงในบทที่ 10 และบทที่ 11 นั้น ใช้ในการค้นคว้าเพื่อเปรียบเทียบในเรื่องเดียว หรือปัญหาเดียวโดยเฉพาะ เช่น ในการทดลองเปรียบเทียบพันธุ์พืช เราก็มุ่งไปในเรื่องพันธุ์เพียงอย่างเดียว ส่วนปัญหาอื่น ๆ เช่น ระดับปุ๋ย ระยะปลูก หรือวันปลูก เป็นปัญหาคงที่ คือใช้เหมือนกันทุกพันธุ์ ถ้าเราต้องการทดสอบอาหารสัตว์สูตรต่าง ๆ ก็ต้องใช้พันธุ์เดียวกันและอายุเท่าๆ กัน ถ้าต้องการทดสอบการเก็บรักษาผลไม้ในอุณหภูมิต่าง ๆ กันก็ต้องใช้ผลไม้ชนิดเดียวกันที่มีอายุเก็บเกี่ยวเหมือนกันเหล่านี้เป็นต้น โดยวิธีนี้เห็นว่าเราต้องทำการทดลองหลาย ๆ ครั้งจึงจะได้ข้อมูลในปัญหาแต่ละอย่างที่ต้องการทราบได้ครบ เช่น ถ้าต้องการทราบว่าปุ๋ยสูตรใดเหมาะกับพืชพันธุ์ใด อาหารสัตว์สูตรใดเหมาะกับสัตว์พันธุ์ใด และอุณหภูมิการเก็บรักษาระดับไหนเหมาะกับผลไม้ชนิดใด ก็ต้องแยกไปทำการทดลองใหม่ ซึ่งนับว่าเป็นการเสียเวลาและทุนทรัพย์ การทดลองแบบแฟกตอเรียลที่จะกล่าวถึงในบทนี้ อำนวยให้สามารถศึกษาถึงปัญหาต่าง ๆ ได้พร้อมกันคราวละสองหรือสามอย่าง และจัดเป็นส่วนขยายเพิ่มเติมจากแผนการทดลองที่กล่าวไว้แล้วนั่นเอง

13.2 การทดลองแบบแฟกตอเรียล

ในการทดลองที่เราศึกษาปัญหามากกว่าหนึ่งชนิดพร้อม ๆ กันนั้น อาจจัดปัญหาเข้าชุดเพื่อศึกษาได้พร้อม ๆ กัน เช่น ถ้าการศึกษานั้นมี 2 ปัญหา เป็นต้นว่าต้องการที่จะศึกษาผลของการใส่ปุ๋ยและปูนขาวแก่พืช ถ้าหากว่ามีปุ๋ย 4 ระดับ คือ 0, 25, 50 และ 100 กก./ไร่ ส่วนปูนมี 3 ระดับคือ 0, 100 และ 200 กก./ไร่ ในกรณีนี้อาจจัดได้ทั้งสิ้น 4 x 3 = 12 ชุด ทั้งนี้ปุ๋ยแต่ละระดับได้รับการศึกษากับปูนทุกระดับ การจัดปัญหาเช่นนี้เรียกว่าเป็นการจัดแบบแฟกตอเรียล[1] ซึ่งมีความหมายว่าเป็นการจัดเข้าชุด ส่วนปัญหาที่จะจัดเข้าชุดเรียกว่า แฟกเตอร์ [2] เช่น ในกรณีดังกล่าวถ้าเราให้ระดับของปุ๋ยเป็นแฟกเตอร์ A ระดับย่อย ๆ มีสัญลักษณ์เป็น a_0 , a_1 , a_2 และ a_3 ส่วนปูนเป็นแฟกเตอร์ B ระดับย่อยมีสัญลักษณ์เป็น b_0 , b_1 และ b_2 และเมื่อจัดแฟกตอเรียลก็ได้ 12 ทรีตเมนต์ ดังนี้

ระดับปูนและสัญลักษณ์		ระดับปุ๋ยและสัญลักษณ์ 0 a_0	25 a_1	50 a_2	100 a_3
0	b_0	a_0b_0	a_1b_0	a_2b_0	a_3b_0
100	b_1	a_0b_1	a_1b_1	a_2b_1	a_3b_1
200	b_2	a_0b_2	a_1b_2	a_2b_2	a_3b_2

เมื่อได้จัดทรีตเมนต์เรียบร้อยแล้ว ก็นำไปทดลองโดยใช้แผนการทดลองในแบบต่าง ๆ ต่อไป เช่น อาจทดลองโดยใช้ CRD, RCB หรือลาตินสแควร์ จึงเห็นได้ว่าแฟกตอเรียลไม่ได้เป็นแผนการทดลอง แต่เป็นเพียงวิธีการจัดทรีตเมนต์เท่านั้น

ในการจัดทรีตเมนต์แบบแฟกตอเรียลนั้น เห็นได้ว่าจำนวนชุดของปัญหา หรือจำนวนทรีตเมนต์ขึ้นอยู่กับจำนวนแฟกเตอร์และระดับของแต่ละแฟกเตอร์ ในกรณีของตัวอย่างข้างบนจำนวนชุดของปัญหา คือ 4 x 3 ชุด ในกรณีที่แต่ละแฟกเตอร์มีระดับเท่ากัน เช่น มี 3 แฟกเตอร์ แฟกเตอร์ละ 2 ระดับ จำนวนปัญหาเท่ากับ 2 x 2 x 2 หรือ 2^3 ชุด ถ้ามี 3 แฟกเตอร์ สองแฟกเตอร์แรกมีอย่างละ 3 ระดับ แฟกเตอร์ที่สามมี 2 ระดับ จำนวนปัญหาก็เท่ากับ 3^2 x 2 ชุด ดังนี้เป็นต้น

เนื่องจากในการทดลองแบบแฟกตอเรียล มีการจัดชุดของแฟกเตอร์คละกัน ในการวิเคราะห์ผลการทดลองเราสามารถหาคำตอบได้ 2 ชนิด ชนิดแรกคือคำตอบเกี่ยวกับแต่ละแฟกเตอร์เดี่ยว ๆ เช่น คำตอบเกี่ยวกับผลของการใช้ปุ๋ยและการใช้ปูน คำตอบเช่นนี้เหมือนกับที่เราได้รับจากการนำปุ๋ยและปูนไปทดลองแยกกันใน CRD หรือ RCB นั่นเอง เราเรียกคำตอบดังกล่าวนี้ว่า **อิทธิพลหลัก**[3] คำตอบอีกชนิดหนึ่งที่ได้จากการวิเคราะห์คือ **ขนาดของปฏิกิริยาระหว่างแฟกเตอร์**[4] **ซึ่งอาจอธิบาย**ไว้เป็นเบื้องต้นว่า อิทธิพลหลักคือผลสนองตอบที่เพิ่มขึ้นหรือลดลงเนื่องจากการเปลี่ยนแปลงระดับของแฟกเตอร์หนึ่ง ๆ ส่วนปฏิกิริยาระหว่างแฟกเตอร์หมายถึงการที่ผลสนองตอบที่ไม่คงที่ของแฟกเตอร์หนึ่งเมื่อเปลี่ยนระดับของอีกแฟกเตอร์หนึ่ง

ในการทดลองที่มี 2 แฟกเตอร์ (A และ B) แต่ละแฟกเตอร์มี 2 ระดับ (0 และ 1) อาจจัดชุดปัญหาดังนี้

แฟกเตอร์ B	แฟกเตอร์ A	
	a_0	a_1
b_0	a_0b_0	a_1b_0
b_1	a_0b_1	a_1b_1

เมื่อนำชุดปัญหานี้ไปทดสอบ เราสามารถแสดงอิทธิพลหลักและปฏิกิริยาดังนี้

(1) อิทธิพลหลักของ A คือค่าที่เพิ่มขึ้นเนื่องจากการเปลี่ยนระดับของ A จาก a_0 เป็น a_1 ซึ่งเราใช้สัญลักษณ์ว่า A คือ

$$\begin{aligned} A &= \left[\frac{a_1b_1 + a_1b_0}{2}\right] - \left[\frac{a_0b_1 + a_0b_0}{2}\right] \\ &= \frac{1}{2}\left[(a_1b_1 + a_1b_0) - (a_0b_1 + a_0b_0)\right] \qquad \text{...(13-1)} \end{aligned}$$

(2) อิทธิพลหลักของ B คือค่าที่เพิ่มขึ้นเนื่องจากการเปลี่ยนระดับของ B จาก b_0 เป็น b_1 ซึ่งหาได้เช่นเดียวกับแฟกเตอร์ A ว่า

$$B = \left[\frac{a_1b_1 + a_0b_1}{2}\right] - \left[\frac{a_1b_0 + a_0b_0}{2}\right]$$

$$= \frac{1}{2}\left[(a_1b_1 + a_0b_1) - (a_1b_0 + a_0b_0)\right] \qquad \ldots(13\text{-}2)$$

(3) ปฏิกิริยาระหว่างแฟกเตอร์ AB คือการไม่คงที่ของอัตราสนองตอบของแฟกเตอร์หนึ่ง เมื่อเปลี่ยนแปลงระดับอีกแฟกเตอร์หนึ่ง ถ้ากำหนดให้

$(AB)_0 = a_1b_0 - a_0b_0$ เป็นการเพิ่มเมื่อใช้ A ที่ระดับ b_0

$(AB)_1 = a_1b_1 - a_0b_1$ เป็นการเพิ่มเมื่อใช้ A ที่ระดับ b_1

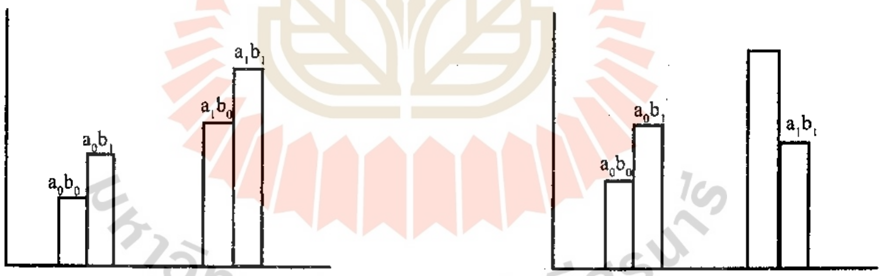
ดังนั้นปฏิกิริยา (AB) เมื่อวัดจากแฟกเตอร์ A คือค่าเฉลี่ยของ $(AB)_1$ - $(AB)_0$ คือ

$$AB = \left[\frac{a_1b_1 - a_0b_1}{2}\right] - \left[\frac{a_1b_0 - a_0b_0}{2}\right] \qquad \ldots(13\text{-}3)$$

$$= 1/2\ [(a_1b_1 - a_0b_1) - (a_1b_0 - a_0b_0)]$$

$$= 1/2\ [(a_1b_1 + a_0b_1 + a_1b_0)]$$

ถ้าไม่มีปฏิกิริยาระหว่างแฟกเตอร์ คือการเพิ่มเมื่อใช้ A ในระดับ b_0 และ b_1 เท่ากัน ดังนั้น $(AB)_0 = (AB)_1$ รูปข้างล่างแสดงผลของชุดปัญหาที่ไม่มีและมีปฏิกิริยาระหว่างแฟกเตอร์ตามลำดับ เป็นความแตกต่าง เมื่อเพิ่มแฟกเตอร์ B จาก b_0 เป็น b_1



ไม่มีปฏิกิริยา AB **มีปฏิกิริยา AB**

รูป 13.2.1 ผลจากปฏิกิริยาระหว่างแฟกเตอร์ A และ B

13.3 การวิเคราะห์วาเรียนซ์ในการทดลองที่ใช้สองแฟกเตอร์

ในการทดลองที่มีปัญหา 2 แฟกเตอร์ เช่น ทดลองเกี่ยวกับผลของการใส่ปุ๋ยและปูนขาวต่อพืชตระกูลถั่ว อาจให้ปุ๋ยเป็นแฟกเตอร์ A ซึ่งมี 4 ระดับ (a_0, a_1, a_2 และ a_3) และปูนขาวเป็นแฟกเตอร์ B ซึ่งมี 3 ระดับ (b_0, b_1 และ b_2) เมื่อจัดปุ๋ยและปูนทุกระดับเข้าชุดก็จะได้ 4 x 3 = 12 คือ a_0b_0, a_0b_1, a_0b_2,..., a_3b_2 ดังตาราง 13.3.1 แต่ละชุดจะกลายเป็นปัญหาแต่ละทรีตเมนต์ เราก็นำทรีตเมนต์นี้

ไปทดลองโดยใช้แผนการทดลองที่เหมาะสมต่อไป เช่น จากผลการทดลองแบบ CRD ดังแสดงไว้ในตาราง 13.3.2 จากตารางดังกล่าวนี้ ระดับปุ๋ย a_0, a_1, a_2 และ a_3 เท่ากับ 0, 25, 50 และ 100 กก./ไร่ ตามลำดับ สำหรับระดับปุ๋ย b_0,b_1 และ b_2 เท่ากับ 0, 100 และ 200 กก./ไร่ ตามลำดับ

ตาราง 13.3.1 การจัดชุดแฟกตอเรียลที่มี 2 แฟกเตอร์ คือ A และ B

		แฟกเตอร์ A			
		a_0	a_1	a_2	a_3
แฟกเตอร์ B	b_0	a_0b_0	a_1b_0	a_2b_0	a_3b_0
	b_1	a_0b_1	a_1b_1	a_2b_1	a_3b_1
	b_2	a_0b_2	a_1b_2	a_2b_2	a_3b_2

แบบจำลองทางคณิตศาสตร์ของการทดลองแบบแฟกตอเรียล

ก่อนที่ทำการวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 13.3.2 เราต้องศึกษาถึงรายละเอียดที่จำเป็นก่อนทำการวิเคราะห์เสียก่อน

ในกรณีที่การทดลองมี 2 แฟกเตอร์ คือ A มี a ระดับ และ B มี b ระดับ และใช้แผนการทดลองแบบ CRD ดังกล่าวนี้ ก็อาจแสดงแบบจำลองทางคณิตศาสตร์ได้ดังนี้

ในกรณีที่ทดลองแบบ CRD ได้สมการ

$$X_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad \ldots(13\text{-}4)$$

ในสมการนี้ τ เป็นผลของทรีตเมนต์ เมื่อขยายการทดลองเป็นแบบแฟกตอเรียล τ_i ประกอบด้วยแฟกเตอร์ α และ β ก็ได้สมการ

$$X_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \qquad \ldots(13\text{-}5)$$

เมื่อให้

μ เป็นค่าเฉลี่ยของประชากร

α, β เป็นอิทธิพลหลักของแฟกเตอร์ A และ B ตามลำดับ ซึ่งเป็นค่าเบี่ยงเบนจากค่าเฉลี่ยของประชากร

$\alpha\beta$ เป็นปฏิกิริยาระหว่างอิทธิพลหลัก

$i = 1, 2, \ldots, a$ เมื่อ a เป็นจำนวนระดับของเฟกเตอร์ A

$j = 1, 2, \ldots, b$ เมื่อ b เป็นจำนวนระดับของเฟกเตอร์ B

$k = 1, 2, \ldots, n$ เมื่อ n เป็นจำนวนค่าสังเกตแต่ละทรีตเมนต์

ตาราง 13.3.2 ผลผลิตฝักสดของถั่วลิสงเมื่อได้รับปุ๋ยสูตร 12-24-12 จำนวน 4 ระดับ และปูนขาว 3 ระดับ โดยใช้แผนการทดลองแบบ CRD จำนวน 4 ซ้ำ (ข้อมูลสมมติ)

แฟกเตอร์ A	แฟกเตอร์ B	ซ้ำ				รวม
ปุ๋ย (กก./ไร่)	ปูนขาว (กก./ไร่)	1	2	3	4	
0	0	4	3	4	6	17
	100	4	4	5	6	19
	250	8	7	6	10	31
15	0	3	6	5	5	19
	100	4	7	2	6	21
	200	6	4	8	9	27
30	0	5	7	5	6	23
	100	4	8	6	8	26
	200	9	10	12	10	41
45	0	7	10	8	9	34
	100	10	7	10	12	39
	200	9	9	5	8	31
รวม						328

ในการทดลองนี้ให้แฟกเตอร์ A และ B เป็นปัจจัยคงที่ ดังนั้น $\sum \alpha_i = 0, \sum \beta_j = 0$ และปฏิกิริยาระหว่างแฟกเตอร์ คือ $\sum \alpha\beta_{ij} = 0$ เช่นกัน ส่วนสมมุติฐานที่ทดสอบมีดังนี้

แฟกเตอร์ A $\quad H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_a = 0$

$\quad H_1 :$ อย่างน้อยมี 1 $\alpha \neq 0$

แฟกเตอร์ B $\quad H_0 : \beta_1 = \beta_2 = \ldots = \beta_b = 0$

$\quad H_1 :$ อย่างน้อยมี 1 $\beta \neq 0$

และสำหรับปฏิกิริยาระหว่างแฟกเตอร์

$\quad H_0 : (\alpha\beta)_{ij} = 0$ สำหรับ ij

$\quad H_1 :$ อย่างน้อยมี 1 $\alpha\beta \neq 0$

การแยก sum of square (SS)

การแยก sum of square เป็นขั้นตอนต่อเนื่องจากการวิเคราะห์วาเรียนซ์ ในสมการ (8-1) และ (8-2) อาจแยก SSTr ดังนี้

$$\text{SSTr} = \sum(\overline{X}_{ij}. - \overline{X}..)^2$$

$$= \sum_i\sum_j\sum_k[(\overline{X}_i.. - \overline{X}...) + (\overline{X}_{.j.} - \overline{X}...) + (\overline{X}_{ij}. - \overline{X}_i.. - \overline{X}_{.j.} + X...)]^2$$

$$= bn\sum_i(\overline{X}_i... - \overline{X}...)^2 + an\sum_j(\overline{X}_{.j.} - \overline{X}...)^2 + n\sum_i\sum_j(\overline{X}_{ij}. - \overline{X}_i.. - \overline{X}_{.j.} + \overline{X}...)^2$$

$$= \underbrace{\frac{(\sum X_i^2..)}{bn} - CF}_{SS(A)} + \underbrace{\frac{\sum X_{.j.}^2}{an} - CF}_{SS(B)} + \underbrace{SSTr - SS(A) - SS(B)}_{SS(AB)} \qquad ...(13\text{-}6)$$

ดังนั้นจากสมการ (8-1) เราขยายได้เป็นสมการพร้อม df ดังนี้

$$TSS = SS(A) + SS(B) + SS(AB) + SSE$$

ซึ่งมี df เท่ากับ $a-1, b-1, (a-1)(b-1)$ และ $ab(n-1)$ ซึ่งนำลงตารางวิเคราะห์วาเรียนซ์ได้ดังตาราง 13.3.3

ตาราง 13.3.3 ตารางวิเคราะห์วาเรียนซ์ที่มี 2 แฟกเตอร์ ในแผนการทดลอง CRD

Sources	df	SS	MS		F	EMS
A	$a-1$	SS(A)	MS(A)	M4	M4/M1	$\sigma^2 + bnK_a^2$
B	$b-1$	SS(B)	MS(B)	M3	M3/M1	$\sigma^2 + anK_b^2$
AB	$(a-1)(b-1)$	SS(AB)	MS(AB)	M2	M2/M1	$\sigma^2 + nK_{ab}^2$
Error	$ab(n-1)$	SSE	MSE	M1		σ^2
Total	$nab-1$	TSS				

ดังกล่าวไว้แล้วว่า สมมุติว่าการทดลองนี้ A, B และ AB เป็นปัจจัยคงที่ ดังนั้นจากตาราง 13.3.3 แสดงได้ว่า

$$MS(A) \text{ หรือ } M4 = \sigma^2 + \frac{bn\sum\alpha_i^2}{a-1} = \sigma^2 + bnK_a^2$$

$$MS(B) \text{ หรือ } M3 = \sigma^2 + \frac{an\sum\beta_j^2}{b-1} = \sigma^2 + anK_b^2$$

$$MS(AB) \text{ หรือ } M2 = \sigma^2 + \frac{n\sum(\alpha\beta)^2}{(a-1)(b-1)} = \sigma^2 + nK_{ab}^2$$

$$MSE \text{ หรือ } M1 = \frac{SSE}{ab(n-1)} = \sigma^2$$

การวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 13.3.3

ขั้นที่ 1 ทำการวิเคราะห์แบบ CRD ธรรมดา โดยไม่คำนึงว่าเป็นการจัดแฟกเตอร์แบบแฟกตอเรียล

$$\text{Correction factor (CF)} = \frac{(\sum X_{ijk})^2}{abn} = \frac{(328)^2}{48} = 2{,}241.33$$

$$\text{Total SS(TSS)} = 4^2 + 3^2 + \ldots + 8^2 - 2{,}241.33 = 264.63$$

$$\text{Treatment SS(SSTr)} = \frac{[17^2 + 19^2 + \ldots + 31^2]}{4} - 2{,}241.33 = 175.17$$

$$SSE = TSS - SSTr$$

$$= 264.63 - 175.17 = 89.46$$

โดยมี df ดังนี้

df ของ total = nab – 1 = (4)(4)(3) – 1 = 47

df ของ treatment = ab – 1 = (4)(3) – 1 = 11

df ของ error = ab(n – 1) = (4)(3)(3) = 36

ขั้นที่ 2 วิเคราะห์เพื่อแยก SSTr ออกมาเป็นอิทธิพลหลัก A, B และปฏิกิริยาระหว่างแฟกเตอร์ AB ก่อนการวิเคราะห์ต้องนำผลบวกของทรีตเมนต์ ซึ่งเกิดจากการจัดแฟกตอเรียล มาจัดระเบียบเป็นตารางแบบแถวและคอลัมน์ที่มีแฟกเตอร์ A และ B ตัดกัน ดังตาราง 13.3.4 เราเรียกตารางนี้ว่าตาราง AB

ตาราง 13.3.4 ตาราง AB จากข้อมูลในตาราง 13.3.2

B = ปูนขาว (กก./ไร่)	A = ปุ๋ย (กก./ไร่)				รวม
	0	15	30	45	
0	17	19	23	34	93
100	19	21	26	39	105
200	31	27	41	31	130
รวม	67	67	90	104	328

จากสมการ (13-6) เรากำหนดใหม่เพื่อความสะดวกในการคำนวณว่า

$$SS(A) = \frac{\sum X_{i\cdot}^2}{nb} - CF = \frac{\sum A^2}{nb} - CF$$

$$SS(B) = \frac{\sum X_{\cdot j}^2}{na} - CF = \frac{\sum B^2}{na} - CF$$

$$SS(AB) = \frac{\sum X_{ij}^2}{n} - CF - SS(A) - SS(B)$$

$$= \frac{\sum (AB)^2}{n} - CF - SS(A) - SS(B)$$

ดังนั้นจากตาราง 13.3.4

$$SS(A) = \frac{\sum A^2}{nb} - CF$$

$$= \frac{67^2 + 67^2 + \ldots + 104^2}{(4)(3)} - 2{,}241.33$$

$$= 83.17$$

$$SS(B) = \frac{\sum B^2}{na} - CF$$

$$= \frac{93^2 + 105^2 + 130^2}{(4)(4)} - 2{,}241.33$$

$$= 44.55$$

$$SS(AB) = \frac{\sum (AB)^2}{n} - CF - SS(A) - SS(B)$$

$$= \frac{(17^2 + 19^2 + \ldots + 31^2)}{4} - 2{,}241.33 - 83.17 - 44.55$$

$$= 47.45$$

ในการคำนวณ SS(AB) นี้ ถ้าหาค่า SSTr ไว้แล้ว ก็อาจหาได้โดยตรงโดยใช้ผลต่างดังนี้

$$SS(AB) = SSTr - SS(A) - SS(B)$$

ในขั้นสุดท้ายคือหา SSE ถ้าไม่ได้วิเคราะห์ไว้ในขั้นที่ 1 ก็สามารถหาโดยใช้สมการดังนี้

$$SSE = TSS - SS(A) - SS(B) - SS(AB)$$

$$= 264.63 - 83.17 - 44.55 - 47.45 = 89.46$$

ขั้นที่ 3 นำผลการวิเคราะห์ลงตารางวิเคราะห์วาเรียนซ์ดังตาราง 13.3.5

การคำนวณ MS และค่า F กระทำโดยวิธีการที่กล่าวมาแล้ว คือสมมุติว่าเป็นปัจจัยคงที่ ก็ใช้ MSE หารค่าบนทุกค่า เมื่อคำนวณค่า F แล้วก็เปิดตารางเพื่อเปรียบเทียบ ซึ่งปรากฏว่าอิทธิพลหลัก A มีความแตกต่างกันในทางสถิติอย่างมีนัยสำคัญยิ่ง และปฏิกิริยา AB มีความแตกต่างอย่างมีนัยสำคัญ การที่ปฏิกิริยามีความแตกต่างนี้แสดงว่า ผลของปุ๋ยแต่ละระดับให้ผลแตกต่างกันตามระดับของปูน

ตาราง 13.3.5 **ผลการวิเคราะห์ของข้อมูลในตาราง 13.3.1**

Sources	df	SS	MS	F	F (table)	
				(model 1)	5%	1%
Treatments	11	(175.17)				
A	3	83.17	27.22	11.13**	2.86	4.39
B	2	44.45	22.22	8.92**	3.26	5.25
AB	6	47.45	7.91	3.18*	2.36	3.35
Error	36	89.46	2.49			
Total	47	264.63				

** = แตกต่างทางสถิติในระดับ 1 เปอร์เซ็นต์

ต่อจากนั้นก็อาจเปรียบเทียบค่าเฉลี่ยของแฟกเตอร์ A และ B ซึ่งจะอธิบายในตัวอย่างจากการทดลองแบบ RCB ในการคำนวณส่วนเบี่ยงเบนมาตรฐานนั้น ใช้ตัวหารตามความเหมาะสม เช่น เปรียบเทียบระหว่างปุ๋ย

$$s_{\bar{x}} = \sqrt{\frac{MSE}{nb}} = \sqrt{\frac{2.49}{12}} = 0.46 \text{ กก.}$$

เปรียบเทียบระหว่างปุ๋ยเมื่อใส่ปูนขาว 200 กก.

$$s_{\bar{x}} = \sqrt{\frac{MSE}{n}} = \sqrt{\frac{2.49}{4}} = 0.79 \text{ กก.}$$

การทดลองที่ใช้สองแฟกเตอร์ใน RCB

ในการทดลองที่มี 2 แฟกเตอร์ หลังจากการจัดกลุ่มของแฟกเตอร์แล้วอาจนำไปทดสอบโดยใช้แผนการทดลองแบบ RCB ซึ่งแสดงความปรวนแปร และวิธีการวิเคราะห์วาเรียนซ์ดังตาราง 13.3.6 ตัวอย่างดังเช่นข้อมูลในตาราง 13.3.7 ซึ่งเป็นผลจากการทดลองอาหารสุกรที่ผสมไลซีน 3 ระดับ (คือ 0, 0.6 และ 0.9 เปอร์เซ็นต์) และโปรตีน 4 ระดับ คือ (10, 12, 14 และ 16 เปอร์เซ็นต์)

ตาราง 13.3.6 ตารางวิเคราะห์วาเรียนซ์ของแผนการทดลอง แบบแฟกตอเรียลใน RCB

Sources	df	SS	MS
Blocks	n – 1	SSB	SSB/n – 1
Treatments	ab – 1	SSTr	
A	a – 1	SS(A)	SS(A)/a – 1
B	b – 1	SS(B)	SS(B)/b – 1
AB	(a – 1)(b – 1)	SS(AB)	SS(AB)/(a – 1)(b – 1)
Error	(ab – 1)(n – 1)	SSE	SSE/(ab – 1)(n – 1)
Total	abn – 1		

ตาราง 13.3.7 หนักเพิ่มของสุกรที่เลี้ยงด้วยอาหารเสริมไลซีนและมีระดับโปรตีนต่าง ๆ กัน (กก./20 วัน)

ไลซีน(%) (Lysine)	โปรตีน (%)	บลอก I	II	III	IV	รวม
0	10	9	10	9	10	38
	12	12	9	10	12	43
	14	10	8	12	16	46
	16	14	12	8	11	45
0.6	10	9	15	20	18	62
	12	12	8	18	19	57
	14	16	16	22	18	72
	16	15	18	15	20	68
0.9	10	12	20	14	21	67
	12	14	18	17	18	67
	14	13	16	13	22	64
	16	15	18	17	15	65
รวม		151	168	175	200	694

แบบจำลองทางคณิตศาสตร์

ในการทดลองที่มี 2 แฟกเตอร์ และใช้แผนการทดลองแบบ RCB ก็อาจแสดงแบบจำลองทางคณิตศาสตร์ ได้ดังนี้

$$X_{ijk} = \mu + B_k + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \qquad \text{...(13-11)}$$

เมื่อให้

μ เป็นค่าเฉลี่ยของประชากร

B เป็นผลของการใช้บลอก

α, β เป็นอิทธิพลของแฟกเตอร์ A และ B ตามลำดับ

$\alpha\beta$ เป็นปฏิกิริยาระหว่างอิทธิพลหลัก

ε เป็นความคลาดเคลื่อนในการทดลอง

i = 1, 2, 3, ... a เมื่อ a เป็นจำนวนระดับของแฟกเตอร์ A

j = 1, 2, 3, ... b เมื่อ b เป็นจำนวนระดับของแฟกเตอร์ B

k = 1, 2, 3, ... n เมื่อ n เป็นจำนวนบลอก

ในกรณีที่เป็นแผนการทดลองแบบ RCB สามารถแยกแหล่งของความปรวนแปรดังตาราง 13.3.6

นำผลการทดลองดังแสดงในตาราง 13.3.7 มาวิเคราะห์วาเรียนซ์ตามขั้นตอนดังต่อไปนี้

ขั้นที่ 1 วิเคราะห์ แบบ RCB

$$CF = \frac{(\sum X_{ijk})^2}{nab} = \frac{(694)^2}{48} = 10{,}034.08$$

$$\begin{aligned}
\text{Total SS(TSS)} &= \sum X_{ijk}^2 - CF \\
&= 9^2 + 12^2 + 10^2 + \ldots + 15^2 - 10{,}034.08 \\
&= 10{,}792 - 10{,}034.08 \\
&= 757.92
\end{aligned}$$

$$\begin{aligned}
\text{Block SS(SSB)} &= (\sum B_k^2)/ab - CF \\
&= \frac{151^2 + 168^2 + 175^2 + 200^2}{(3)(4) = 12} - 10{,}034.08 \\
&= 10{,}137.50 - 10{,}034.08 \\
&= 103.42
\end{aligned}$$

$$\text{Treatment SS(SSTr)} = (\sum X_{ij})^2/n - CF$$
$$= \frac{38^2 + 43^2 + ... + 65^2}{4} - 10{,}034.08$$
$$= 10{,}408.50 - 10{,}034.08$$
$$= 374.42$$

$$\text{Error SS(SSE)} = TSS - SSB - SSTr$$
$$= 757.92 - 103.42 - 374.42$$
$$= 280.08$$

ขั้นที่ 2 วิเคราะห์เพื่อแยก SSTr ออกเป็นส่วนย่อย ๆ ดังแสดงไว้ในตาราง 13.3.6 ในการวิเคราะห์ขั้นนี้ ต้องจัดทำตาราง AB เสียก่อน ดังตาราง 13.3.8

ตาราง 13.3.8 ตาราง AB จากการใช้ไลซีนและโปรตีนในอาหารสุกร

โปรตีน	ไลซีน (%)			Total
(%)	0	0.6	0.9	
10	38	62	67	167
12	43	57	67	167
14	46	72	64	182
16	45	68	65	178
	172	259	263	694

แล้ววิเคราะห์ผลดังนี้

$$SS(A) = SS(\text{Lysine}) = \frac{\sum X_i^2}{nb} - CF = \frac{\sum A^2}{nb} - CF$$
$$= \frac{172^2 + 259^2 + 263^2}{16} - 10{,}034.08$$
$$= 330.54$$

$$SS(B) = SS(\text{Protein}) = \frac{\sum X_j^2}{an} - CF = \frac{\sum B^2}{an} - CF$$
$$= \frac{167^2 + 167^2 + ... + 178^2}{12} - 10{,}034.08$$
$$= 14.75$$

$$SS(AB) = SS(\text{Lysine} \times \text{Protein}) = SSTr - SS(A) - SS(B)$$

ถ้าหากว่าเราไม่หาค่า SSTr ไว้ก่อน ก็อาจคำนวณ SS (AB) ได้โดยตรงดังนี้

$$
\begin{aligned}
SS(AB) &= \frac{\sum X_{ij}^2}{n} - CF - SS(A) - SS(B) \\
&= \frac{\sum AB^2}{n} - CF - SS(A) - SS(B) \\
&= \frac{38^2 + 43^2 + ... + 65^2}{4} - 10{,}034.08 - 330.54 - 14.75 \\
&= 29.13
\end{aligned}
$$

ขั้นที่ 3 นำผลของการวิเคราะห์ลงตารางวิเคราะห์วาเรียนซ์ดังตาราง 13.3.9

ตาราง 13.3.9 ผลการวิเคราะห์วาเรียนซ์ของข้อมูลในตาราง 13.3.5

Sources	df	SS	MS	F	F(table) 5 or 1%
Blocks	3	103.42	34.47	4.06*	2.89 (.05)
Treatments	11	374.42	34.03	4.01**	2.84 (.01)
Lysine (L)	2	330.54	165.27	19.47**	5.32 (.01)
Protein (P)	3	14.75	4.91	0.58^{ns}	2.89 (.05)
L x P	6	29.13	4.85	0.57^{ns}	2.39 (.05)
Error	33	280.08	8.49		
Total	47	757.92			

***, **, ns แตกต่างที่ระดับ 5, 1 เปอร์เซ็นต์ และไม่แตกต่างตามลำดับ**

ขั้นที่ 4 เปรียบเทียบค่า F ที่คำนวณได้กับค่า F ในตาราง ในการเปิดหาค่า F ในตารางนั้น ให้ดูจาก df ของตัวตั้งและตัวหาร เช่น F(Lysine) ให้เปรียบเทียบกับค่าจากตารางซึ่งเปิดหาค่าที่ df ตัวตั้ง 2 และตัวหาร 33 ที่ระดับ 0.01 มีค่าเท่ากับ 5.32

ขั้นที่ 5 คำนวณหาสัมประสิทธิ์ของความปรวนแปร (CV)

$$
\begin{aligned}
CV(\%) &= \frac{\sqrt{MSE}}{\text{Grand mean}} \times 100 \\
&= \frac{\sqrt{8.48}}{694/48} \times 100 \\
&= 20.10\%
\end{aligned}
$$

ขั้นที่ 6 การเปรียบเทียบค่าเฉลี่ยของแฟกเตอร์ต่าง ๆ คือ ไลซีนและโปรตีน ในขั้นนี้ต้องจัดสร้างตารางค่าเฉลี่ยโดยใช้ข้อมูลในตาราง 13.3.8 ดังนี้

ตาราง 13.3.9 ค่าเฉลี่ยของโปรตีน ไลซีน และปฏิกิริยา

โปรตีน (%)	ไลซีน (%)			เฉลี่ย
	0	0.6	0.9	
10	9.50	15.50	16.67	13.92
12	10.75	14.25	16.75	13.92
14	11.50	18.00	16.00	15.17
16	11.25	17.00	16.25	14.83
เฉลี่ย	10.75	16.19	16.44	14.46

จากตารางวิเคราะห์วาเรียนซ์ ผลของโปรตีนไม่แตกต่างกันจึงไม่จำเป็นต้องเปรียบเทียบระหว่างค่าเฉลี่ย แต่ผลของไลซีนมีความแตกต่าง ดังนั้นอาจคำนวณค่า LSD ดังนี้

$$LSD_{(0.05)} = t_{(0.05)} s_{\bar{d}}$$

ค่า $t_{0.05}$ เปิดจากตาราง t ที่ df 33 มีค่าเท่ากับ 2.036 ส่วน $s_{\bar{d}}$ สำหรับการเปรียบเทียบระหว่างไลซีนหาได้จากสมการ

$$s_{\bar{d}} = \sqrt{\frac{2MSE}{nb}} = \sqrt{\frac{2(8.49)}{4x4}} = 1.03$$

ดังนั้น

$$LSD_{(0.05)} = (2.036)(1.03) = 2.097 \text{ กก./20 วัน}$$

ในทำนองเดียวกันหาได้ว่า

$$LSD_{(0.01)} = 2.815 \text{ กก./20 วัน}$$

เมื่อนำไปเปรียบเทียบกับไลซีนระดับต่าง ๆ พบว่าการไม่ใช้ไลซีนแตกต่างกับการใช้ทุกระดับ แต่ไลซีนระดับ 0.6 และ 0.9 ไม่แตกต่างกันไม่ว่าที่โปรตีนระดับใดก็ตาม

13.4 การวิเคราะห์วาเรียนซ์ใน 2^2 แฟกตอเรียล

แฟกตอเรียลที่มี 2 แฟกเตอร์ แต่ละแฟกเตอร์มี 2 ระดับ ซึ่งมีจำนวน 2 x 2 แฟกตอเรียล อาจทำการวิเคราะห์โดยใช้สัมประสิทธิ์ของแฟกตอเรียล[(5)] ในการทดลองแบบนี้ ให้แฟกเตอร์ A มีระดับเป็น a_0, a_1 ให้แฟกเตอร์ B มีระดับ b_0, b_1 ดังนั้น เมื่อจัดเข้าชุดกันได้แฟกตอเรียลในลำดับ

a_0b_0, a_1b_0, a_0b_1 และ a_1b_1 ซึ่งอาจเขียนให้กะทัดรัดเป็น (1), a, b, ab โดยที่ a_0b_0 = (1), a_1b_0 = a, a_0b_1 = b และ a_1b_1 = ab ในตอน 13.2 ได้อธิบายเกี่ยวกับอิทธิพลหลักและปฏิกิริยาไว้ดังนี้

$$A = \frac{1}{2}[(a_1b_1 + a_1b_0) - (a_0b_1 + a_0b_0)]$$

ซึ่งอาจใช้สัญลักษณ์ใหม่ดังกล่าวแล้วข้างบนดังนี้

$$A = \frac{1}{2}[ab + a - b - (1)] \qquad \ldots(13\text{-}7)$$

ในทำนองเดียวกันก็เขียนได้ว่า

$$B = \frac{1}{2}[ab + b - a - (1)] \qquad \ldots(13\text{-}8)$$

$$AB = \frac{1}{2}[ab - b - a + (1)] \qquad \ldots(13\text{-}9)$$

ซึ่งสอดคล้องกับสมการ (13-1), (13-2) และ (13-3) ทุกประการ

จากสมการข้างบนถ้าแยกออกมาเฉพาะสัมประสิทธิ์ก็จะได้ผลดังตาราง 13.4.1

ตาราง 13.4.1 สัมประสิทธิ์ของ A, B และ AB ใน 2^2 แฟกตอเรียล

บรรทัด	อิทธิพล	แฟกตอเรียลและสัมประสิทธิ์			
		(1)	a	b	ab
1	A	-1	+1	-1	+1
2	B	-1	-1	+1	+1
3	AB	+1	-1	-1	+1

สัมประสิทธิ์ในตาราง 13.4.1 เขียนได้ดังนี้

บรรทัดที่ 1 อิทธิพล A ในคอลัมน์ที่มี a ใส่บวก (+1) ไม่มี a ใส่ลบ (-1)

บรรทัดที่ 2 อิทธิพล B ในคอลัมน์ที่มี b ใส่บวก ไม่มี b ใส่ลบ

บรรทัดที่ 3 นำสัมประสิทธิ์ของอิทธิพลทั้งสองคูณกัน เช่น (-1)(-1) = +1

การคำนวณค่า sum of squares

ต่อไปนี้เราก็ใช้สัมประสิทธิ์ดังแสดงในตาราง 13.4.1 มาคำนวณค่า sum of squares (SS) โดยนำสัมประสิทธิ์มาคูณแบบจับคู่กับผลรวมของทรีตเมนต์ แล้วยกกำลังสอง และหารด้วยผลรวมของสัมประสิทธิ์ยกกำลังสองและคูณด้วยจำนวนซ้ำ ดังนี้

$$SS(X_i) = \frac{[\sum C_i T_i]^2}{(\sum C_i^2)(n)} \qquad \ldots(13\text{-}10)$$

เช่น ถ้าจะหา SS(A) ก็อาจคำนวณดังนี้

$$SS(A) = \frac{[(-1)(T_1)+(+1)(T_2)+(-1)(T_3)+(+1)(T_4)]^2}{[(-1)^2+1^2+(-1)^2+1^2]\times n} = \frac{\sum A^2}{4n}$$

ทั้งนี้ค่า -1, 1, -1 และ 1 คือสัมประสิทธิ์ของ A ส่วน T_1, T_2, T_3, T_4 คือผลรวมทรีตเมนต์ (1), a, b และ ab ตามลำดับ ทำนองเดียวกันหาได้ว่า

$$SS(B) = \frac{\sum B^2}{4n}$$

$$SS(AB) = \frac{\sum AB^2}{4n}$$

ตัวอย่าง ในการทดลองติดตาระหว่างส้ม 2 ชนิด โดยใช้แฟกเตอร์ดังนี้

a_0 = ตาไม่มีเนื้อไม้ b_0 = ตัดต้นตอทิ้งทันทีเมื่อตางอก

a_1 = ตามีเนื้อไม้ b_1 = ตัดทิ้งเมื่อตางอกแล้ว 5 ซม.

เมื่อสำรวจเปอร์เซ็นต์การติดตาที่ประสบผลสำเร็จได้ผลดังตาราง 13.4.2

การวิเคราะห์วาเรียนซ์ดำเนินตามขั้นตอนดังนี้

ขั้นที่ 1 วิเคราะห์วาเรียนซ์ เช่น RCB ที่มี 4 ทรีตเมนต์ และ 4 ซ้ำ ตามวิธีที่กล่าวมาแล้ว

$$\text{Correction factor (CF)} = \frac{\sum(X_{ijk})^2}{nab} = \frac{(570.75)^2}{(4)(2)(2)} = 20{,}359.72$$

$$\text{Total SS (TSS)} = 3{,}949.51 \; (df = 15)$$

$$\text{Block SS (SSB)} = 208.14 \; (df = 3)$$

$$\text{Treatment SS (SSTr)} = 3{,}213.64 \; (df = 3)$$

$$\text{Error SS (SSE)} = 527.73 \; (df = 9)$$

ตาราง 13.4.2 เปอร์เซ็นต์ผลสำเร็จจากการติดตาระหว่างส้ม 2 พันธุ์

บล็อก	ทรีตเมนต์				รวม
	a_0b_0(1)	a_1b_0(a)	a_0b_1(b)	a_1b_1(ab)	
I	6.42	28.66	33.21	53.13	121.42
II	35.06	21.97	45.00	60.00	162.03
III	24.35	20.27	39.82	60.67	145.11
IV	18.44	35.06	30.00	58.69	142.19
รวม	84.27	105.96	148.03	232.49	570.75

ขั้นที่ 2 วิเคราะห์อิทธิพลหลัก A, B และปฏิกิริยา AB ตามวิธีที่แสดงไว้ในตาราง 13.4.1 ก่อนวิเคราะห์ให้จัดตารางโดยนำผลรวมของทรีตเมนต์วางบนหัวตารางดังนี้

อิทธิพลหลักและปฏิกิริยา	ทรีตเมนต์และผลรวม				Sum of Squares (SS)
	$a_0b_0(1)$	$a_1b_0(a)$	$a_0b_1(b)$	$a_1b_1(ab)$	
	84.27	105.96	148.03	232.49	
A	-1	1	-1	1	SS(A) = 704.24
B	-1	-1	1	1	SS(B) = 2,263.14
AB	1	-1	-1	1	SS(AB) = 246.25

ซึ่งอาจคำนวณได้ว่า

$$SS(A) = \frac{[(-1)(84.27)+(1)(105.96)+(-1)(148.03)+(1)(232.49)]^2}{[(-1)^2+1^2+(1)^2+1^2]\times 4}$$

$$= \frac{(106.15)^2}{16}$$

$$= 704.24$$

ในทำนองเดียวกันก็คำนวณได้ว่า

$$SS(B) = 2,263.14$$

$$SS(AB) = 246.25$$

ทั้งนี้ SS(A) + SS(B) + SS(AB) ที่คำนวณโดยวิธีนี้จะเท่ากับ SSTr ที่หาได้ในขั้นที่ 1 ที่น่าสังเกตก็คือในการวิเคราะห์แบบนี้ ทั้งอิทธิพลหลักและปฏิกิริยามี df เท่ากับ 1 เสมอ และเมื่อรวมกันแล้วก็ได้เท่ากับ df ของทรีตเมนต์นั่นเอง

ขั้นที่ 3 นำผลการวิเคราะห์วาเรียนซ์ลงตารางวิเคราะห์วาเรียนซ์ ดังตาราง 13.4.4

ตาราง 13.4.4 ผลของการวิเคราะห์วาเรียนซ์ของผลการติดตาของส้ม 2 พันธุ์ จากตาราง 13.4.2

Source	df	SS	MS	F	F (table) 5%	F (table) 1%
Blocks	3	208.14				
Treatments	(3)	(3,213.64)				
A	1	704.24	704.24	12.01**	5.12	10.56
B	1	2,263.14	2,263.14	38.60**	5.12	10.56
AB	1	246.25	246.25	4.20	5.12	10.56
Error	9	527.25	58.64			
Total	15	3,949.51				

** = แตกต่างทางสถิติที่ระดับ 1 เปอร์เซ็นต์

เมื่อเปรียบเทียบกับค่า $F_{0.01}$ df 1, 9 = 10.55 ก็ปรากฏว่าความแตกต่างในอิทธิพลหลัก A, B มีนัยสำคัญยิ่งในทางสถิติ ส่วนปฏิกิริยาระหว่าง A และ B ไม่มีนัยสำคัญแต่อย่างใด

13.5 แฟกตอเรียลที่มีสามแฟกเตอร์

ในการทดลองโดยใช้แฟกตอเรียลอาจเพิ่มจำนวนปัญหาที่ศึกษา หรือแฟกเตอร์เป็น 3 ชนิดก็ได้ แต่ละชนิดมีระดับต่าง ๆ กัน เช่น ถ้ามีแฟกเตอร์ A 2 ระดับ, B 3 ระดับ และ C 2 ระดับ ก็จัดทรีตเมนต์เปนแฟกตอเรียลได้ทั้งสิ้น 2 x 3 x 2 = 12 ทรีตเมนต์ เมื่อมีปัญหามากขึ้นวิธีการวิเคราะห์ก็เพิ่มความสลับซับซ้อนขึ้นไปด้วย เช่น การทดลองเปรียบเทียบพันธุ์พืช a พันธุ์ กับปุ๋ย b ระดับ และระยะปลูก c ระยะ โดยใช้แผนการทดลองแบบ RCB ก็อาจแสดงแหล่งของความปรวนแปรได้ดังตาราง 13.5.1 ทั้งนี้แตกต่างจากการวิเคราะห์วาเรียนซ์ที่มี 2 แฟกเตอร์ โดยที่มีปฏิกิริยาระหว่าง แฟกเตอร์เพิ่มขึ้น ทั้งนี้ให้ n = จำนวนบลอกหรือซ้ำ a, b และ c = จำนวนระดับของแฟกเตอร์ A, B และ C ตามลำดับ และมีแบบจำลองทางคณิตศาสตร์ ดังสมการ (13-11)

ตาราง 13.5.1 แสดงแหล่งของความปรวนแปรและ df ในการทดลองโดยใช้แฟกตอเรียลของ 3 แฟกเตอร์

Sources	df	SS	MS	F	F(table)
Blocks	n – 1				
Treatments	abc – 1				
Varieties (A)	a – 1	$\sum A^2 / nbc - CF$			
Fertilizers (B)	b – 1	$\sum B^2 / nac - CF$			
Spacings (C)	c – 1	$\sum C^2 / nab - CF$			
AB	(a – 1)(b – 1)	$\sum AB^2 / nc - CF$			
AC	(a – 1)(c – 1)	$\sum AC^2 / nb - CF$			
BC	(b – 1)(c – 1)	$\sum BC^2 / na - CF$			
ABC	(a – 1)(b – 1)(c – 1)	$\sum ABC^2 / n - CF$			
Error	(n – 1)(abc – 1)				
Total	nabc – 1				

ตัวอย่าง ในการทดลองเปรียบเทียบพันธุ์อ้อย 2 พันธุ์ (V_1, V_2) ใช้ปุ๋ย 3 อัตรา (F_1, F_2, F_3) และระยะปลูก 2 ระยะ (S_1, S_2) โดยใช้แผนการทดลองแบบ RCB จำนวน 4 บลอก ได้ผลดังตาราง 13.5.2

การจัดแฟกตอเรียล 3 แฟกเตอร์ใช้วิธีการทำตารางตาหมากรุกดังนี้

(1) พันธุ์ x ปุ๋ย

		ปุ๋ย		
		F_1	F_2	F_3
พันธุ์	V_1	V_1F_1	V_1F_2	V_1F_3
	V_2	V_2F_1	V_2F_2	V_2F_3

(2) ระยะปลูก x (พันธุ์ x ปุ๋ย)

		พันธุ์ - ปุ๋ย					
		V_1F_1	V_1F_2	V_1F_3	V_2F_1	V_2F_2	V_2F_3
ระยะปลูก	S_1	$V_1F_1S_1$	$V_1F_2S_1$	$V_1F_3S_1$	$V_2F_1S_1$	$V_2F_2S_1$	$V_2F_3S_1$
	S_2	$V_1F_1S_2$	$V_1F_2S_2$	$V_1F_3S_2$	$V_2F_1S_2$	$V_2F_2S_2$	$V_2F_3S_2$

แล้วนำชุดแฟกตอเรียลในตารางสุดท้ายจำนวน 12 ชุด ซึ่งเทียบเท่ากับจำนวน 12 ทรีตเมนต์ ไปทดลองโดยใช้แผนการทดลองแบบ RCB ดังที่ได้อธิบายมาแล้วในบทที่ 10 ต่อไป และเก็บข้อมูลดังแสดงในตาราง 13.5.2 การทดลองมีแบบจำลองทางคณิตศาสตร์ดังนี้

$$X_{ijk} = \mu + B_i + \alpha_j + \beta_k + \gamma_\ell + (\alpha\beta)_{jk} + (\alpha\gamma)_{j\ell} + (\beta\gamma)_{k\ell} + (\alpha\beta\gamma)_{jk\ell} + \varepsilon_{ijk\ell} \quad \ldots(13\text{-}11)$$

เมื่อกำหนดให้

μ คือค่าเฉลี่ยของประชากร

B, α, β, γ คือผลของบลอกและอิทธิพลหลักเนื่องจากแฟกเตอร์ A, B และ C ตามลำดับ

$\alpha\beta, \alpha\gamma, \beta\gamma, \alpha\beta\gamma$ คือปฏิกิริยาระหว่างแฟกเตอร์ในระดับ 2 และ 3 แฟกเตอร์

$\varepsilon_{ijk\ell}$ คือความคลาดเคลื่อนของการทดลอง

ทั้งนี้ $i = 1, 2, \ldots, n$ โดยที่ n คือจำนวนบลอก

$j = 1, 2, \ldots, a$ เมื่อ a เป็นจำนวนระดับแฟกเตอร์ A (พันธุ์พืช)

$k = 1, 2, \ldots, b$ เมื่อ b เป็นจำนวนระดับแฟกเตอร์ B (ระดับปุ๋ย)

$\ell = 1, 2, \ldots, c$ เมื่อ c เป็นจำนวนระดับแฟกเตอร์ C (ระยะปลูก)

ถ้าปัญหาที่ใช้ทดลองเป็นปัจจัยคงที่ การตั้งสมมุติฐานก็คล้ายกับที่กล่าวมาแล้วในตอน 13.3

การวิเคราะห์วาเรียนซ์ ข้อมูลในตาราง 13.5.2

ขั้นที่ 1 วิเคราะห์เช่นเดียวกับการทดลอง RCB ได้ผลดังนี้

$$\text{Corection factor (CF)} = \frac{\sum X_{ijkl}^2}{nabc} = \frac{(429)^2}{48} = 3{,}834.19$$

$$\text{Total SS(TSS)} = 4{,}369.00 - 3{,}834.19 = 530.81$$

$$\text{Block SS (SSB)} = 3{,}841.75 - 3{,}834.19 = 7.56$$

$$\text{Treatment SS(SSTr)} = 4{,}235.75 - 3{,}834.19 = 401.56$$

$$\text{SSE} = \text{TSS} - \text{SSB} - \text{SSTr}$$

$$= 530.81 - 7.56 - 401.56 = 121.69$$

อย่างไรก็ดีในการทดลองแบบแฟกตอเรียลไม่จำเป็นต้องวิเคราะห์ SSTr และ SSE ในขั้นนี้ก็ได้

ขั้นที่ 2 วิเคราะห์เพื่อแยก SSTr ออกเป็นส่วนย่อย ๆ ดังแสดงในตาราง 13.5.1 ก่อนการวิเคราะห์ต้องจัดทำตารางปฏิกิริยาระหว่างปัจจัยในครบดังแสดงในตาราง 13.5.3 เสียก่อนแล้วดำเนินการต่อไป ดังนี้

$$\text{SS(A)(Varieties)} = \frac{\sum X_j^2}{nbc} - CF = \frac{\sum A^2}{nbc} - CF$$

$$= \frac{(161^2 + 268^2)}{24} - 3{,}834.19$$

$$= 238.52$$

$$\text{SS(B)(Fertilizers)} = \frac{\sum X_k^2}{nac} - CF = \frac{\sum B^2}{nac} - CF$$

$$= \frac{(130^2 + 162^2 + 137^2)}{16} - 3{,}834.19$$

$$= 35.37$$

$$\text{SS(C) (Spacings)} = \frac{\sum X_\ell^2}{nab} - CF = \frac{\sum C^2}{nab} - CF$$

$$= \frac{(184^2 + 245^2)}{24} - 3{,}834.19$$

$$= 77.52$$

ตาราง 13.5.2 ผลผลิตของอ้อย 2 พันธุ์ เมื่อใช้ปุ๋ย 3 ระดับ และใช้ระยะปลูก 2 ระยะ ใช้แผนการทดลอง RCB

พันธุ์ (V)	ปุ๋ย (F)	ระยะปลูก (S)	บลอก I	บลอก II	บลอก III	บลอก IV	รวม
V_1	F_1	S_1	4	4	3	4	15
		S_2	8	4	10	7	29
	F_2	S_1	8	7	8	10	33
		S_2	8	12	7	8	35
	F_3	S_1	5	7	5	6	23
		S_2	7	8	5	6	26
V_2	F_1	S_1	9	6	7	9	31
		S_2	15	14	16	10	55
	F_2	S_1	11	9	13	10	43
		S_2	13	14	11	13	51
	F_3	S_1	10	14	6	9	39
		S_2	14	13	12	10	49
	รวม		112	112	103	102	429

ตาราง 13.5.3 ปฏิกิริยาระหว่างแฟกเตอร์

ก. A x B

		ปุ๋ย (B) F_1	F_2	F_3	รวม
พันธุ์ (A)	V_1	44	68	49	161
	V_2	86	94	88	268
	รวม	130	162	137	

ข. A x C

		ระยะปลูก (C) S_1	S_2	รวม
พันธุ์ (A)	V_1	71	90	161
	V_2	113	155	268
		184	245	

ค. **B x C**

		ระยะปลูก (C)		รวม
		S_1	S_2	
ปุ๋ย (B)	F_1	46	84	130
	F_2	76	86	162
	F_3	62	75	137

$$
\begin{aligned}
SS(AB) &= \frac{\sum X_{jk}^2}{nc} - CF - SS(A) - SS(B) \\
&= \frac{\sum AB^2}{nc} - CF - SS(A) - SS(B) \\
&= \frac{(44^2 + 68^2 + \ldots + 88^2)}{8} - 3{,}834.19 - 238.52 - 35.37 \\
&= 9.04
\end{aligned}
$$

$$
\begin{aligned}
SS(AC) &= \frac{\sum X_{j\ell}^2}{nb} - CF - SS(A) - SS(C) \\
&= \frac{\sum AC^2}{nb} - CF - SS(A) - SS(C) \\
&= \frac{(71^2 + 90^2 + \ldots + 155^2)}{12} - 3{,}834.19 - 238.52 - 77.52 \\
&= 11.02
\end{aligned}
$$

$$
\begin{aligned}
SS(BC) &= \frac{\sum X_{k\ell}^2}{na} - CF - SS(B) - SS(C) \\
&= \frac{\sum BC^2}{na} - CF - SS(B) - SS(C) \\
&= \frac{(46^2 + 76^2 + \ldots + 75^2)}{8} - 3{,}834.19 - 35.37 - 77.52 \\
&= 29.54
\end{aligned}
$$

$$
\begin{aligned}
SS(ABC) &= \frac{\sum X_{jk\ell}^2}{n} - CF - SS(A) - SS(B) - SS(C) - SS(AB) - SS(AC) \\
&\quad - SS(BC) \\
&= SSTr - SS(A) - SS(B) - SS(C) - SS(AB) - SS(AC) \\
&= 401.56 - 238.52 - 35.37 - 77.52 - 9.04 - 11.02 - 29.54 \\
&= 0.55
\end{aligned}
$$

ขั้นที่ 3 นำผลการวิเคราะห์ทั้งหมดลงตารางวิเคราะห์วาเรียนซ์ ดังตาราง 13.5.4 ปฏิกิริยาระหว่างปัจจัย AB, AC และ BC เรียกว่า ปฏิกิริยาระหว่าง 2 ปัจจัย[7] ส่วนปฏิกิริยา ABC เรียกว่า ปฏิกิริยา 3 ปัจจัย[8]

ตาราง 13.5.4 ผลการวิเคราะห์วาเรียนซ์ของข้อมูลในตาราง 13.5.2

Sources	df	SS	MS	F[1]
Blocks	3	7.56	2.52	
Treatments	11	401.56	36.51	9.90*
Varieties (A)	1	238.52	238.52	64.63**
Fertilizers (B)	2	35.38	17.69	4.79
Spacings (C)	1	77.52	77.52	21.02**
AB	2	9.04	4.52	1.22
AC	1	11.02	11.02	2.99
BC	2	29.54	14.77	4.00*
ABC	2	0.55	0.28	0.00
Error	33	121.68	3.69	
Total	47	530.81		

***, ** แตกต่างในทางสถิติในระดับ 5 และ 1 เปอร์เซ็นต์**

[1] สมมุติว่าทุกปัจจัยเป็นปัจจัยคงที่ ถ้าเป็นปัจจัยสุ่มวิธีการทดสอบแสดงไว้ในตาราง 13.7.3

ในการวิเคราะห์วาเรียนซ์ พบว่าพันธุ์ ระยะปลูก และปฏิกิริยาระหว่างปุ๋ยและระยะปลูกมีความแตกต่างในทางสถิติ

13.6 การวิเคราะห์ 2^3 แฟกตอเรียล

เราอาจใช้วิธีการที่คล้ายคลึงกับวิธีในตอน 13.4 เพื่อวิเคราะห์แฟกตอเรียลที่เกิดจากแฟกเตอร์มากกว่าสองชนิด ชนิดละสองระดับ คือการทดลองที่มีทรีตเมนต์ 2^n แฟกตอเรียล ตัวอย่างเช่น การทดลองเปรียบเทียบแฟกเตอร์ A, B และ C แต่ละแฟกเตอร์มี 2 ระดับ คือ $A(a_0, a_1)$, $B(b_0, b_1)$ และ $C(c_0, c_1)$ ดังนั้นเราอาจจัดชุดของทรีตเมนต์ได้ดังนี้

$$a_0b_0c_0,\ a_1b_0c_0,\ a_0b_1c_0,\ a_1b_1c_0,\ a_0b_0c_1,\ a_1b_0c_1,\ a_0b_1c_1,\ a_1b_1c_1$$

ซึ่งได้ 8 ชุด เราอาจจะเขียนให้ดูง่ายขึ้น โดยตัดเลขห้อยท้ายออก ก็จะได้เป็นลำดับ (ตามข้างบน) ดังนี้

$$(1),\ a,\ b,\ ab,\ c,\ ac,\ bc,\ abc \qquad \text{...(13-12)}$$

คือ $a_0b_0c_0 = (1), a_1b_0c_0 = a, ..., a_1b_1c_1 = abc$ ต่อจากนี้ก็กำหนดสัมประสิทธิ์ของอิทธิพลหลักและปฏิกิริยา ดังตาราง 13.6.1

ตาราง 13.6.1 วิธีกำหนดสัมประสิทธิ์ของชุดเปรียบเทียบ 2^3 แฟกตอเรียล

บรรทัด	อิทธิพลหลักและฏิกิริยา	ชุดแฟกตอเรียล							
		(1)	a	b	ab	c	ac	bc	abc
1	A	−1	+1	−1	+1	−1	+1	−1	+1
2	B	−1	−1	+1	+1	−1	−1	+1	+1
3	AB	+1	−1	−1	+1	+1	−1	−1	+1
4	C	−1	−1	−1	−1	+1	+1	+1	+1
5	AC	+1	−1	+1	−1	−1	+1	−1	+1
6	BC	+1	+1	−1	−1	−1	−1	+1	+1
7	ABC	−1	+1	+1	−1	+1	−1	−1	+1

วิธีการเขียนสัมประสิทธิ์ให้ถือหลักดังนี้ คือ ถ้าไม่มีอิทธิพลหลักในทรีตเมนต์ใด ให้ใส่ −1 ถ้ามีให้ใส่ +1 เช่น A ในบรรทัดที่ 1 ใส่ −1, +1, −1 ,..., +1 เป็นต้น ส่วนสัมประสิทธิ์ของปฏิกิริยาเกิดจากผลคูณของสัมประสิทธิ์ของอิทธิพลหลักนั่นเอง ชุดเปรียบเทียบทั้งหมดเป็น orthogonal contrast คือผลบวกของแต่ละชุดเท่ากับศูนย์ และผลบวกของผลคูณระหว่างชุดเปรียบเทียบแต่ละคู่ก็เท่ากับศูนย์ เช่น

$$\begin{aligned}\sum A &= (-1)+(+1)+(-1)+(+1)+(-1)+(+1)+(-1)+(+1)=0\\ &= (-1)+(-1)+(+1)+(+1)+(-1)+(-1)+(-1)+(+1)=0\\ &= (-1)(-1)+(+1)(-1)+(-1)(+1)+(+1)(+1)+(-1)(-1)+(+1)(-1)+(-1)(+1)\\ &\quad +(+1)(+1)=0\end{aligned}$$

การวิเคราะห์วาเรียนซ์โดยใช้วิธีการดังกล่าวอาจแสดงโดยใช้ข้อมูลในตาราง 13.6.2

ตาราง 13.6.2 ขนาดของเมล็ดถั่วเขียวที่ได้รับปุ๋ย N, P และ K (กรัม/100 เมล็ด)

ชุดทรีตเมนต์ (แฟกตอเรียล)	บลอก I	บลอก II	บลอก III	รวม
(1)(000)[1]	6	5	5	16
N(100)	6	6	5	17
P(010)	7	6	6	19
NP(110)	7	7	6	20
K(001)	7	6	7	20
NK(101)	5	6	5	16
PK(011)	6	5	6	17
NPK(111)	6	5	6	17
รวม	50	46	46	142

[(1)]**หมายเหตุ สำหรับปุ๋ย N, P และ K ; 0 = ไม่ใส่ ; 1 ใส่อัตราที่เหมาะสม**

ซึ่งอาจนำผลการทดลองมาทำการวิเคราะห์ได้ดังนี้:

ขั้นที่ 1 ทำการวิเคราะห์เหมือนการทดลอง RCB ปกติ

$$\text{Correction factor (CF)} = \frac{\sum X_{ijk}^2}{nabc} = \frac{142^2}{24} = 840.17$$

$$\text{Total SS(TSS)} = 6^2 + 5^2 + \ldots + 6^2 - 840.17 = 11.83$$

$$\text{Block SS(SSB)} = \frac{[50^2 + 46^2 + 46^2]}{8} - 840.17 = 1.33$$

$$\text{Treatment SS (SSTr)} = \frac{[16^2 + 17^2 + \ldots + 17^2]}{3} - 840.17 = 6.50$$

$$\text{Error SS(SSE)} = \text{TSS} - \text{SSB} - \text{SSTr}$$

$$= 11.83 - 1.33 - 6.50 = 4.00$$

ขั้นที่ 2 ทำการวิเคราะห์เพื่อหาผลของอิทธิพลหลักและปฏิกิริยา คือวิเคราะห์เพื่อแยก SSTr ออกเป็น SS(N), SS(P), SS(K),…, SS(NPK) โดยใช้สมการเช่น

$$SS(N) = \frac{\sum c_i T_i^2}{(\sum c_i^2)(n)} \qquad \ldots(13\text{-}13)$$

เมื่อ

c_i = สัมประสิทธิ์ของชุดแฟกตอเรียล

T_i = ผลบวกของทรีตเมนต์ที่ i

n = จำนวนบลอกหรือซ้ำ

ในกรณีของตัวอย่างในตาราง 13.6.2, $n = 3$, $\sum c_i^2 = 8$ เสมอเพราะมีสัมประสิทธิ์ 8 ค่า ดู ตาราง 13.6.1 เช่น

สัมประสิทธิ์ของ $SS(N) = (-1)^2 + 1^2 + (-1)^2 + 1^2 + (-1)^2 + 1^2 + (-1)^2 + 1^2 = 8$

ก่อนการวิเคราะห์โดยใช้วิธีดังกล่าว ต้องนำผลรวมของทรีตเมนต์มาเรียงลำดับ แล้วนำผลของอิทธิพลหลักหรือปฏิกิริยาเข้าไปจับคู่ดังตัวอย่าง สัมประสิทธิ์ของชุดเปรียบเทียบที่สมบูรณ์ดูในตาราง 13.6.1

อิทธิพล	ทรีตเมนต์หรือแฟกตอเรียล								อธิบาย
	(1)	N	P	PN	K	NK	PK	NPK	
ผลรวม T_i	16	17	19	20	20	16	17	17	
N	−1	+1	−1	+1	−1	+1	−1	+1	มี N ใส่ +1
P	−1	−1	+1	+1	−1	−1	+1	+1	มี P ใส่ +1
NP	+1	−1	−1	+1	+1	−1	−1	+1	N คูณ P

แล้วสามารถวิเคราะห์โดยใช้สมการ (13-13) ดังนี้

$$SS(N) = \frac{1}{(8)(3)}[(-1)(16) + (1)(17) + \ldots + (1)(17)]^2 = 0.17$$

$$SS(P) = \frac{1}{(8)(3)}[(-1)(16) + (-1)(17) + \ldots + (1)(19)]^2 = 0.67$$

$$SS(NP) = \frac{1}{(8)(3)}[(1)(16) + (-1)(17) + \ldots + (1)(17)]^2 = 0.67$$

$$SS(K) = \frac{1}{(8)(3)}[(-1)(16) + (-1)(17) + \ldots + (1)(17)]^2 = 0.17$$

$$SS(NK) = \frac{1}{(8)(3)}[(1)(16) + (-1)(17) + \ldots + (1)(17)]^2 = 1.50$$

$$SS(PK) = \frac{1}{(8)(3)}[(1)(16)+(1)(17)+...+(1)(17)]^2 = 2.67$$

$$SS(NPK) = \frac{1}{(8)(3)}[(-1)(16)+(1)(17)+...+(1)(17)]^2 = 0.67$$

ถ้ายังไม่หา SSE ไว้ก่อน ก็คำนวณต่อไปว่า

$$SSE = TSS - SSB - SS(N) - SS(P) = ... - SS(NPK)$$

$$= 11.83 - 1.33 - 0.17 - ... - 0.67 = 4.00$$

ขั้นที่ 3 นำค่าต่าง ๆ ลงตารางวิเคราะห์วาเรียนซ์ 13.6.2 ต่อไป

ตาราง 13.6.2 ผลการวิเคราะห์วาเรียนซ์ของข้อมูลในตาราง 13.6.2

Sources	df	SS	MS	F
Blocks	n - 1 = 2	1.33	0.67	2.39
N	2 - 1 = 1	0.17	0.17	0.61
P	2 - 1 = 1	0.67	0.67	20.39
K	2 - 1 = 1	0.17	0.17	0.61
NP	(2 - 1)(2 - 1) = 1	0.67	0.67	2.39
NK	(2 - 1)(2 - 1) = 1	1.50	1.50	5.36
PK	(2 - 1)(2 - 1) = 1	2.67	2.67	9.34
NPK	(2 - 1)(2 - 1)(2 - 1) = 1	0.67	0.67	2.39
Error	(n - 1)(abc-1) = 14	4.00	0.28	
Total	n(abc) - 1 = 23	11.83		

13.7 ค่าคาดหมายของมีนสแควร์

ค่ามีนสแควร์ที่ได้จากการวิเคราะห์วาเรียนซ์ประกอบด้วยส่วนย่อย ๆ หลายชนิด มีนสแควร์ที่แสดงส่วนย่อย ๆ นี้เรียกว่าคาดหมายของมีนสแควร์[6] ซึ่งเราอาจเขียนย่อ ๆ ว่า EMS

ในการทดลองแบบแฟกตอเรียลที่มี 2 แฟกเตอร์ ธรรมชาติของปัจจัยอาจเป็นได้หลายแบบ คือทั้ง A และ B เป็นปัจจัยคงที่ หรือทั้งสองชนิดเป็นปัจจัยสุ่ม หรือฝ่ายใดฝ่ายหนึ่งเป็นปัจจัยสุ่ม เช่นนี้เราสามารถหาค่าคาดหมายมีนสแควร์โดยใช้วิธีการคล้ายกับที่กล่าวถึงในตอน 9.6 ในกรณีของแฟกตอเรียลนั้นถ้า A และ B หรือเพียงฝ่ายใดฝ่ายหนึ่งเป็นปัจจัยสุ่ม ปฏิกิริยาระหว่าง A และ B ก็เป็นปัจจัยสุ่มไปด้วย ค่าคาดหมายของมีนสแควร์ต่าง ๆ แสดงไว้ในตาราง 13.7.1

ตาราง 13.7.1 แสดง EMS ของการทดลองแบบแฟกตอเรียลที่มี 2 แฟกเตอร์ เมื่อเป็นปัจจัยคงที่ (Model I) ปัจจัยสุ่ม (Model II) และแบบผสม (Model III)

Sources	Model I (A, B fixed)	Model II (A, B random)	Model III (A fixed, B random)
A	$\sigma^2 + nbK_a^2$	$\sigma^2 + n\sigma_{ab}^2 + nb\sigma_a^2$	$\sigma^2 + n\sigma_{ab}^2 + nbK_a^2$
B	$\sigma^2 + naK_b^2$	$\sigma^2 + n\sigma_{ab}^2 + nb\sigma_b^2$	$\sigma^2 + na\sigma_b^2$
AB	$\sigma^2 + nK_{ab}^2$	$\sigma^2 + n\sigma_{ab}^2$	$\sigma^2 + n\sigma_{ab}^2$
Error	σ^2	σ^2	σ^2

ในการทดสอบโดยใช้ F-test นั้น ถ้า A และ B เป็นปัจจัยคงที่ก็ใช้มีนสแควร์ของ error เป็นตัวหารทุกรายการ ถ้า A และ B เป็นปัจจัยสุ่มก็ใช้มีนสแควร์ของ error หรือปฏิกิริยา AB เป็นตัวหาร ส่วนในโมเดลผสมก็เลือกตัวหารตามความเหมาะสม เห็นได้ว่าการกำหนดตัวหารขึ้นอยู่กับจำนวนองค์ประกอบย่อยที่ใช้เป็นตัวตั้ง เช่น ในโมเดลที่ 2 เราต้องเริ่มต้นด้วยการทดสอบ MS(AB) เสียก่อน

$$F = \frac{MS(AB)}{MSE} \quad \text{(vs F(.01, .05) df AB และ error)}$$

ถ้าไม่แตกต่างในทางสถิติก็แสดงว่า σ_{ab}^2 ไม่สำคัญ ก็สามารถใช้ MSE ทดสอบได้ทุกรายการ แต่ถ้าแตกต่างทางสถิติก็ต้องใช้ MS(AB) ทดสอบ MS(A) และ MS(B) ต่อไป

วิธีการเขียน EMS

ก. วิธีที่ 1

การเขียน EMS ตามที่แสดงไว้ในบทที่ 9 (ตอน 9.6) เสียเวลาและมีความยุ่งยากมาก โดยเฉพาะอย่างยิ่ง เมื่อมีส่วนประกอบของวาเรียนซ์มากขึ้น

วิธีการดังต่อไปนี้ เป็นวิธีการที่สะดวกและง่าย และอาจขยายจำนวนได้ตามต้องการ ซึ่งแสดงวิธีการ โดยการใช้แฟกตอเรียลที่มี 2 แฟกเตอร์ ดังนี้

แบบจำลองทางคณิตศาสตร์

$$X_{ijk} = \mu + A_i + B_j + AB_{ij} + \varepsilon_{(ij)k}$$

1. เขียนความแปรวนแปรในแบบจำลองทางคณิตศาสตร์ ลงในสดมภ์แรกของตารางดังนี้

A_i
B_j
AB_{ij}
$\varepsilon_{(ij)k}$

2. เขียนตัวห้อยท้าย i, j, และ k ของ A, B และ $\varepsilon_{(ij)k}$ ลงในหัวสดมภ์ แล้วเขียน F ถ้าผลของทรีตเมนต์เป็นแบบคงที่ และเขียน R ถ้าเป็นอย่างสุ่ม และเขียนจำนวนตัวแปร (a, b, n) ของแต่ละค่าห้อยท้าย ดังนี้

	a F i	b R j	n R k
A_i			
B_j			
AB_{ij}			
$\varepsilon_{(ij)k}$			

3. ในแต่ละแถว ใส่จำนวนตัวแปร (a, b, n) ที่ไม่เกี่ยวข้องกับตัวห้อยท้าย

	a F i	b R j	n R k
A_i		b	n
B_j	a		n
AB_{ij}			n
$\varepsilon_{(ij)k}$			

4. สำหรับตัวห้อยท้ายที่อยู่ในวงเล็บ ใส่ 1 (ตรงตำแหน่งที่ตรงกับตัวห้อยท้ายชนิดเดียวกันในแถวบน)

	a F i	b R j	n R k
A_i		b	n
B_j	a		n
AB_{ij}			n
$\varepsilon_{(ij)k}$	1	1	

5. ในช่องว่างใส่ 0 ถ้าหากปัจจัยคงที่ (F) และใส่ 1 ถ้าเป็นปัจจัยสุ่ม (R)

	a	b	n
	F	R	R
	i	j	k
A_i	0	b	n
B_j	a	1	n
AB_{ij}	0	1	n
$\varepsilon_{(ij)k}$	1	1	1

6. ในการหา EMS ของแต่ละแฟกเตอร์ในแบบจำลองทางคณิตศาสตร์ อาจหาสัมประสิทธิ์ของวาเรียนซ์แต่ละค่าดังนี้

(1) ปิดสดมภ์ทีละสดมภ์ ปิดเฉพาะสดมภ์ที่ตัวห้อยท้ายไม่อยู่ในวงเล็บ เพื่อหาสัมประสิทธิ์ เช่น สัมประสิทธิ์ของ A_i ก็ปิดสดมภ์ i

(2) หลังจากปิดสดมภ์ดังกล่าวแล้วก็คูณระหว่างสดมภ์ที่เหลือ ผลคูณเหล่านี้คือ สัมประสิทธิ์ของวาเรียนซ์ที่เกี่ยวกับแฟกเตอร์นั้น เช่น เมื่อปิดสดมภ์ i แล้วคูณกันก็ได้ bn, n, n และ 1 นำค่านี้ไปคูณกับวาเรียนซ์ แล้วบวกกันก็จะได้ EMS ของ A_i ดังนี้ : $n\sigma^2_a + n\sigma^2_{ab} + 1.\sigma^2$ หรือเท่ากับ $\sigma^2 + n\sigma^2_{ab} + nb\sigma^2_a$ ทั้งนี้ไม่ได้ใช้ n ตัวแรก เพราะไม่มี i ในตัวห้อยท้ายของ B เมื่อดำเนินการในขั้นตอนกล่าวแล้วข้างบนจนครบก็ได้ EMS ดังนี้

Factor	a F i	b R j	n R k	EMS
A_i	0	b	1	$\sigma^2 + n\sigma^2_{ab} + nbk^2_a$
B_j	a	1	n	$\sigma^2 + na\sigma^2_b$
AB_{ij}	0	1	n	$\sigma^2 + n\sigma^2_{ab}$
$\varepsilon_{(ij)k}$	1	1	1	σ^2

โดยใช้วิธีการเดียวกันนี้ ก็สามารถหา EMS ของปัจจัยคงที่และปัจจัยสุ่มได้ทุกชนิด เช่น ถ้าเป็นปัจจัยสุ่มทั้งหมด ก็หาได้ดังนี้

Factor	a F i	b R j	n R k	EMS
A_i	1	b	n	$\sigma^2 + n\sigma^2_{ab} + nbK^2_a$
B_j	a	1	n	$\sigma^2 + n\sigma^2_{ab} + na\sigma^2_b$
AB_{ij}	1	1	n	$\sigma^2 + n\sigma^2_{ab}$
$\varepsilon_{(ij)k}$	1	1	1	σ^2

ข. วิธีที่ 2

สมมติว่าทำการศึกษา 3 แฟกเตอร์ คือ A, B และ C และมี a, b และ c ระดับตามลำดับ การทดลองกระทำ n ซ้ำ ก่อนเขียน EMS ต้องแยกเสียก่อนว่าแฟกเตอร์ใดเป็นปัจจัยคงที่ แฟกเตอร์ใดเป็นปัจจัยสุ่ม วิธีการเขียน EMS อาจแยกได้ตามคุณสมบัติของแฟกเตอร์ดังนี้

1. เมื่อ A, B และ C ต่างก็เป็นปัจจัยสุ่ม

(1) เขียนรายการความปรวนแปรหรือ Source โดยใช้อักษรตัวใหญ่ คือ A, B, C, AB, AC, BC และ ABC (เหมือนสดมภ์แรกของตาราง 13.5.4) ลงในตาราง 13.7.2

(2) เขียนวาเรียนซ์ของแต่ละแหล่งความปรวนแปรในข้อ (1) โดยใช้อักษรตัวเล็กเป็นตัวกำหนดห้อยท้าย เช่น σ^2_a แสดงวาเรียนซ์ของแฟกเตอร์ A, σ^2_{abc} แสดงวาเรียนซ์ของ ABC

(3) วาเรียนซ์ทุกค่าในข้อ (2) มีสัมประสิทธิ์ สัมประสิทธิ์ของวาเรียนซ์ใดเท่ากับจำนวนค่าสังเกตทั้งหมด (nabc) หารด้วยแหล่งของความปรวนแปรนั้น เช่น สัมประสิทธิ์ของ σ^2_a = nabc/A = nbc อย่างไรก็ดีสัมประสิทธิ์ของวาเรียนซ์ของความคลาดเคลื่อน หรือ σ^2 เท่ากับ 1 เสมอ

(4) ทุก ๆ แหล่งของความปรวนแปรมีส่วนประกอบย่อย ๆ ดังนี้คือ : (1) วาเรียนซ์ของความคลาดเคลื่อน (σ^2), (2) วาเรียนซ์ของตัวเอง และ (3) วาเรียนซ์อื่น ๆ ที่แหล่งของความปรวนแปรเป็นกำหนดห้อยท้ายอย่างน้อย 1 ค่า เช่น EMS ของ $A = \sigma^2 + n\sigma^2_{abc} + nb\sigma^2_{ac} + nc\sigma^2_{ab} + nbc\sigma^2_a$

อาจแสดงวิธีการข้อ (1) ถึง (4) ดังตาราง 13.7.2

EMS แต่ละบรรทัดในตาราง 13.7.2 เกิดจากวาเรียนซ์ตามเลขที่ในสดมภ์แรกบวกกัน เช่น EMS ของ A = 8 + 7 + 4 + 5 + 1 = $\sigma^2 + n\sigma^2_{abc} + nc\sigma^2_{ab} + nb\sigma^2_{ac} + nbc\sigma^2_a$ ซึ่งอาจสรุปทุกค่าได้ดังตาราง 13.7.3 ซึ่งจัดเป็น EMS ของการทดลองแบบแฟกตอเรียลที่มี 3 แฟกเตอร์ แต่ละแฟกเตอร์เป็นปัจจัยสุ่ม

ตาราง 13.7.2 แสดงวาเรียนซ์ย่อย ๆ ของแต่ละแหล่งของความปรวนแปร

บรรทัดที่	Sources	Variance และสัมประสิทธิ์	EMS ประกอบด้วยวาเรียนซ์ย่อยบรรทัดที่ ...
1	A	$nbc\sigma_a^2$	8 + 7 + 5 +4 + 1
2	B	$nac\sigma_b^2$	8 + 7 + 6 + 4 + 2
3	C	$nab\sigma_c^2$	8 + 7 + 6 +5 + 3
4	AB	$nc\sigma_{ab}^2$	8 + 7 + 4
5	AC	$nb\sigma_{ac}^2$	8 + 7 + 5
6	BC	$na\sigma_{bc}^2$	8 + 7 + 6
7	ABC	$n\sigma_{abc}^2$	8 + 7
8	Error	σ^2	8

ตาราง 13.7.3 EMS ของการทดลองแบบแฟกตอเรียลที่มี 3 แฟคเตอร์ เป็นปัจจัยสุ่ม

Source	MS	EMS
A	M8	$\sigma^2 + n\sigma_{abc}^2 + nc\sigma_{ab}^2 + nb\sigma_{ac}^2 + nbc\sigma_a^2$
B	M7	$\sigma^2 + n\sigma_{abc}^2 + nc\sigma_{ab}^2 + na\sigma_{bc}^2 + nac\sigma_b^2$
C	M6	$\sigma^2 + n\sigma_{abc}^2 + nb\sigma_{ac}^2 + na\sigma_{bc}^2 + nab\sigma_c^2$
AB	M5	$\sigma^2 + n\sigma_{abc}^2 + nc\sigma_{ab}^2$
AC	M4	$\sigma^2 + n\sigma_{abc}^2 + nb\sigma_{ac}^2$
BC	M3	$\sigma^2 + n\sigma_{abc}^2 + na\sigma_{bc}^2$
ABC	M2	$\sigma^2 + n\sigma_{abc}^2$
Error	M1	σ^2

จากตาราง 13.7.3 เห็นได้ว่าค่า F สำหรับการทดสอบสมมุติฐาน จะใช้ MSE หารจากล่างขึ้นไปยังทุกรายการข้างบนไม่ได้อีกต่อไป เพราะไม่ใช่เป็นปัจจัยคงที่ แต่ต้องดัดแปลงตามความเหมาะสมดังนี้

F (ABC) = M2 / M1

F(BC) : ถ้า F (ABC) แตกต่างก็ใช้ M3 / M2

: ถ้า F (ABC) ไม่แตกต่างใช้ M3 / M1

F (AB, AC) : วิธีการเช่นเดียวกับการทดสอบ BC

อย่างไรก็ดี การทดสอบ A, B และ C ใช้ approximate test (F′) ของ Satterthwaite (1946, Cockran and Cox, 1957) เช่น

$$F'(C) = (M6 + M2)/(M4 + M3)$$

โดยมี df ดังนี้

$$\text{df ตัวตั้ง} = (M6 + M2)^2/[(M6^2/V6) + (M2^2/V2)]$$

$$\text{df ตัวหาร} = (M4 + M3)^2/[(M4^2/V4) + (M3^2/V3)]$$

ทั้งนี้ V2, V3, V4, และ V6 คือ df ของแหล่งความปรวนแปรนั้น ๆ ซึ่งจะเห็นได้ว่าเป็นขั้นตอนที่สลับซับซ้อน

2. เมื่อ A, B และ C ต่างก็เป็นปัจจัยคงที่

(1) เขียนรายการความปรวนแปรตามข้อ 1(1)

(2) ทุกแหล่งของความปรวนแปรมีวาเรียนซ์ของความคลาดเคลื่อน (σ^2) อยู่ด้วยเสมอ และมีวาเรียนซ์ของตัวเองพร้อมกับสัมประสิทธิ์เท่านั้น แต่แทน σ^2 ด้วย K^2 ดังตาราง 13.7.4

ตาราง 13.7.4 EMS เมื่อ A, B และ C เป็นปัจจัยคงที่

Sources	EMS
A	$\sigma^2 + nbcK_a^2$
B	$\sigma^2 + nacK_b^2$
C	$\sigma^2 + nabK_c^2$
AB	$\sigma^2 + ncK_{ab}^2$
AC	$\sigma^2 + nbK_{ac}^2$
BC	$\sigma^2 + naK_{bc}^2$
ABC	$\sigma^2 + nK_{abc}^2$
Error	σ^2

การทดสอบในกรณีนี้ไม่มีความยุ่งยาก เพราะสามารถใช้ MSE ทดสอบได้ทุกค่า

3. เมื่อแฟกเตอร์เป็นแบบผสม

แฟกเตอร์แบบผสมคือ บางปัจจัยเป็นแบบคงที่ บางปัจจัยเป็นแบบสุ่ม เช่น ให้ A เป็นปัจจัยสุ่ม และ B, C เป็นแบบคงที่ อาจหา EMS ดังนี้

(1) ให้เริ่มจาก EMS ของปัจจัยสุ่มทุกแฟกเตอร์ดังตาราง 13.7.3

(2) วาเรียนซ์ใด ๆ มีอักษรห้อยท้ายที่เหมือนกับชื่อแหล่งของความปรวนแปรบรรทัดนั้นรวมอยู่ด้วยอย่างน้อย 1 ตัว ให้ตัดอักษรห้อยท้ายตัวนั้นออก เช่น $n\sigma^2_{abc}$ เมื่อตัด a ออกก็เหลือ σ^2_{bc} แต่เราใช้วิธีขีดเส้นใต้ไว้

(3) เมื่อตัดอักษรห้อยท้ายดังข้อ (2) แล้ว ถ้าอักษรห้อยท้ายที่เหลือเป็นของปัจจัยสุ่มทั้งหมด ก็ให้คงวาเรียนซ์ย่อยนั้นไว้ โดยใส่ตัวห้อยท้ายที่ตัดออกคืนที่เดิม แต่ถ้าอักษรที่เหลือเป็นปัจจัยคงที่ทั้งหมดหรือบางส่วน ก็ให้ตัดวาเรียนซ์ย่อยนั้นออก

(4) ให้คงวาเรียนซ์ของแหล่งความปรวนแปรนั้นไว้ในทุกกรณี เช่น ในบรรทัด AB ต้องคง $nc\sigma^2_{ab}$ ไว้

(5) ถ้าอักษรห้อยท้ายทุกตัวเป็นปัจจัยคงที่ ก็ให้ใช้ K^2 แทน ซึ่งอาจแสดง EMS แบบผสมดังตาราง 13.7.5

ส่วนทางด้านซ้ายมือของตาราง 13.7.5 เป็น EMS ของปัจจัยสุ่ม อักษรห้อยท้ายที่ขีดเส้นใต้คือตัวที่ถูกตัดออก เช่น Source A เราตัด a ออกจากอักษรห้อยท้ายของวาเรียนซ์ย่อย 3 ค่า คือจาก abc, ab และ ac ก็เหลือ bc, b และ c ซึ่งต่างก็เป็นอักษรห้อยท้ายของปัจจัยคงที่ เมื่อเป็นเช่นนี้ก็ตัดวาเรียนซ์เหล่านี้ออกทั้งหมด เหลือไว้เฉพาะ $\sigma^2 + nbc\sigma^2_{ab}$ ดังแสดงในส่วนทางด้านขวามือเท่านั้น ในบรรทัดอื่น ๆ ก็ใช้หลักวิธีเดียวกัน เช่นนี้ เราสามารถหา EMS ที่มีแฟกเตอร์แบบผสมได้ดังตาราง 13.7.5

ตาราง 13.7.5 แสดง EMS โดยเป็นปัจจัยสุ่ม B และ C เป็นปัจจัยคงที่*

Sources	EMS (A, B, C เป็นปัจจัยสุ่ม)	EMS (A สุ่ม ; B & C คงที่)
A	$\sigma^2 + n\sigma^2_{\underline{a}bc} + nc\sigma^2_{\underline{a}b} + nb\sigma^2_{\underline{a}c} + nbc\sigma^2_a$	$= \sigma^2 + nbc\sigma^2_a$
B	$\sigma^2 + n\sigma^2_{a\underline{b}c} + nc\sigma^2_{a\underline{b}} + na\sigma^2_{\underline{b}c} + nac\sigma^2_b$	$= \sigma^2 + nc\sigma^2_{ab} + nacK^2_b$
C	$\sigma^2 + n\sigma^2_{ab\underline{c}} + nb\sigma^2_{a\underline{c}} + na\sigma^2_{b\underline{c}} + nab\sigma^2_c$	$= \sigma^2 + nb\sigma^2_{ac} + nabK^2_c$
AB	$\sigma^2 + n\sigma^2_{abc} + nc\sigma^2_{ab}$	$= \sigma^2 + nc\sigma^2_{ab}$
AC	$\sigma^2 + n\sigma^2_{\underline{a}b\underline{c}} + nb\sigma^2_{ac}$	$= \sigma^2 + nb\sigma^2_{ac}$
BC	$\sigma^2 + n\sigma^2_{a\underline{b}\underline{c}} + na\sigma^2_{bc}$	$= \sigma^2 + n\sigma^2_{abc} + naK^2_{bc}$
ABC	$\sigma^2 + n\sigma^2_{abc}$	$= \sigma^2 + n\sigma^2_{abc}$
Error	σ^2	$= \sigma^2$

* **อักษรห้อยท้ายที่ขีดเส้นใต้คืออักษรที่ตัดออก ถ้าอักษรที่เหลือเป็นตัวห้อยท้ายของปัจจัยคงที่ก็ให้ตัดวาเรียนซ์ย่อยนั้นออกดังในบรรทัด A**

จากตาราง 13.7.5 เห็นได้ว่าการทดสอบผลของแฟกตอเรียลต่าง ๆ ต้องดัดแปลงตามความเหมาะสม คือมีทั้งสามารถใช้ MSE หารโดยตรงและวิธีอื่น ๆ

13.8 รายละเอียดเพิ่มเติมเกี่ยวกับแฟกตอเรียล

1. เมื่อไม่มีปฏิกิริยาระหว่างแฟกเตอร์

ในบางครั้ง การทดลองที่มี 2 แฟกเตอร์อาจไม่มีปฏิกิริยาระหว่างแฟกเตอร์ หรือมีก็น้อยมาก ซึ่งสามารถพบเห็นได้เสมอ จากแบบจำลองทางคณิตศาสตร์

$$X_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

เมื่อไม่มีปฏิกิริยาก็จะกลายเป็น

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$$

ซึ่งคล้ายกับการทดลองแบบ RCB ตัวอย่างเช่นข้อมูลในตาราง 13.3.2 ถ้าไม่มีปฏิกิริยาระหว่าง AB คือระหว่างปุ๋ยกับปูนขาว ก็สามารถวิเคราะห์เหมือนกับการทดลองแบบ RCB ได้ผลดังตาราง 13.8.1

ตาราง 13.8.1 การวิเคราะห์ข้อมูลในตาราง 13.3.2 เมื่อไม่มีปฏิกิริยาระหว่างแฟกเตอร์

Sources	df	SS	MS
A	3	14.17	4.72
B	2	123.04	61.52**
Error	42	127.42	3.03
Total	47	264.63	

2. การวิเคราะห์ 2^n แฟกตอเรียลโดยวิธีของ Yates

ในปี 1937 Yates ได้พัฒนาวิธีการวิเคราะห์การทดลอง 2^n แฟกตอเรียล โดยวิธีง่าย ๆ คือใช้วิธีการบวก-ลบ เรียกวิธีนี้ว่า Yates' algorithm ซึ่งมีวิธีการดังนี้

(1) เรียงชุดแฟกตอเรียลตามลำดับตัวอักษรและการคูณ เช่น แฟกตอเรียล 2^2, 2^3 และ 2^4 มีลำดับทรีตเมนต์ดังนี้

แฟกตอเรียล	ลำดับที่ 1	2	3	4	5	6	7	8	9	10	....
2^2	(1)	a	b	ab							
2^3	(1)	a	b	ab	c	ac	bc	abc			
2^4	(1)	a	b	ab	c	ac	bc	abc	d	ad	....

(2) จัดระเบียบของข้อมูลตามลำดับในข้อ 1 เช่น ถ้าเป็น 2^3 แฟกตอเรียล จำนวน 3 ซ้ำ อาจแสดงดังตาราง 13.8.2

(3) ดำเนินการบวก-ลบ จำนวน n ครั้ง (n = จำนวนแฟกเตอร์) ในแต่ละครั้งครึ่งบนเกิดจากการบวกเป็นคู่ ๆ ชิดกัน ครึ่งล่างเกิดจากการลบเป็นคู่ ๆ ชิดกัน ในการลบนั้นให้ข้อมูลค่าล่างของแต่ละคู่เป็นตัวตั้ง

(4) เมื่อดำเนินการครบแล้ว นำผลที่ได้ไปหาค่า sum of square โดยหารด้วย $n(\sum c_i^2)$, n = จำนวนซ้ำหรือบลอก, $\sum c_i^2 = 1^2 + 1^2 + \ldots + 1^2 = 8$ หรือ $2^3 = 8$ ตัวอย่างเช่น

$$SS(A) = \frac{(\sum A)^2}{n\sum c_i^2} = \frac{(-2)^2}{24} = 0.17$$

$$SS(B) = \frac{4^2}{24} = 0.67$$

ผลของการคำนวณแสดงไว้ในตาราง 13.8.2 ทั้งนี้ค่าอื่น ๆ คำนวณโดยวิธีปกติดังนี้

$$\text{Correcton factor (CF)} = \frac{60^2}{24} = 150$$

$$\text{Total SS (TSS)} = 2^2 + 2^2 + \ldots + 2^2 - 150 = 12.00$$

$$\text{Block SS (SSB)} = \frac{21^2 + 18^2 + 21^2}{8} - 150 = 0.75$$

$$\text{SSE} = \text{TSS} - \text{SS อื่น ๆ} = 3.40$$

แล้วนำผลการวิเคราะห์วาเรียนซ์ลงตารางต่อไปตามปกติ

ตาราง 13.8.2 ข้อมูล 2^3 แฟกตอเรียลจากการทดลอง 3 ซ้ำ

ทรีตเมนต์	ซ้ำที่			รวม	(1)	(2)	(3)	อิทธิพล (ผลรวมทั้งหมด)	SS
	1	2	3						
(1)	2	1	2	5	12	25	60		
a	2	2	3	7	13	35	−2	A	0.17
b	3	2	3	8	16	−1	4	B	0.67
ab	2	2	1	5	19	1	6	AB	1.50
c	3	3	2	8	2	1	10	C	4.17
ac	3	2	3	8	−3	3	0	AC	0.00
bc	3	3	4	10	0	−5	2	BC	0.17
abc	3	3	3	9	−1	−1	4	ABC	0.67
รวม	21	18	21	60					

อธิบายขั้นตอน (1)

12 = 5 + 7	2 = 7 − 5
13 = 8 + 5	−3 = 5 − 8
16 = 8 + 8	0 = 8 − 8
19 = 10 + 9	−1 = 9 − 10

อธิบายขั้นตอน (2)

25 = 12 + 13	1 = 13 − 12
35 = 16 + 19	3 = 19 − 16
−1 = 2 + (−3)	−5 = −3 − 2
−1 = 6 + (−1)	−1 = −1 − 6

3. การทดลอง 1 ซ้ำใน 2^n แฟกตอเรียล

ในกรณีที่ n มีค่าสูง เช่น $2^6 = 64$, $2^8 = 256$ ทำให้การทดลองใหญ่เกินไป เช่น ถ้าเราต้องการทราบผลของปุ๋ยและธาตุอาหาร N, P, K, Mg, Zn, Fe, Ca ในการทดลองครั้งเดียว ก็มีถึง 128 แฟกตอเรียล ถ้าต้องทำ 2-4 ซ้ำ ก็จะลงทุนมาก อาจปรับปรุงโดยทดลองเพียง 1 ซ้ำ ในวิธีนี้เราไม่มีการประมาณความคลาดเคลื่อน แต่ความคลาดเคลื่อนที่ใช้ทดสอบได้จากการรวม mean square ของแฟกตอเรียลในระดับสูง ๆ เช่น abcd, abcde, ฯลฯ เช่น จากการทดลอง 2^3 แฟกตอเรียลดังตาราง 13.8.3 อาจวิเคราะห์ลงตารางดังตาราง 13.8.4

ตาราง 13.8.3 การทดลอง 2^3 แฟกตอเรียล 1 ซ้ำ

แฟกตอเรียล	ผล	(1)	(2)	(3)	SS
(1)	1	4	9	24	
a	3	5	15	0	0
b	3	7	1	2	0.5
ab	2	8	−1	−4	2.0
c	4	2	1	6	4.5
ac	3	−1	1	−2	0.5
bc	4	−1	−3	0	0
abc	4	0	−1	2	0.5

ตาราง 13.8.4 ผลการวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 13.8.3

Source	df	SS		MS
A	1	0		0
B	1	0.5		0.5
C	1	4.5		4.5
AB	1	2.0		
AC	1	0.5	3.00	0.75 (df 4)
BC	1	0		
ABC	1	0.5		

จากตาราง 13.8.3 การวิเคราะห์ปัจจัยหลักอาจใช้วิธีปกติก็ได้ ดังนี้

$$\text{Correction factor (CF)} = 72$$

$$\text{Total SS(TSS)} = 1^2 + 3^2 + \ldots + 4^2 - 72 = 8$$

$$SS(A) = \frac{(2A-G)^2}{8} = \frac{[2(3+2+3+4)-24]^2}{8} = 0.0$$

$$SS(B) = \frac{(2B-G)^2}{8} = \frac{[2(3+2+4+4)-24]^2}{8} = 0.5$$

$$SS(C) = \frac{(2C-G)}{8} = \frac{[2(4+3+4+4)-24]^2}{8} = 4.5$$

$$SSE = TSS - SS(A) - SS(B) - SS(C)$$

$$= 8.0 - 0 - 0.5 - 4.5 = 3.00$$

ทั้งนี้ 8 ที่ใช้เป็นตัวหารคือ $\sum c_i^2$ หรือ $2^3 = 8$, A คือผลบวกของแฟกตอเรียลที่มีแฟกเตอร์ A คือ a, ab, ac และ abc จากตาราง 13.8.4 เห็นได้ว่าเราอาจใช้ MS ที่เกิดจาก AB, AC, BC และ ABC เป็นค่าที่ใช้ทดสอบผลของอิทธิพล A, B, C อย่างไรก็ดี ถึงแม้เราไม่ทดสอบก็ทราบได้ว่าแฟกเตอร์ C สำคัญกว่า A และ B

13.9 3^n แฟกตอเรียล

ในการทดลองที่มี n แฟกเตอร์ แฟกเตอร์ละ 3 ระดับ จำนวนแฟกตอเรียลเท่ากับ 3^n เช่น มี 2 แฟกเตอร์ก็ได้ $3^2 = 9$ แฟกตอเรียล เช่น แฟกเตอร์ A และ B อย่าง 3 ระดับ ก็จะได้แฟกตอเรียลดังนี้

แฟกเตอร์ B	2	02	12	22
	1	01	11	21
	0	00	10	20
		0	1	2
			แฟกเตอร์ A	

รูป 13.9.1 การจัดชุดทรีตเมนต์ 3^2 แฟกตอเรียล

คือ ได้แฟกตอเรียล 00, 01, 02, ..., 22 ซึ่งมีสมการคณิตศาสตร์

$$X_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

เมื่อให้

α, β เป็นผลของแฟกเตอร์ A และ B ตามลำดับ

$\alpha\beta$ เป็นปฏิกิริยาระหว่างแฟกเตอร์

$i = 1, 2, 3, \ldots, a$ เมื่อ a จำนวนระดับของแฟกเตอร์ A

$j = 1, 2, 3, \ldots, b$ เมื่อ b จำนวนระดับของแฟกเตอร์ B

ในการจัดชุดแฟกตอเรียลของแฟกเตอร์ที่มี 3 ระดับนี้แตกต่างจากวิธีการจัด 2^n แฟกตอเรียลเล็กน้อย คือไม่นิยมใช้ระดับแฟกเตอร์เป็น a_i และ b_j แต่ใช้ตัวเลข 0, 1, 2 แทน ดังตาราง 13.9.1 คือเลขหลักแรกเป็นระดับของ A หลักที่สองเป็นระดับของ B และถ้าเพิ่มเป็น 3^3 แฟกตอเรียลก็จะมีหลักที่ 3 เป็นระดับของ C ไปเป็นลำดับ ดังนั้นทรีตเมนต์แฟกตอเรียลของ 3^2 แฟกตอเรียลได้แก่ลำดับดังนี้

00, 01, 02, 10, 11, 12, 20, 21, 22

ซึ่งสอดคล้องกับ 2^2 แฟกตอเรียลเดิมที่จัดลำดับเป็น (1), a, b และ ab จากจำนวน 9 ชุด แฟกตอเรียลนี้ทำให้เป็นการทดลองที่มี 9 ทรีตเมนต์ คือ $3^2 = 9$ ดังนั้น ในการทดสอบอิทธิพลหลัก A และ B แต่ละอิทธิพลมี df = 2 ส่วนปฏิกิริยา AB มี df = 4 ถ้าการทดลองมี n ซ้ำ ก็ได้การทดลอง $n3^2 - 1$ df และมี $3^2(n - 1)$ df สำหรับ error

ตัวอย่าง การทดลองเปรียบเทียบผลของปุ๋ยฟอสฟอรัส (P) 3 ระดับ คือ 0, 15 และ 30 กก./ไร่ และปูนขาว 3 ระดับ คือ 0, 100 และ 200 กก./ไร่ ต่อการให้ผลผลิตของถั่วเหลือง โดยทำการทดลอง 3 ซ้ำดังตาราง 13.9.1 ซึ่งแทนระดับปุ๋ยและปูนขาวด้วย 0,1 และ 2 ตามลำดับ

ตาราง 13.9.1 ผลผลิตของพืชได้รับปุ๋ยฟอสฟอรัสและปูนขาว (สมมุติ)

ปุ๋ย	ปูนขาว	ซ้ำ 1	ซ้ำ 2	ซ้ำ 3	รวม
0	0	2	1	1	4
	1	2	2	1	5
	2	4	4	4	12
1	0	2	2	1	5
	1	5	4	6	15
	2	4	5	6	15
2	0	3	3	2	8
	1	2	3	3	8
	2	2	1	2	5
					77

ก่อนดำเนินการวิเคราะห์วาเรียนซ์ ควรจัดทำตารางปฏิกิริยาปุ๋ย x ปูนขาวดังตาราง 13.9.2

ตาราง 13.9.2 ตารางปฏิกิริยาปุ๋ย x ปูนขาว

		ปุ๋ย (P) 0	1	2	รวม
ปูนขาว	0	4	5	8	17
(L)	1	5	15	8	28
	2	12	15	5	32
รวม		21	35	21	

ผลรวมของปุ๋ย $P_0 = 4 + 5 + 12 = 21, P_1 = 35, P_2 = 21$; ผลของปูนขาว $L_0 = 17, L_1 = 28, L_2 = 32$ ดำเนินการวิเคราะห์วาเรียนซ์ดังนี้

$$\text{Correction factor (CF)} = \frac{77^2}{27} = 219.59$$

$$\text{Total SS(TSS)} = 2^2 + 1^2 + \ldots + 2^2 - 219.59 = 59.41$$

$$SS(P) = \frac{(21^2 + 35^2 + 21^2)}{9} - 219.59 = 14.52$$

$$SS(L) = \frac{(17^2 + 28^2 + 21^2)}{9} - 219.59 = 13.41$$

$$SS(PL) = \frac{(4^2 + 5^2 + \ldots + 5^2)}{3} - 219.59 - 14.52 - 13.41 = 23.48$$

$$SSE = TSS - SS(A) - SS(B) - SS(AB)$$

$$= 59.41 - 14.52 - 13.41 - 23.48 = 8.00$$

ตาราง 13.9.3 ผลการวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 13.9.1

Sources	df	SS	MS	F
P	2	14.52	7.26	16.50**
L	2	13.41	6.71	15.25**
PL	4	23.48	5.87	13.34**
Error	18	8.00	0.44	
Total	26	59.41		

ในการทดสอบปุ๋ยและปูนขาว จะเห็นได้ว่าเราได้เพิ่มอัตราปุ๋ยในอัตราเท่ากัน จาก 0 เป็น 15 และ 30 กก./ไร่ ในทำนองเดียวกันการใส่ปูนขาวก็ใส่ 0, 100 และ 200 กก./ไร่ ซึ่งเป็นลำดับที่ชัดเจน เมื่อพบว่าปุ๋ยและปูนขาวทำให้ผลผลิตแตกต่างกัน การเพิ่มหรือลดของผลผลิตเป็นแบบเส้นตรงหรือไม่ หรือเป็นแบบเส้นโค้ง หรือแบบอื่น ๆ เราอาจวิเคราะห์เพื่อศึกษาผล โดยใช้การวิเคราะห์ออร์โธโก-นัลโพลีโนเมียล[7] ซึ่งใช้ได้ในกรณีที่ระดับของแฟกเตอร์เพิ่มในระดับที่เท่า ๆ กัน เป็นขั้น ๆ จากต่ำไปหาสูง การเพิ่มเช่นนี้เราอาจคาดหวังได้ว่า ผลสนองตอบจะเพิ่มหรือลดแบบ Linear, Quadratic, Cubic, หรือแบบอื่น ๆ ซึ่งมีค่าสัมประสิทธิ์แสดงไว้ในตารางภาคผนวก 11 เช่น ถ้าเราต้องการจะแยกว่าผลของฟอสฟอรัสเป็นแบบ Linear (P_L) หรือแบบ Quadratic (P_Q) ก็อาจนำสัมประสิทธิ์ของโพลีโนเมียลดังตาราง 13.9.4 ไปคูณกับผลรวมของทรีตเมนต์แล้วคำนวณ sum of square ดังสมการ (13-14) เป็นตัวอย่าง โดยให้ n, ℓ และ p เป็นจำนวนซ้ำ ระดับของปูนขาวและปุ๋ยตามลำดับ

ตาราง 13.9.4 สัมประสิทธิ์ของ orthogonal polynomial

ปุ๋ย P	ผลผลิต	Orthogonal Contrast	
		Linear	Quadratic
0	21	−1	+1
15	35	0	−2
30	21	+1	+1

$$\text{SS (อิทธิพลหลัก)} = \frac{(\sum c_i T_i)^2}{(\sum c_i^2)(\ell n)} \qquad \ldots(13\text{-}14)$$

ซึ่งตัวหารและเปลี่ยนแปลงตามชนิดของอิทธิพลที่คำนวณ ซึ่งอาจเป็น $n\ell$ หรือ np ก็ได้

$$SS(P_L) = \frac{[(-1)(21)+(0)(35)+(1)(21)]^2}{(\sum c_i^2)(\ell n)} = \frac{0^2}{(2)(9)} = 0.00$$

$$SS(P_Q) = \frac{[(1)(21)+(-2)(35)+(1)(21)]^2}{(\sum c_i^2)(\ell n)} = \frac{(-28)^2}{(6)(9)} = 14.52$$

ในทำนองเดียวกันอาจคำนวณได้ว่า $SS(L_L) = 12.50$ และ $SS(L_Q) = 0.91$

ซึ่งเห็นได้ว่าเราสามารถแยก SS เนื่องจากปุ๋ยและปูนขาวออกได้เป็นส่วนย่อย ๆ คือ Linear และ Quadratic เท่านั้น แต่ละส่วนมี 1 df ไม่มี df มากพอให้แยกได้มากกว่านั้น

ต่อไปก็อาจแยกปฏิกิริยาระหว่างปุ๋ยและปูนขาวออกเป็น 4 ส่วนย่อย ๆ ตามจำนวน df คือ P_LL_L, P_LL_Q, P_QL_L และ P_QL_Q โดยการหาผลคูณของสัมประสิทธิ์ของแต่ละแฟกเตอร์ คือ หา $P_L \times L_L$, $P_L \times L_Q$, $P_Q \times L_L$ และ $P_Q \times L_Q$ ดังนี้

		P_L			P_Q		
		−1	0	+1	+1	−2	+1
L_L	−1	+1	0	−1	−1	+2	−1
	0	0	0	0	0	0	0
	+1	−1	0	+1	+1	−2	+1
L_Q	+1	−1	0	+1	+1	−2	+1
	−2	+2	0	−2	−2	+4	−2
	+1	−1	0	+1	+1	−2	+1

แล้วจัดลงตารางดังตาราง 13.9.5 และดำเนินการคำนวณต่อไป

ตาราง 13.9.4 สัมประสิทธิ์ของปฏิกิริยาระหว่างแฟกเตอร์ใน 3^2 แฟกตอเรียล

แฟกเตอร์	ชุดแฟกตอเรียล									
	00	01	02	10	11	12	20	21	22	
ผลรวม	4	5	12	5	15	15	8	8	5	
$P_L L_L$	+1	0	−1	0	0	0	−1	0	+1	4
$P_L L_Q$	−1	+2	−1	0	0	0	+1	−2	+1	12
$P_Q L_L$	−1	0	+1	+2	0	−2	−1	0	+1	12
$P_Q L_Q$	+1	−2	+1	−2	+4	−2	+1	−2	+1	36

เมื่อหา sum of squares โดยใช้สมการ $SS(AB) = \dfrac{(\sum AB_{ij})^2}{n(\sum c_i^2)}$ ก็ได้

$$SS(P_L L_L) = \frac{1}{3(4)}[1(4)+0(5)-1(12)+0(5)+0(15)+0(15)-1(8)+0(8)+1(5)]^2$$

$$= \frac{1}{3(4)}(-11)^2 = 10.08$$

$$SS(P_L L_Q) = \frac{1}{3(12)}[-1(4)+2(5)-1(12)+0(5)+0(15)+0(15)+1(8)-2(8)+1(5)]^2$$

$$= \frac{1}{3(12)}(-9)^2 = 2.25$$

$$SS(P_Q L_L) = \frac{1}{3(12)}[-1(4)+0(5)+1(12)+2(5)+0(15)-2(15)-1(8)+0(8)+1(5)]^2$$

$$= \frac{1}{3(12)}(-15)^2 = 6.25$$

$$SS(P_Q L_Q) = \frac{1}{3(36)}[1(4)-2(5)+1(12)-2(5)+4(15)-2(15)+1(8)-2(8)+1(5)]^2$$

$$= \frac{1}{3(36)}(23)^2 = 4.90$$

ผลการวิเคราะห์ปรากฏดังตาราง 13.9.5 การทดสอบค่า MSTr ต่าง ๆ ในตารางนี้ ใช้ MSE เป็นตัวหาร ซึ่งจะเห็นได้ว่าปฏิกิริยาระหว่างแฟกเตอร์แบบสำคัญ คือแบบ $P_L L_L$ รองลงมา คือแบบ $P_Q L_L$ จากตารางนี้ สามารถสรุปได้ว่า ผลของ P เป็นแบบโค้ง ผลของปูนขาวเป็นแบบเส้นตรง ส่วนปฏิกิริยา $P_L L_L$ สำคัญกว่าแบบอื่น ๆ

ตาราง 13.9.5 ผลการวิเคราะห์วาเรียนซ์ของข้อมูลในตาราง 13.9.1

Source	df	SS	MS
P	2	14.52	
P_L	1	0.00	0.00
P_Q	1	14.52	14.52**
L	2	13.41	
L_L	1	12.50	12.50**
L_Q	1	0.91	0.91
PL	4	23.48	
P_LL_L	1	10.08	10.08**
P_LL_Q	1	2.25	2.25
P_QL_L	1	6.25	6.25*
P_QL_Q	1	4.90	4.90
Error	18	8.00	0.44

การใช้วิธี Yates' algorithm เพื่อวิเคราะห์ 3^k แฟกตอเรียล

การวิเคราะห์วาเรียนซ์โดยวิธีของ Yates ดังที่กล่าวมาแล้วในตอน 13.8 อาจนำมาดัดแปลงเพื่อใช้กับ 3^k แฟกตอเรียลก็ได้ ซึ่งอาจแสดงโดยใช้ตัวอย่างใน 13.9.1 ในการวิเคราะห์วิธีนี้เราต้องจัดลำดับแฟกตอเรียลเสียใหม่ดังนี้

00 10 20 01 11 21 02 12 22

แล้วดำเนินการวิเคราะห์ดังนี้

ตาราง 13.9.7

1. เลข 3 ค่าบนเกิดจากผลรวมของ 3 ค่าเป็นลำดับคือ $4 + 5 + 8 = 17$, $5 + 15 + 8 = 28$ และ $12 + 15 + 5 = 32$

2. เลข 3 ค่ากลางหาดังนี้ : แบ่งผลรวมออกเป็น 3 ชุด ๆ ละ 3 ค่า จากบนลงล่าง ในแต่ละชุดให้นำค่าที่ 3 ลบด้วยค่าที่ 1 ดังนั้น $8 - 4 = 4$, $8 - 5 = 3$ และ $5 - 12 = -7$

3. เลข 3 ค่าล่างหาดังนี้ : จากผลรวมแต่ละชุด ให้นำค่าที่ 1 ตั้ง ลบด้วย 2 คูณด้วยค่าที่ 2 และบวกด้วยค่าที่ 3 ดังนั้น $4 - (2)(5) + 8 = 2$, $5 - (2)(15) + 8 = -17$, $12 - (2)(15) + 5 = -13$

ตาราง 13.9.7 แสดงการใช้วิธี Yates' algorithm กับข้อมูลในตาราง 13.9.1

แฟกตอเรียล	ผลรวม	(1)	(2)	ผล	ตัวหาร	SS
00	4	17	77	-	-	-
10	5	28	0	A_L	$2^1.3^1.3 = 18$	0.00
20	8	32	–28	A_Q	$2^1.3^2.3 = 54$	14.52
01	5	4	15	B_L	$2^1.3^1.3 = 18$	12.50
11	15	3	–11	AB_{LL}	$2^2.3^0.3 = 12$	10.08
21	8	–7	–15	AB_{QL}	$2^2.3^1.3 = 36$	6.25
02	12	2	–7	B_Q	$2^1.3^2.3 = 54$	0.91
12	15	–17	–9	AB_{LQ}	$2^2.3^1.3 = 36$	2.25
22	5	–13	23	AB_{QQ}	$2^2.3^2.3 = 108$	4.90

สดมภ์ (2) กระทำเช่นเดียวกับสดมภ์ (1)

สดมภ์ - ผล เป็นสดมภ์ที่เปลี่ยนแฟกตอเรียลให้เป็นผลของทรีตเมนต์ ให้ 1 เป็น linear, ให้ 2 เป็น quadratic เช่น 10 เป็น A_L, 01 เป็น B_L, 20 เป็น A_Q, 02 เป็น B_Q, 22 เป็น AB_{QQ}

ตัวหาร คือ $2^a\ 3^b\ n$ เมื่อ a เป็นจำนวนแฟกเตอร์ในชุดนั้น, b เป็นจำนวนแฟกเตอร์ที่ทดสอบลบด้วยจำนวน linear และ n เป็นจำนวนซ้ำ หรือจำนวนบลอก

สดมภ์ SS ค่า SS ของแต่ละรายการได้จากผลของสดมภ์ที่ 2 ยกกำลังและหารด้วยตัวหารในบรรทัดเดียวกัน เช่น

จากตารางดังกล่าวจึงอาจหาได้ว่า

$$SS(A_Q) = \frac{(-28)^2}{54} = 14.52$$

แล้วนำค่า SS ที่คำนวณให้ลงตาราง 13.9.7 ต่อไป
จากตารางดังกล่าวอาจหาได้ว่า

$$SS(A) = SS(A_L) + SS(A_Q) = 0.00 + 14.52 = 14.52$$

$$SS(B) = SS(B_L) + SS(B_Q) = 12.50 + 0.91 = 13.41$$

$$SS(AB) = SS(AB_{LL}) + SS(AB_{LQ}) + SS(AB_{QL}) + SS(AB_{QQ})$$
$$= 10.08 + 6.25 + 2.25 + 4.90 = 23.48$$

ซึ่งผลการวิเคราะห์ตรงกับค่าในตาราง 13.9.5 ทุกประการ

13.10 แบบฝึกหัด

1. ในการทดลองเปรียบเทียบทรีตเมนต์ที่เกิดจากแฟกเตอร์ต่าง ๆ 2 ชนิด (A, B) โดยใช้แผนการทดลองแบบ CRD ปรากฏผลดังนี้

		Replication			
		1	2	3	4
A_1	B_1	52	51	62	31
	B_2	85	69	59	62
A_2	B_1	57	40	29	48
	B_2	36	54	38	49

(1) จงแสดงแบบจำลองทางคณิตศาสตร์

(2) จงแสดงการวิเคราะห์วาเรียนซ์เพื่อทดสอบปัจจัยหลัก A, B และปฏิกิริยา AB

(3) จงแสดงวิธีการคำนวณ $S_{\bar{x}}$ ของอิทธิพล A, B และ AB

2. จากการทดลองเปรียบเทียบการใช้ปุ๋ย P และ K ในถั่วเหลือง โดยใช้แผนการทดลอง RCB ปรากฏว่าได้ผลผลิต (กก. / ไร่) ดังนี้

ทรีตเมนต์	บลอก				รวม
	I	II	III	IV	
ไม่ใส่ปุ๋ย	183	176	291	254	904
P	356	300	301	271	1,228
K	224	258	244	271	943
P + K	329	283	308	326	1,246
รวม	1,092	1,017	1,144	1,068	4,321

จงวิเคราะห์วาเรียนซ์โดยแยกผลของ P, K และ PK ออกมาให้ชัดเจน และทดสอบโดยใช้ F-test และเปรียบเทียบความแตกต่างของผลของการใส่ P และ K

3. จงวิเคราะห์วาเรียนซ์ข้อมูลในข้อ 1 และ 2 โดยวิธีสัมประสิทธิ์เปรียบเทียบ

4. ในการเปรียบเทียบการใช้สารเคมี 4 ชนิด (C) เพื่อจะเพิ่มความงอกแก่เมล็ดพืชชนิดหนึ่ง จำนวน 2 พันธุ์ (V) เมื่อนำเมล็ดพืชดังกล่าว 100 เมล็ด มาคลุกด้วยสารเคมีแล้วเพาะความงอก จากการนับจำนวนเมล็ดที่งอกได้ผลดังนี้

ทรีตเมนต์	บลอกที่			รวม
	1	2	3	
C_1V_1	−1	1	5	−5
C_1V_2	−4	−4	−7	−15
C_2V_1	−6	−7	−10	−23
C_2V_2	2	0	−2	0
C_3V_1	4	3	2	9
C_3V_2	0	2	2	4
C_4V_1	1	−2	−4	−5
C_4V_2	5	6	5	16
รวม	1	−1	−19	−19

(ในการวิเคราะห์วาเรียนซ์กรณีที่ไม่ใช้เครื่องคิดเลข การนำค่าใดค่าหนึ่งมาลบออก ทำให้ค่าสังเกตน้อยลง ทำให้ง่ายแก่การคำนวณ ส่วนผลการวิเคราะห์วาเรียนซ์ไม่เปลี่ยนแปลงแต่อย่างใด ในกรณีของข้อมูลนี้ เรานำ 25 มาลบออกจากค่าสังเกตของทุกข้อมูลย่อย)

(1) จงวิเคราะห์วาเรียนซ์และทดสอบสมมุติฐานว่าผลของสารเคมี พันธุ์ และปฏิกิริยาไม่แตกต่างกัน

(2) จงคำนวณค่า LSD เพื่อเปรียบเทียบระหว่างสารเคมี พันธุ์พืช และทรีตเมนต์ต่าง ๆ

5. ในการทดลองที่มี 3 แฟกเตอร์ที่มีอิทธิพลของ A, B, และ C จงแสดงค่าคาดหมายมีนสแควร์ (EMS) ของอิทธิพลดังนี้

(1) C เมื่อทุกอิทธิพลเป็นปัจจัยสุ่ม

(2) AB เมื่อ A, B เป็นปัจจัยสุ่ม และ C เป็นปัจจัยคงที่

(3) AB เมื่อ A, B เป็นปัจจัยคงที่

(4) ถ้าเพิ่มแฟกเตอร์ D อีก 1 แฟกเตอร์จงแสดง EMS ของ A เมื่อ A, B, C เป็นปัจจัยสุ่ม แต่ D เป็นปัจจัยคงที่

6. ในการทดสอบผลของธาตุ N, P และ K โดยการใส่ และไม่ใส่ธาตุเหล่านี้ต่อผลผลิตของถั่วเหลือง ทำการทดลอง 3 ซ้ำ ได้ผลผลิต (โดยลบด้วย 220) ดังนี้

N (A)	P (B)	K (C)	ชุดทรีตเมนต์	1	2	3
–	–	–	(1)	2	11	5
+	–	–	a	12	23	9
–	+	–	b	15	14	30
+	+	–	ab	25	27	26
–	–	–	c	24	25	18
+	–	+	ac	20	17	16
–	+	+	bc	40	30	34
+	+	+	abc	19	21	27

(1) จงแสดงการวิธีวิเคราะห์วาเรียนซ์ และทดสอบสมมุติฐาน ถ้าทุกแฟกเตอร์เป็นปัจจัยคงที่

(2) จงทดลองผลของ N, P, K โดยแสดง F-test ถ้าแฟกเตอร์เหล่านี้เป็นปัจจัยสุ่ม

7. จงทำวิเคราะห์ข้อมูลในแบบฝึกหัดที่ 6 โดยวิธีสัมประสิทธิ์ของการเปรียบเทียบ
8. จงวิเคราะห์ข้อมูลในแบบฝึกที่ 6 โดยวิธี Yates' algorithm
9. ในการเปรียบเทียบระยะแฟกเตอร์ A และ B อย่างละ 3 ระดับ โดยทำการทดลอง 2 ซ้ำ ได้ผลดังนี้

แฟกเตอร์ B	แฟกเตอร์ A 0	1	2
0	0.8	1.5	2.5
	2.8	3.2	4.2
1	1.0	1.6	1.8
	1.6	1.8	1.0
2	2.0	1.5	2.5
	2.2	0.8	4.0

(1) จงแสดงแบบจำลองทางคณิตศาสตร์

(2) จงแสดงสมมุติฐานถ้าปัจจัย A คงที่ ปัจจัย B เป็นแบบสุ่ม

(3) จงแสดงการวิเคราะห์วาเรียนซ์

10. จากแบบฝึกหัดที่ 9 ซึ่งระดับของแฟกเตอร์ห่างเท่ากัน จงวิเคราะห์เพื่อแยกอิทธิพลหลักและปฏิกิริยาในรูปของ orthogonal polynomial
11. จงวิเคราะห์ข้อมูลในข้อ 9 โดยวิธี Yates' algorithm

คำในบท

(1) factorial (2) factor (3) main effect (4) interaction (5) coefficient, orthogonal contrast

(6) expected mean square (7) orthogonal polynomial

บทที่ 14
คอนฟาวด์แฟกตอเรียล

14.1 คำนำ

ในการทดลองแบบแฟกตอเรียลนั้น ถ้าเราทดสอบครั้งละหลายแฟกเตอร์ ก็ย่อมทำให้มีจำนวนชุดของแฟกตอเรียลมากขึ้น เช่น ถ้ามี 4 แฟกเตอร์ แต่ละแฟกเตอร์มี 2 ระดับ ก็มีแฟกตอเรียล 2^4 = 16 ชุด ถ้าเพิ่มเป็น 6 แฟกเตอร์ก็ได้ 64 ชุด ดังนั้นถ้าเป็นการทดลองทางพืช เช่น เปรียบเทียบสูตรปุ๋ย เมื่อมีชุดทรีตเมนต์มากมายเช่นนี้ แต่ละบลอกก็จะมีขนาดใหญ่มาก ไม่สามารถหาพื้นที่สม่ำเสมอที่จะจัดเป็นบลอกได้อย่างมีประสิทธิภาพ ทำให้ควบคุมความคลาดเคลื่อนได้ยาก อย่างไรก็ดี ปัญหาดังกล่าวนี้อาจแก้ไขได้โดยการแบ่งบลอกใหญ่ออกเป็นบลอกย่อยที่มีขนาดเล็กลง เพื่อให้มีความสม่ำเสมอภายในบลอก ซึ่งทำให้สามารถควบคุมความคลาดเคลื่อนได้ง่ายขึ้น

ในการแบ่งบลอกใหญ่ออกเป็น 2 บลอกย่อยนั้น เราต้องใช้แฟกตอเรียลบางชุดเป็นตัวแบ่งแยก[1] เมื่อใช้เป็นตัวแบ่งแยกเสียแล้ว ผลของแฟกตอเรียลชุดนั้นก็จะปนอยู่กับผลของบลอกย่อย ไม่สามารถแยกมาศึกษาได้ การที่ผลของแฟกตอเรียลนั้นปนอยู่กับผลของบลอกเรียกว่าแฟกตอเรียลนั้นได้รับการคอนฟาวด์[2] ตัวอย่างเช่นในการทดลอง 3 แฟกเตอร์ A, B และ C อย่างละ 2 ระดับ ถ้าใช้แฟกตอเรียล ABC เป็นตัวแบ่งแยกออกเป็น 2 บลอกย่อย ผลของ ABC ก็จะถูกคอนฟาวด์

14.2 คอนฟาวด์ 2^n แฟกตอเรียล

แฟกตอเรียล 2^n คือแฟกตอเรียลที่มี n แฟกเตอร์ แฟกเตอร์ละ 2 ระดับ เช่น มีแฟกเตอร์ A, B, C ก็มีระดับดังนี้

A มีระดับ a_0 , a_1

B มีระดับ b_0 , b_1

C มีระดับ c_0 , c_1

เมื่อนำมาจัดเข้าชุดเป็นแฟกตอเรียล ก็จะได้แฟกตอเรียล 8 ชุดดังนี้

$$a_0b_0c_0,\ a_1b_0c_0,\ a_0b_1c_0,\ a_1b_1c_0,\ a_0b_0c_1,\ a_1b_0c_1,\ a_0b_1c_1,\ a_1b_1c_1$$

ถ้าให้ $a_0b_0c_0$ = (1), $a_1b_0c_0$ = a,... ก็ได้แฟกตอเรียลดังนี้

(1), a, b, ab, c, ac, bc, abc

เมื่อเรานำแฟกตอเรียลเหล่านี้ไปแยกเพื่อกำหนดผลของแฟกเตอร์และปฏิกิริยาระหว่างแฟกเตอร์ ดังปรากฏในตาราง 14.2.1 ก็จะได้กลุ่มบวกและกลุ่มลบ เช่น A, AB, ABC มีกลุ่มบวก-ลบ ดังนี้

แฟกเตอร์	กลุ่มบวก	กลุ่มลบ
A	a, ab, ac, abc	(1), b, c, bc
AB	(1), ab, c, abc	a, b, ac, bc
ABC	a, b, c, abc	(1), ab, ac, bc

ถ้าเรานำกลุ่มบวก-ลบเหล่านี้ไปแยกทดลองคนละบลอก เช่น นำกลุ่มที่เกิดจากแฟกตอเรียล ABC ไปแยกทดลองก็ได้ผลดังนี้

ตาราง 14.2.1 สัมประสิทธิ์ของการทดลอง 2^3 แฟกตอเรียล

แฟกตอเรียล	ทรีตเมนต์							
	(1)	a	b	ab	c	ac	bc	abc
A	−	+	−	+	−	+	−	+
B	−	−	+	+	−	−	+	+
AB	+	−	−	+	+	−	−	+
C	−	−	−	−	+	+	+	+
AC	+	−	+	−	−	+	−	+
BC	+	+	−	−	−	−	+	+
ABC	−	+	+	−	+	−	−	+

กลุ่มบวก	กลุ่มลบ
a	(1)
b	ab
c	ac
abc	bc
บลอก 1	บลอก 2

ในกรณีนี้กล่าวได้ว่า ABC คือตัวแบ่งแยก ผลของ ABC ปนอยู่กับผลของบลอก คือความแตกต่างระหว่างบลอกบวกและบลอกลบ แยกไม่ได้ว่าเกิดจากผลของบลอกหรือผลของ ABC เช่นนี้จึงกล่าวได้ว่าผลของ ABC ถูกคอนฟาวด์ ในทำนองเดียวกันเราอาจใช้แฟกตอเรียลอื่น ๆ เช่น AB, AC หรือ BC หรือแม้แต่ A, B, และ C เป็นตัวแบ่งแยกก็ได้ คือนำชุดบวกและลบมาทดลองคนละบลอกก็ได้เช่นกัน แต่หลักการที่ถูกต้องคือการใช้ปฏิกิริยาระดับสูง ๆ เป็นตัวคอนฟาวด์ เพราะในการศึกษานั้นเราต้องการทราบผลของแฟกเตอร์หลัก ๆ และปฏิกิริยาระดับต่ำ ๆ

การทดลองแบบคอนฟาวด์อาจแยกได้ 2 แบบคือ **แบบสมบูรณ์**[3] **คือใช้แฟกตอเรียลตัวเดียวคอนฟาวด์ทุกซ้ำ** และ**คอนฟาวด์บางส่วน**[4] คือตัวคอนฟาวด์ในซ้ำต่าง ๆ ไม่เหมือนกัน การคอนฟาวด์แบบสมบูรณ์นั้นเราไม่อาจประเมินผลของแฟกตอเรียลที่ใช้คอนฟาวด์ได้เลย เพราะรวมอยู่กับผลของบลอกทั้งหมด ในกรณีของแฟกตอเรียล 2^3 นั้น ถ้าใช้ ABC คอนฟาวด์ก็ไม่อาจแยกผลของ ABC ออกมา แต่ในการคอนฟาวด์บางส่วน ถ้าทดลอง 4 ซ้ำ ใช้ ABC, AB, AC และ BC คอนฟาวด์ในซ้ำที่ 1, 2, 3 และ 4 ตามลำดับ ดังรูป 14.2.1 ทำให้ทราบผลของปฏิกิริยาระหว่างปัจจัยเหล่านี้ คือผลของ ABC, AB, AC และ BC ได้บ้างแต่ไม่เต็ม 100 เปอร์เซ็นต์

ก. คอนฟาวด์แบบสมบูรณ์ จำนวน 4 ซ้ำ (ปฏิกิริยา ABC ถูกคอนฟาวด์)

ซ้ำที่	1		2		3		4	
บลอกที่	1	2	3	4	5	6	7	8
	abc	ab	abc	ab	abc	ab	abc	ab
	a	ac	a	ac	a	ac	a	ac
	b	bc	b	bc	b	bc	b	bc
	c	(1)	c	(1)	c	(1)	c	(1)

ข. คอนฟาวด์บางส่วน จำนวน 4 ซ้ำ มี (ABC, AB, AC และ BC ได้รับการคอนฟาวด์ในซ้ำที่ 1, 2, 3 และ 4 ตามลำดับ

ซ้ำที่	1		2		3		4	
บลอกที่	1	2	3	4	5	6	7	8
	abc	ab	(1)	a	(1)	b	(1)	b
	a	ac	ab	b	ac	c	bc	c
	b	bc	c	ac	b	ab	a	ab
	c	(1)	abc	bc	abc	bc	abc	ac
คอนฟาวด์	ABC		AB		AC		BC	

รูป 14.2.1 แฟกตอเรียล 2^3 ที่มีการคอนฟาวด์โดยใช้ปฏิกิริยาบางตัว

ตาราง 14.2.2 แหล่งของความปรวนแปรในการวิเคราะห์วาเรียนซ์ ของการคอนฟาวด์แบบสมบูรณ์ และคอนฟาวด์บางส่วน

แฟกตอเรียล ไม่คอนฟาวด์		แฟกตอเรียล คอนฟาวด์ ABC		แฟคตอเรียลคอนฟาวด์ AB, AC, BC และ ABC	
Sources	df	Sources	df	Sources	df
Replication	3	Blocks	7	Blocks	7
A	1	A	1	A	1
B	1	B	1	B	1
C	1	C	1	C	1
AB	1	AB	1	AB	$1'$
AC	1	AC	1	AC	$1'$
BC	1	BC	1	BC	$1'$
ABC	1	Error	18	ABC	$1'$
Error	21			Error	17

ในคอนฟาวด์แบบสมบูรณ์นั้น เราไม่อาจแยกผลของแฟกตอเรียลที่คอนฟาวด์ (ในกรณีนี้ได้แก่ ABC) ออกจากผลของบลอก แต่ถ้าคอนฟาวด์เพียงบางส่วนก็จะทำให้ทราบผลของปฏิกิริยาได้บ้าง แต่ไม่เต็ม 100 เปอร์เซ็นต์ เช่น ในกรณีของตัวอย่างนี้เราอาจวิเคราะห์ AB, AC, BC และ ABC จาก 3 ซ้ำ ใน 4 ซ้ำ

14.3 การทำคอนฟาวด์ 2^n แฟกตอเรียลโดยวิธีอื่น ๆ

การจัดนำคอนฟาวด์แฟกตอเรียลโดยใช้บลอกบวกและบลอกลบ ดังที่กล่าวมาแล้วในตอน 14.2 มีข้อจำกัดในกรณีที่มีแฟกเตอร์เป็นจำนวนมาก ทำให้เสียเวลาและเกิดความสับสน อย่างไรก็ดีเราอาจใช้วิธีอื่น ซึ่งทำได้เร็วกว่าและง่ายกว่า

ก. วิธีชุดคู่-คี่

วิธีคู่-คี่ คือการนำตัวแบ่งแยกมาจัดชุดแบบคู่-คี่ให้ครบทุกชุด โดยที่ชุดคู่จะอยู่ในบลอกหนึ่ง และชุดคี่จะอยู่ในอีกบลอกหนึ่ง เช่น ถ้าเป็นการทดลองแฟกเตอร์ A, B, C และ D อย่างละ 2 ระดับ ถ้าให้ ABCD เป็นตัวแบ่งแยกก็จัดเป็นชุดคู่-คี่ ได้ดังนี้

ชุดคู่คือ (1) ab ac ad bc bd cd abcd

ชุดคี่คือ a b c d abc abc acd bcd

ในกรณีนี้พบว่าชุดคู่เป็นชุดบวก ชุดคี่เป็นชุดลบ เรานำชุดคู่และคี่ไปทดลองคนละบลอก แต่ถ้าใช้ ABC เป็นตัวแบ่งแยกก็ได้ผลดังนี้

ชุดคู่คือ (1) ab ac bc d abd acd bcd

ชุดคี่คือ a b c abc ad bd cd abcd

ในกรณีที่ตัวแบ่งแยกเป็นคี่นั้นชุดคู่เป็นชุดลบ ส่วนชุดคี่เป็นชุดบวก ในการเขียนชุดคู่หรือคี่ในตัวอย่างหลังนี้ จะเห็นได้ว่าแฟกตอเรียลตัวที่ห้าในแต่ละชุดเกิดจากการนำเอาแฟกเตอร์ที่ไม่ได้รวมอยู่ในตัวแบ่งแยกเข้ามาคูณ เช่น d เกิดจาก d x (1) , abd เกิดจาก d x ab เป็นต้น

การจัดแบบชุดคู่-คี่นี้อาจเรียกว่าเป็นการจัดแบบชุดประธาน[5] คือแฟกตอเรียลที่มีแฟกเตอร์ a_0b_0 หรือ $a_0b_0c_0$ ซึ่งแทนด้วย (1) เรียกว่าเป็นชุดประธาน ชุดประธานนี้จะอยู่ในชุดแฟกตอเรียลคู่เสมอ

ข. การแยกบลอกโดยวิธีฐานเลขคณิตและสมการเส้นตรง[6]

การใช้วิธีฐานเลขคณิต[7] คือใช้ประโยชน์เลขที่เหลือจากการหารเลขค่าหนึ่งด้วยอีกค่าหนึ่ง เช่น เลขมีค่า I เมื่อหารด้วย m แล้วได้ผลลัพธ์ q และเศษ r ดังนั้น

$$I = qm + r$$

ตัวอย่างเช่น 20 หารด้วย 3 เหลือเศษ 2 อ่านว่า 20 modulo 3 คือ 2 หรือมีความหมายง่าย ๆ ว่า 20 หาร 3 เหลือเศษ 2 ซึ่งเขียนว่า

$$20 \pmod 3 = 2$$

หรือชุดอื่น ๆ เช่น

$$5 \pmod 2 = 1$$
$$4 \pmod 2 = 0$$
$$3 \pmod 5 = 3$$

คือ r ที่เกิดจาก mod 2 = 0, 1, r ที่เกิดจาก mod 3 = 0, 1, 2, ฯลฯ

ตัวตั้งที่ให้เศษเหมือนกันเรียกว่าเป็นเลขในกลุ่มเดียวกัน[8] เช่น

$$18 \pmod 3 = 0$$
$$7 \pmod 3 = 1$$
$$3 \pmod 3 = 0$$
$$4 \pmod 3 = 1$$

ดังนั้น 18 และ 3, 7 และ 4 อยู่ในคนละกลุ่ม เราใช้หลักอันนี้ไปจัดแบ่งบลอกในการทำคอนฟาวด์นั่นเอง

ในการทดลองแบบแฟกตอเรียลนั้น ความสัมพันธ์ระหว่างปัจจัยเป็นแบบเส้นตรงคือ

$$L = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_k x_k \qquad \ldots(14\text{-}1)$$

เมื่อ α_i = คือค่ายกกำลัง ซึ่งเท่ากับชนิดของปฏิกิริยา ซึ่งถ้าแฟกเตอร์มี 2 ระดับ (2^n) $\alpha_i = 1$, x_i คือระดับของแฟกเตอร์คือ 0 และ 1 เช่น a_0 , a_1

ตัวอย่าง ในการทดสอบ 3 แฟกเตอร์คือ A, B และ C อย่างละ 2 ระดับ ได้แฟกตอเรียล (1), a, b, ab, c, ac, bc, abc อาจแยกออกเป็น 2 บลอก โดยใช้ ABC เป็นตัวแบ่งแยก

จากสมการ (14-1) $\alpha_1 = \alpha_2 = \alpha_3 = 1$; $A = x_1$, $B = x_2, C = x_3$ ดังนั้น ได้สมการ

$$L \;=\; (1)x_1 + (1)x_2 + 1(x_3)$$

เมื่อใช้วิธีฐานทางเลขคณิตเพื่อแยกเป็น 2 บลอก ก็กระทำได้ดังนี้

$$\begin{array}{rcl}
(1) & : & L = 1(0)+1(0)+1(0) = 0 \pmod 2 \;\; (\text{คือ } (1) = a_0b_0c_0) \\
a & : & L = 1(1)+1(0)+1(0) = 1 \pmod 2 \;\; (\text{คือ } a = a_1b_0c_0) \\
b & : & L = 1(0)+1(1)+1(0) = 1 \pmod 2 \\
ab & : & L = 1(1)+1(1)+1(0) = 0 \pmod 2 \\
c & : & L = 1(0)+1(0)+1(1) = 1 \pmod 2 \\
ac & : & L = 1(1)+1(0)+1(1) = 0 \pmod 2 \\
bc & : & L = 1(0)+1(1)+1(1) = 0 \pmod 2 \\
abc & : & L = 1(1)+1(1)+1(1) = 1 \pmod 2
\end{array}$$

ต่อไปก็จัดบลอกตามเศษเหลือ คือ L = 0 ไว้บลอกหนึ่ง L = 1 ไว้อีกบลอกหนึ่ง

L = 0	L = 1
(1)	a
ab	b
ac	c
bc	abc
บลอก 1	บลอก 2

อย่างไรก็ดีในการแยกออกเป็น 2 บลอกนั้น เมื่อทราบทรีตเมนต์ใน บลอกแรกแล้ว บลอกที่ 2 อาจหาโดยใช้วิธีง่าย ๆ โดยการนำแฟกตอเรียลเดี่ยว ๆ ตัวใดตัวหนึ่งเข้าไปคูณในบลอกแรก เช่น นำ 6 เข้าไปคูณใน L_0

$$(1).b = b,\;\; ab.b = ab^2 = a\;,\; ac.b = abc,\; bc.b = b^2c = c$$

ก็จะได้บลอก L_1 ทั้งนี้ค่ายกกำลังสอง เช่น b^2 จะให้ค่าเท่ากับ 1

14.4 การวิเคราะห์วาเรียนซ์ในคอนฟาวด์ 2^n แฟกตอเรียล

ก. แฟกตอเรียลที่ไม่ใช้ซ้ำ

เนื่องจากการทดลองแบบแฟกตอเรียลมักเป็นการทดลองขนาดใหญ่ ดังนั้นบางครั้งเราทดลองเพียง 1 ซ้ำ หรือเรียกว่าการทดลองแบบไม่ใช้ซ้ำ เช่น ถ้าทดสอบ 4 แฟกเตอร์คือ A, B, C, D และใช้ ABCD เป็นตัวแบ่งแยกก็ได้ 2 บลอก คือ

บลอก 1 : (1) ab ac ad bc bc cd abcd

บลอก 2 : a b c d abc abd acd bcd

นำบลอก 1 หรือ 2 ไปทดลองเพียง 1 ครั้ง เมื่อนำมาวิเคราะห์ก็ได้ผลดังแสดงในตาราง 14.4.1 ซึ่งพบว่าไม่มี SSE โดยตรง แต่อาจนำปฏิกิริยา 3 แฟกเตอร์มารวมกันเป็น SSE

ตาราง 14.4.1 การวิเคราะห์ 2^4 แฟกตอเรียลที่มี ABCD เป็นตัวแบ่งแยก

Sources		df	
A		1	
B		1	
C		1	
D		1	
AB		1	
AC		1	
AD		1	
BC		1	
BD		1	
CD		1	
ABC	error	1	4 df
ABD		1	
ACD		1	
BCD		1	
Blocks or ABCD		1	

ตัวอย่าง ในการทดลองแฟกตอเรียลที่เกิดจาก 3 แฟกเตอร์ โดยใช้ ABC เป็นตัวแบ่งแยกการทดลองออกเป็น 2 บลอก ดังตาราง 14.4.2

ตาราง 14.4.2 ผลการทดลอง 2^3 แฟกตอเรียล ที่ทดลอง 2 บลอก โดยมี ABC เป็นตัวแบ่งแยก

บลอก 1	บลอก 2
(1) = 7	a = 9
ab = 10	b = 9
ac = 13	c = 12
bc = 12	abc = 12
รวม 42	รวม 42

ซึ่งสามารถวิเคราะห์โดยใช้วิธี Yates' algorithm ดังตาราง 14.4.3 และนำผลลงตารางได้ดังตาราง 14.4.4

ตาราง 14.4.3 การวิเคราะห์ข้อมูลในตาราง 14.4.2

ทรีตเมนต์	ผล	(1)	(2)	(3)	SS
(1)	7	16	35	84	
a	9	19	49	4	2.00
b	9	25	3	2	0.50
ab	10	29	1	– 2	0.50
c	12	2	3	14	24.50
ac	13	1	– 1	– 2	0.50
bc	12	1	– 1	– 4	2.00
abc	12	0	– 1	0	0.00

ตาราง 14.4.4 ผลการวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 14.4.2

Sources	df	SS	MS
A	1	2.00	2.00
B	1	0.500	0.50
C	1	24.50	24.50
Error หรือ AB, AC, BC	3	3.00	1.00
Block หรือ ABC	1	0.00	

ข. การวิเคราะห์วาเรียนซ์เมื่อมีซ้ำและคอนฟาวด์แบบสมบูรณ์

การคอนฟาวด์แบบสมบูรณ์คือพวกที่ใช้ตัวแบ่งแยกเพียง 1 ตัว ในทุก ๆ ซ้ำ ดังนั้นทรีตเมนต์ในแต่ละบลอกประกอบด้วยชุดแฟกตอเรียลที่เหมือนกัน ตัวอย่างวิธีการจัดบลอกในแต่ละซ้ำและผลการวิเคราะห์วาเรียนซ์แสดงไว้ในรูป 14.2.1 และตาราง 14.2.2

ตัวอย่าง การทดลองเพื่อทดสอบแฟกตอเรียลที่เกิดจาก 3 แฟกเตอร์ แฟกเตอร์ละ 2 ระดับ แบ่งแยกออกเป็น 2 บลอก โดยใช้ ABC เป็นตัวแบ่งแยก และทดลอง 4 ซ้ำ ดังตาราง 14.4.5 และ 14.4.6 และสามารถวิเคราะห์ผลได้ดังนี้

$$\text{Correction factor (CF)} = \frac{120^2}{32} = 450.00$$

$$\text{Total SS (TSS)} = 3^2 + 5^2 + \ldots + 4^2 - 450.00 = 24.00$$

$$\text{Block SS(SSB)} = \frac{18^2 + 12^2 + \ldots + 11^2}{4} - 450.00 = 11.50$$

ตาราง 14.4.5 การทดลอง 2^3 แฟกตอเรียลที่ทดลอง ซ้ำละ 2 บลอก โดยมี ABC เป็นตัวแบ่งแยก

ซ้ำที่ 1

บลอกที่ 1

ทรีตเมนต์	b	abc	c	a
ข้อมูล	5	4	4	5
รวม				18

บลอกที่ 2

ทรีตเมนต์	(1)	bc	ac	ab
ข้อมูล	3	3	2	4
รวม				12

ซ้ำที่ 2

บลอกที่ 3

ทรีตเมนต์	(1)	ab	ac	bc
ข้อมูล	4	4	3	4
รวม				15

บลอกที่ 4

ทรีตเมนต์	a	c	b	abc
ข้อมูล	3	4	4	5
รวม				16

ซ้ำที่ 3

บลอกที่ 5

ทรีตเมนต์	ac	(1)	bc	ab
ข้อมูล	2	4	4	4
รวม				14

บลอกที่ 6

ทรีตเมนต์	a	c	b	abc
ข้อมูล	5	5	4	4
รวม				18

ซ้ำที่ 4

บลอกที่ 7

ทรีตเมนต์	abc	b	a	c
ข้อมูล	4	5	4	3
รวม				16

บลอกที่ 8

ทรีตเมนต์	bc	ac	(1)	ab
ข้อมูล	3	3	4	2
รวม				11

ตาราง 14.4.6 การจัดข้อมูลเพื่อการวิเคราะห์วาเรียนซ์จากการทดลองแบบคอนฟาวด์ ABC

ทรีตเมนต์	ซ้ำที่ 1	2	3	4	รวม	เฉลี่ย
(1)	3	4	4	3	14	3.50
a	5	3	5	4	17	4.25
b	5	4	5	5	19	4.27
c	4	4	4	3	15	3.75
ab	4	4	4	2	14	3.50
ac	2	3	2	3	10	2.50
bc	3	4	4	3	14	3.50
abc	4	5	4	4	17	4.50
รวม	30	31	32	27	120	

วิเคราะห์ A, B, C, AB, AC, BC ดูวิธีการหาสัมประสิทธิ์ในตาราง 13.6.1 หรือจำง่าย ๆ ว่าใช้ความแตกต่างระหว่างกลุ่มบวกและกลุ่มลบ ดังนี้

$$\begin{aligned}
SS(A) &= [(a + ab + ac + abc) - (1 + b + c + bc)]^2 / (\textstyle\sum c^2)(n) \\
&= [(17 + 14 + 10 + 17) - (14 + 19 + 15 + 14)]^2 / 32 \\
&= (-4)^2 / 32 = 0.5 \\
SS(B) &= [(b + ab + bc + abc) - (1 + a + c + ac)]^2 / (\textstyle\sum c^2)(n) \\
&= [(19 + 14 + 14 + 17) - (14 + 17 + 15 + 10)]^2 / 32 \\
&= 8^2 / 32 = 2.0 \\
SS(C) &= [(c + ac + bc + abc) - (1 + a + b + ab)]^2 / (\textstyle\sum c^2)(n) \\
&= [(15 + 10 + 14 + 17) - (14 + 17 + 19 + 14)]^2 / 32 \\
&= (-8)^2 / 32 = 2.0 \\
SS(AB) &= [(1 + ab + c + abc) - (a + b + ac + bc)]^2 / (\textstyle\sum c^2)(n) \\
&= [(14 + 14 + 15 + 17) - (17 + 19 + 10 + 14)]^2 / 32 \\
&= 0^2 / 32 = 0.0 \\
SS(AC) &= [(1 + ac + b + abc) - (a + c + ab + bc)]^2 / (\textstyle\sum c^2)(n) \\
&= [(14 + 10 + 19 + 17) - (17 + 15 + 14 + 14)]^2 / 32 \\
&= 0^2 / 32 = 0.0 \\
SSC(BC) &= [(1 + bc + a + abc) - (b + c + ab + ac)]^2 / (\textstyle\sum c^2)(n) \\
&= [(14 + 14 + 17 + 17) - (19 + 15 + 14 + 10)]^2 / 32 \\
&= 4^2 / 32 = 0.5 \\
\text{Error SS} &= \text{Total SS} - \text{SS อื่น ๆ} = 7.5
\end{aligned}$$

จงสังเกตว่ามิได้มีการคำนวณ SS ของ ABC เพราะคอนฟาวด์อยู่กับบลอก แล้วนำลงตาราง วิเคราะห์วาเรียนซ์ ดังตาราง 14.4.7

ตาราง 14.4.7 ผลของการวิเคราะห์วาเรียนซ์จากข้อมูลในตาราง 14.4.6

Sources	df	SS	MS	F
Blocks	7	11.5	1.64	4.15
A	1	0.5	0.5	1.25
B	1	2.0	2.0	5.00
C	1	2.0	2.0	5.00
AB	1	0.0	0.0	0.00
AC	1	0.0	0.0	0.00
BC	1	0.5	0.5	1.25
Error	18	7.5	0.4	
Total	31			

ในกรณีที่มีแนวโน้มที่แสดงว่า ABC ไม่มีนัยสำคัญ (ซึ่งดูได้จากการที่ปฏิกิริยา AB, AC และ BC มีค่าต่ำ) และผลของบลอกมีนัยสำคัญ ควรจะมีการปรับค่าเฉลี่ย ในการปรับค่าเฉลี่ยให้ทำการจัดกลุ่มบลอกตามที่ได้คอนฟาวด์ เช่น ABC เป็นตัวคอนฟาวด์ก็จะจัดดังนี้

บลอก	(1)	ab	ac	bc	a	b	c	abc
ค่าเฉลี่ย	3.50	3.50	2.50	3.50	4.25	4.75	3.75	4.50
ค่าเฉลี่ยทั้งหมด				3.25				4.31
ความแตกต่าง				1.06				
ค่าปรับปรุง				+ 0.53				0.53

เมื่อเรียงค่าเฉลี่ยแล้วหาค่าเฉลี่ยทั้งหมด หาความแตกต่างระหว่างกลุ่มของบลอก (1.06 = 4.31 − 3.25) แล้วหารด้วย 2 เพื่อนำไปลบออกจากบลอกที่สูงกว่า และบวกกับบลอกที่ต่ำกว่า เช่น ทรีตเมนต์ (1) มีค่าเท่ากับ 3.50 + 0.50 = 4.00, ทรีตเมนต์ a = 4.25 − 0.50 = 3.75 ฯลฯ

ค. การวิเคราะห์วาเรียนซ์เมื่อคอนฟาวด์เพียงบางส่วน

จากรูป 14.2.1 ซึ่งแสดงการคอนฟาวด์บางส่วน โดยมี AB, AC, BC และ ABC คอนฟาวด์ในซ้ำที่ 1, 2, 3 และ 4 ตามลำดับ สมมติว่าได้ผลดังตาราง 14.4.8

ตาราง 14.4.8 การทดลองแฟกตอเรียลที่มีตัวแบ่งแยกซ้ำละ 1 ชนิด

	AB(1)		AC(2)		BC(3)		ABC(4)	
	(1) = 3	a = 2	(1) = 3	a = 2	(1) = 3	a = 2	(1) = 3	a = 2
	ab = 2	b = 2	ab = 2	b = 2	ab = 2	b = 2	ab = 2	b = 2
	c = 1	ac = 2	c = 1	ac = 2	c = 1	ac = 2	c = 1	ac = 2
	abc = 1	bc = 3	abc = 1	bc = 3	abc = 1	bc = 3	abc = 1	bc = 3
รวม	7	9	11	12	7	7	12	13
รวม	16		23		14		25	

การวิเคราะห์วาเรียนซ์ ดำเนินใน 2 ขั้นตอนใหญ่ ๆ คือ วิเคราะห์แบบปกติเพื่อหา TSS, SSB, SSE, SS(A), SS(B) และ SS(C) แล้ววิเคราะห์ SS(AB), SS(AC), SS(BC) และ SS(ABC) ต่อไปดังนี้

ชุดแฟกตอเรียล	(1)	a	b	ab	c	ac	bc	abc	รวม
ผลรวม	11	10	12	7	10	10	11	7	78

วิเคราะห์วาเรียนซ์

$$\text{Correction factor (CF)} = \frac{\sum X_{ijkl}^2}{nabc} = \frac{78^2}{32} = 190.125$$

$$\begin{aligned}\text{Total SS(TSS)} &= \sum X_{ijkl}^2 - CF \\ &= 3^2 + 2^2 + \ldots + 3^2 - 190.125 \\ &= 23.875\end{aligned}$$

$$\begin{aligned}\text{Block SS(SSB)} &= \frac{1}{4}[7^2 + 9^2 + \ldots + 13^2] - 190.125 \\ &= 11.375\end{aligned}$$

นำชุดเปรียบเทียบจากตาราง 14.2.1 มาใช้ในการวิเคราะห์ดังนี้

$$\begin{aligned}SS(A) &= \frac{[(a + ab + ac + abc) - ((1) + b + c + bc)]^2}{(\sum c_i^2)(n)} \\ &= \frac{[(10 + 7 + 10 + 7) - (11 + 12 + 10 + 11)]^2}{(8)(4)} \\ &= \frac{(-10)^2}{32} \\ &= 3.120\end{aligned}$$

$$SS(B) = \frac{[(b + ab + bc + abc) - ((1) + a + c + ac)]^2}{(\sum c_i^2)(n)}$$
$$= \frac{[(12 + 7 + 11 + 7) - (11 + 10 + 10 + 10)]^2}{(8)(4)}$$
$$= \frac{(-4)^2}{32}$$
$$= 0.500$$

$$SS(C) = \frac{[(c + ac + bc + abc) - ((1) + a + b + ab)]^2}{(\sum c_i^2)(n)}$$
$$= \frac{[(10 + 10 + 11 + 7) - (11 + 10 + 12 + 7)]^2}{(8)(4)}$$
$$= \frac{2^2}{32}$$
$$= 0.125$$

ในขั้นต่อไปก็วิเคราะห์ SS(AB), SS(AC), SS(BC) และ SS(ABC) โดยลดซ้ำที่แฟกตอเรียลเหล่านี้ได้รับการคอนฟาวด์ตามลำดับ เช่น SS(AB) ให้วิเคราะห์เฉพาะผลรวมจากซ้ำที่ 2, 3 และ 4 ทั้งนี้ผลรวมดังกล่าวแสดงไว้ในตาราง 14.4.9

ตาราง 14.4.9 การลดผลรวมเนื่องจากผลจากการคอนฟาวด์

แฟกตอเรียล	ทุกซ้ำ	ลด AB (1)	ลด AC(2)	ลด BC(3)	ลด ABC (4)
(1)	11	8	8	9	8
a	10	8	7	8	7
b	12	10	9	9	8
ab	7	5	5	6	5
c	10	9	6	8	7
ac	10	8	7	9	6
bc	11	8	8	9	8
abc	7	6	5	6	4
รวม	78	62	55	64	53

$$SS(AB) = \frac{[((1) + ab + c + abc) - (a + b + ac + bc)]^2}{(\sum c_i^2)(n-1)}$$

$$= \frac{[(8+5+9+6)-(8+10+8+8)]^2}{(8)(3)}$$

$$= \frac{(-6)^2}{24}$$

$$= 1.500$$

$$SS(AC) = \frac{[(1) + ac + b + abc) - (a + c + ab + ac)]^2}{(\sum c_i^2)(n-1)}$$

$$= \frac{[(8+7+9+5)-(7+6+5+7)]^2}{(8)(3)}$$

$$= \frac{(-4)^2}{24}$$

$$= 0.667$$

$$SS(BC) = \frac{[((1) + bc + a + abc) - (b + c + ab + ac)]^2}{(\sum c_i^2)(n-1)}$$

$$= \frac{[(9+9+8+6)-(9+8+6-9)]^2}{(8)(3)}$$

$$= \frac{0^2}{24}$$

$$= 0.000$$

$$SS(ABC) = \frac{[(a + b + c + abc) - ((1) + ab + ac + bc)]^2}{(\sum c_i^2)(n-1)}$$

$$= \frac{[(7+8+7+4)-(8+5+6+8)]^2}{(8)(3)}$$

$$= \frac{(-1)^2}{24}$$

$$= 0.042$$

แล้วนำผลการวิเคราะห์วาเรียนซ์ลงตารางดังตาราง 14.4.10 ต่อไป ถ้าจะเราจะแสดงค่าเฉลี่ยของทรีตเมนต์ ก็สมควรที่จะปรับค่าเฉลี่ยเสียก่อน ตามวิธีซึ่งแสดงโดย Gouldon (1952)

ตาราง 14.4.10 ผลการวิเคราะห์วาเรียนซ์ของข้อมูลในตาราง 14.4.8

Sources	df	SS	MS	F
Blocks	7	11.372	1.625	4.221
Treatments	(7)			
A	1	3.120	3.120	8.104
B	1	0.500	0.500	1.297
C	1	0.125	0.125	0.325
AB	1	1.500	1.500	3.896
AC	1	0.662	0.662	1.719
BC	1	0.000	0.000	0.000
ABC	1	0.042	0.042	0.109
Error	17	6.554	0.385	
Total	31	23.875		

14.5 การคอนฟาวด์ 2^n ใน 4 บลอก

เมื่อมีแฟกตอเรียลมากขึ้น การแบ่งจาก 1 ซ้ำเป็น 2 บลอก อาจยังไม่เพียงพอ อาจต้องแบ่งเป็น 4 บลอก โดยเฉพาะอย่างยิ่งเมื่อ $n \geq 4$ ถ้าจำเป็นต้องกระทำเช่นนั้น เราต้องเพิ่มตัวแบ่งแยกอีก 1 ตัว เพื่อช่วยในการแบ่งต่อจาก 2 บลอกเป็น 4 บลอก อย่างไรก็ดี การเพิ่มตัวแบ่งแยกอีก 1 ตัวนั้นจะมีตัวที่ได้รับการคอนฟาวด์ตัวที่ 3 เกิดขึ้นจากผลคูณ เช่นทดสอบแฟกเตอร์ A, B, C, D และ E ถ้าใช้ ABC เป็นตัวแบ่งแยกตัวที่ 1 และ BCE เป็นตัวแบ่งแยกตัวที่ 2 ซึ่ง $ABC \times BCE = AB^2C^2E = AE$ ดังนั้น AE ก็จะได้รับการคอนฟาวด์เพิ่มขึ้นอีก 1 ตัว ถ้าใช้ตัวคอนฟาวด์ ABC และ CDE ก็จะได้ ABDE เป็นตัวคอนฟาวด์ตัวที่สาม ซึ่งโดยปกติแล้วเราควรหลีกเลี่ยงการคอนฟาวด์แฟกเตอร์หลัก A, B, C, D และ E และปฏิกิริยาต่ำ ๆ ที่มี 2 แฟกเตอร์ อย่างไรก็ดีในขั้นนี้เราจะเรียนรู้ถึงวิธีการต่าง ๆ ที่ใช้แบ่ง 1 ซ้ำเป็น 4 บลอก

ตัวอย่าง ในการทดสอบแฟกตอเรียลที่ได้จาก 5 แฟกเตอร์ A, B, C, D และ E จงใช้ ABC และ BCE แบ่งแยกเป็น 4 บลอก

ก. ใช้วิธีการจัดบลอกแบบแฟกตอเรียลคู่-คี่

ในการใช้วิธีนี้ขออธิบายนำเสียก่อน ขอให้สังเกตในตาราง 14.2.1 ว่าแฟกตอเรียลมีการแบ่งได้ 2 แบบ คือแบบคู่ เช่น (1), ab, ac และ bc และแบบคี่ เช่น a, b, c และ abc หรือชุดบวก-ชุดลบ เช่น ถ้า ABC เป็นตัวแบ่งแยกก็ได้ a, b, c และ abc เป็นชุดบวก และ (1), ab, ac และ bc เป็นชุดลบ ถ้า AB เป็นตัวแบ่งแยก (1), ab, c, abc เป็นชุดบวก a, b, ac, bc เป็นชุดลบ ดังนั้นถ้าตัวแบ่งแยกเป็นคี่ แฟกตอเรียลชุดบวกก็เป็นคี่ ถ้าตัวแบ่งแยกเป็นคู่แฟกตอเรียลชุดบวกก็เป็นคู่

การแบ่งแยกจาก 1 ซ้ำ เป็น 4 บลอกใช้วิธี 2 ขั้นตอน คือ ขั้นที่ 1 แบ่งจาก 1 ซ้ำเป็น 2 บลอก ขั้นที่ 2 แยกแฟกตอเรียลร่วมออกมาเป็นหน่วยในบลอกย่อยแต่ละบลอก เช่น ในการทดสอบ 5 แฟกเตอร์ (A, B, C, D, E) อาจแบ่ง 1 ซ้ำ เป็น 4 บลอก โดยใช้ ABC และ BCE เป็นตัวแบ่งแยก

(1) ใช้ ABC เป็นตัวแบ่งแยก

เมื่อใช้ ABC เป็นตัวแบ่งแยกก็ได้บลอก ซึ่งมีสมาชิกเป็นแฟกตอเรียลคู่ (ซึ่งเป็นบลอกลบ) ดังนี้ (แทนว่าเป็นบลอก p)

p : (1) ab ac bc d abd acd bcd e abe ace bce de abde acde bcde

บลอก p อาจอธิบายวิธีการจัดดังนี้ จากตัวแบ่งแยก ABC จัดแฟกตอเรียลคู่ได้ 4 ชุดคือ (1), ab, ac และ bc ต่อจากนั้นก็นำแฟกเตอร์ที่ไม่รวมอยู่ในตัวแบ่งแยกคือ d และ e มาคูณเข้าไปเป็นลำดับ ตัวอย่าง เช่น d x (1) = d, d x ab = abd, ...

เมื่อ ABC เป็นตัวแบ่งแยกก็ได้บลอกที่มีสมาชิกเป็นแฟกตอเรียลคี่ (ซึ่งเป็นบลอกบวก (แทนค่าเป็นบลอก q)

q : a b c abc ad bd cd abcd ae be ce abce ade bde cde abcde

บลอก q เป็นการแยกตัวแบ่งแยกเป็นคี่ คือ a b c และ abc เป็นแฟกตอเรียล 4 ชุดแรก ต่อจากนั้นนำ d และ e เข้าไปคูณเป็นลำดับ

(2) เมื่อใช้ BCE เป็นตัวแบ่งแยกก็ได้บลอก r และ s ดังนี้

r : (1) bc be ce a abc abe ace d bcd bde cde ad abcd abde acde

s : b c e bce ab ac ae abce bd cd de bcde abd acd ade abcde

ในขั้นตอนต่อไปก็หาแฟกตอเรียลร่วมในชุด pr, ps, qr และ qs แฟกตอเรียลร่วมเหล่านี้ก็ได้แก่

pr :	(1)	bc	d	bcd	abe	ace	abde	acde
ps :	e	bce	de	bcde	ab	ac	abd	acd
qr :	a	abc	ad	abcd	be	ce	bde	cde
qs :	ae	abce	ade	abcde	b	c	bd	cd

ในขั้นสุดท้ายนี้อาจใช้วิธีลัด คือ เมื่อหาบลอก pr ได้แล้ว ให้นำแฟกตอเรียลเดี่ยว ๆ ที่ไม่มีในชุด pr ได้แก่ a, b, c และ e เข้าไปคูณบลอก pr ก็จะได้บลอกอื่น ๆ โดยอัตโนมัติ เช่น a x pr = a abc ad abcd be ce bde cde ซึ่งเป็นบลอก qr นั่นเอง เมื่อได้ครบ 4 บลอกแล้วก็นำไปทดลองต่อไป

ข. ใช้วิธีสมการเส้นตรง

ถ้าใช้ ADE และ BCE เป็นตัวแบ่งแยกอาจเขียนสมการดังนี้

$$L_1 = x_1 + x_4 + x_5$$
$$L_2 = x_2 + x_3 + x_5$$

เมื่อแทนระดับของแฟกเตอร์ในสมการ และใช้ modulus 2 ก็ได้ชุดของเศษเหลือตามต้องการ เช่น ถ้าใช้ ADE เป็นตัวแบ่งแยก (L_1)

$$(1)\ :\ L_1 = 1(0) + 1(0) + 1(0) = 0 \pmod 2$$
$$a\ :\ L_1 = 1(1) + 1(0) + 1(0) = 1 \pmod 2$$

คำนวณ $L_1 = 0$ และ $L_1 = 1$ จนครบทุกชุด และกระทำเช่นเดียวกันกับตัวแบ่งแยก BCE จนได้ $L_2 = 0$ และ $L_2 = 1$ ต่อจากนั้นก็จัด 4 บลอก โดยใช้แฟกตอเรียลร่วมของชุด (0,0) , (0,1) , (1, 0) และ (1,1) ได้ผลดังนี้

$L_1 = 0, L_2 = 0$: (1), ad, bc, abcd, abe, ace, cde, bde

$L_1 = 1, L_2 = 0$: a, d, abc, bcd , be, abde, ce, acde

$L_1 = 0, L_2 = 1$: b, abd , c, acd, ae, de, abce, bcde

$L_1 = 1, L_2 = 1$: e, ade, bce, abcde, ab, bd, ac, cd

ชุดของทรีตเมนต์เหล่านี้คือ 4 บลอกของ 1 ซ้ำ

ในการวิเคราะห์วาเรียนซ์นั้น แฟกตอเรียลที่ได้รับการคอนฟาวด์ปนอยู่กับผลของบลอก เช่น ถ้ามีแฟกเตอร์ A, B, C, D และมี ABC และ BCD ได้รับการคอนฟาวด์ ดังนี้มีการคอนฟาวด์ตัวที่สามคือ

$$ABC.BCD = AB^2C^2D = AD$$

คือ แยก df ออกไป 3 ค่า ดังนั้นตารางวิเคราะห์วาเรียนซ์ แสดงได้ดังตาราง 14.5.1

ตาราง 14.5.1 การวิเคราะห์เมื่อคอนฟาวด์ 4 บลอก

Sources	df
A	1
B	1
C	1
D	1
AB	1
AC	1
BC	1
BC	1
CD	1
Error หรือ ABC, ACD, ABCD	3
Blocks หรือ ABC, BCD, AD	3

14.6 การทำคอนฟาวด์ใน 3^n แฟกตอเรียล

ก. การแยกบลอกโดยใช้วิธีออร์โทกันัลลาตินสแควร์[9]

การทดลองที่มีแฟกเตอร์ละ 3 ระดับ หรือ 3^n แฟกตอเรียล ก็สามารถแบ่งการทดลองเป็นบลอกโดยวิธีการทำคอนฟาวด์ได้เช่นเดียวกัน เช่น การทดลองที่มีแฟกเตอร์ A และ B อย่างละ 3 ระดับ เราอาจจัดเป็นตารางได้ดังตาราง 14.6.1

ตาราง 14.6.1 การจัดชุด 3^n แฟกตอเรียล

แฟกเตอร์ B	แฟกเตอร์ A		
	0	1	2
0	00 I_1J_1	10 I_3J_2	20 I_2J_3
1	01 I_2J_2	11 I_1J_3	21 I_3J_1
2	02 I_3J_3	12 I_2J_1	22 I_1J_2

ในการวิเคราะห์วาเรียนซ์พบว่า df ของปฏิกิริยา AB ได้เท่ากับ 4 คือ $(a-1)(b-1) = 4$ จาก df นี้ เราอาจแยกออร์โธโกนัลโพลีโนเมียลอย่าง 1 df ดังกล่าวมาแล้วในตอน 13.9 อย่างไรก็ดีเราอาจแยกปฏิกิริยา AB ที่มี 4 df นี้ออกเป็นองค์ประกอบแบบอื่นก็ได้

ในตาราง 14.6.1 ถ้าเราใส่สัญลักษณ์ I_1, I_2, I_3 และ J_1, J_2 , J_3 แบบออร์โธโกนัลลาติสแควร์ คือแต่ละระดับของ I และ J พบกันเพียง 1 ครั้งเท่านั้น อาจใช้การจัดกลุ่มเช่นนี้เพื่อวิเคราะห์แยกปฏิกิริยา AB ออกเป็น 2 กลุ่ม ๆ ละ 2 df สมมติว่าแฟกตอเรียลต่าง ๆ มีค่าดังตาราง 14.6.2

ตาราง 14.6.2 ข้อมูลจากการทดลองแบบ 3^2 แฟกตอเรียล

แฟกเตอร์ B	แฟกเตอร์ A		
	0	1	2
4	4 I_1J_1	5 I_3J_2	8 I_2J_3
1	5 I_2J_2	15 I_1J_3	8 I_3J_1
2	12 I_3J_3	15 I_2J_1	5 I_1J_2
	$I_1 = 4 + 15 + 5 = 24$	$J_1 = 4 + 15 + 8 = 24$	
	$I_2 = 5 + 15 + 8 = 28$	$J_2 = 5 + 5 + 5 = 15$	
	$I_3 = 12 + 5 + 8 = 25$	$J_3 = 12 + 15 + 8 = 35$	

ในตาราง 14.6.2 นี้เราอาจแยกปฏิกิริยา AB ออกได้เป็น 2 ชนิด คือ ชนิด I เราเรียกว่าเป็นปฏิกิริยา AB^2 และ J เรียกว่าเป็นปฏิกิริยา AB ซึ่งสามารถวิเคราะห์ได้ดังนี้

$$SS(AB^2) \text{ หรือ } I = \frac{\sum I_1^2 + \sum I_2^2 + \sum I_3^2}{3n} - CF$$

$$= \frac{24^2 + 28^2 + 25^2}{9} - \frac{77^2}{27} = 0.97$$

$$SS(AB) \text{ หรือ } J = \frac{\sum J_1^2 + \sum J_2^2 + \sum J_3^2}{3n} - CF$$

$$= \frac{27^2 + 15^2 + 35^2}{9} - \frac{77^2}{27} = 22.51$$

เมื่อนำ $SS(AB^2)$ (I) มารวมกับ SS(AB)(J) = 0.97 + 22.51 = 23.48 ซึ่งเท่ากับ SS(AB) ในตาราง 13.9.2 ทุกประการ จึงสรุปได้ว่า อาจแยก SS(AB) ซึ่งมี df 4 ออกเป็น 2 ส่วนคือ $SS(AB^2)$ และ SS(AB) แต่ละส่วนมี 2 df ส่วนที่แบ่งแยกนี้ไม่ได้มีความหมายอย่างไร อย่างไรก็ดีเราอาจใช้วิธีนี้ทำคอนฟาวด์ คือ การแยกกลุ่ม I และ J เป็นคอนฟาวด์คนละชุด

แบบที่ 1 (AB^2) I

I_1	I_2	I_3
00	01	10
11	12	02
22	20	21

แบบที่ 2 (AB) J

J_1	J_2	J_3
00	01	02
12	10	11
21	22	20

ซึ่งแฟกตอเรียล 3^2 แต่ละซ้ำสามารถแยกได้เป็น 3 บลอก ต่อไปเราจะใช้ชุดใดชุดหนึ่งหรือทั้ง 2 ชุด ในการทดลองก็ได้ หรือถ้ามีหลายซ้ำก็ใช้ทั้ง I และ J อย่างละครึ่ง

ข. การแยกบลอกโดยฐานเลขคณิตและสมการเส้นตรง

การจัดบลอกอาจใช้วิธี modulus ดังที่เคยใช้กับ 2^n แฟกตอเรียลก็ได้ ในกรณี 3^n ที่ใช้ modulus ของ 3 หรือใช้เศษเหลือจากการหารเลขฐาน 10 ด้วย 3 หรือ "r(mod 3)" เช่น 4 = 1(mod 3) คือ 1 เป็นเศษจากการหาร 4 ด้วย 3 และ 1 = 1(mod 3) คือ 1 คือเศษจากการหาร 1 ด้วย 3

จากสมการเส้นตรง (14-1), $L = \alpha_1 x_1 + \alpha_2 x_2$ ในการใช้สมการดังกล่าวนี้ เพื่อจัดทำคอนฟาวด์ 3^2 แฟกตอเรียล ถ้าให้ได้แฟกตอเรียล 2 ชุด เหมือนวิธี I และ J ก็กำหนดค่า α_i ให้ต่างกัน ทั้งนี้พบว่า ถ้า α_1 และ α_2 มีค่า 1 ก็จะได้สมการ $L = x_1 + x_2$ ซึ่งเขียนว่าแฟกตอเรียล AB บลอกที่จัดได้เป็นชุด J ถ้าให้ $\alpha_1 = 1$ และ $\alpha_2 = 2$ ก็ได้สมการ $L = x_1 + 2x_2$ ซึ่งเขียนแฟกตอเรียลเรียกว่า AB^2 และบลอกที่จัดได้เป็นชุด I ดังนั้น

$$(AB)J = AB \; ; \; (AB)I = AB^2 \qquad \ldots(14\text{-}2)$$

ดังนั้นสามารถใช้ modulus 3 จัดบลอกได้ดังนี้

เมื่อปฏิกิริยา AB ก็ใช้สมการ $L = x_1 + x_2$ ดังนี้

$$
\begin{aligned}
00 : L &= 1(0) + 1(0) = 0 \pmod 3\\
01 : L &= 1(0) + 1(1) = 1 \pmod 3\\
02 : L &= 1(0) + 1(2) = 2 \pmod 3\\
10 : L &= 1(1) + 1(0) = 1 \pmod 3\\
11 : L &= 1(1) + 1(1) = 2 \pmod 3\\
12 : L &= 1(1) + 1(2) = 0 \pmod 3\\
20 : L &= 1(2) + 1(0) = 2 \pmod 3\\
21 : L &= 1(2) + 1(1) = 0 \pmod 3\\
22 : L &= 1(2) + 1(2) = 1 \pmod 3
\end{aligned}
$$

แล้วจัดค่า L เดียวกันลงในบลอกเดียวกัน

L_0	L_1	L_2
00	01	11
12	10	02
21	22	20

ซึ่งพบว่าเป็นการจัดในแบบ J ดังนั้นปฏิกิริยา J = AB

ถ้าใช้ AB^2 ก็เขียนสมการเส้นตรง

$$L = x_1 + 2x_2$$

แฟกตอเรียล

$$
\begin{aligned}
00 : L &= 1(0) + 2(0) = 0 \pmod 3\\
01 : L &= 1(0) + 2(1) = 2 \pmod 3\\
02 : L &= 1(0) + 2(2) = 1 \pmod 3\\
10 : L &= 1(1) + 2(0) = 1 \pmod 3\\
11 : L &= 1(1) + 2(1) = 0 \pmod 3\\
12 : L &= 1(1) + 2(2) = 2 \pmod 3\\
20 : L &= 1(2) + 2(0) = 2 \pmod 3\\
21 : L &= 1(2) + 2(1) = 1 \pmod 3\\
22 : L &= 1(2) + 2(2) = 0 \pmod 3
\end{aligned}
$$

ซึ่งจัดบลอกได้ดังนี้

L = 0	L = 1	L = 2
00	00	00
12	12	12
21	21	21

ซึ่งเป็น $(AB)I = AB^2$

ก่อนที่จะจบตอนนี้อธิบายเพิ่มเติมว่า เมื่อใช้แฟกตอเรียลในรูปยกกำลัง A^pB^q กำหนดไว้ว่าค่าแรกให้ยกกำลัง 1 เสมอ ถ้าหากเกิน 1 ใช้วิธียกกำลังสองแล้วหารด้วย 3 หรือใช้ modulus 3 ดังนี้

$$A^2B = (A^2B)^2 = A^4B^2 = AB^2$$

ค. การคอนฟาวด์ 3^3 แฟกตอเรียล

ในกรณีที่มี 3 แฟกเตอร์ แต่ละแฟกเตอร์มี 3 ระดับ ก็จัดได้ $3^3 = 27$ แฟกตอเรียล อาจใช้ปฏิกิริยา ABC, AB^2C, AB^2C^2 เป็นตัวคอนฟาวด์ ทั้งนี้ยกเว้นค่าแรก (A) ที่กำหนดไว้ไม่ให้มีสัมประสิทธิ์อื่นนอกจาก 1 เช่น ถ้าใช้ AB^2C^2 เป็นตัวคอนฟาวด์ก็จะได้สมการ

$$L = 1x_1 + 2x_2 + 2x_3$$

ซึ่งสามารถแสดงการแบ่งแยกโดยใช้ modulus 3 ดังนี้

แฟกตอเรียล

$$000 : L = 1(0) + 2(0) + 2(0) = 0 \pmod 3$$
$$010 : L = 1(0) + 2(1) + 2(0) = 2 \pmod 3$$
$$020 : L = 1(0) + 2(2) + 2(0) = 1 \pmod 3$$
$$100 : L = 1(1) + 2(0) + 2(0) = 1 \pmod 3$$
$$110 : L = 1(1) + 2(1) + 2(0) = 0 \pmod 3$$
$$120 : L = 1(1) + 2(2) + 2(0) = 2 \pmod 3$$

$$\vdots$$

$$202 : L = 1(2) + 2(0) + 2(2) = 0 \pmod 3$$
$$212 : L = 1(2) + 2(1) + 2(2) = 2 \pmod 3$$
$$222 : L = 1(2) + 2(2) + 2(2) = 1 \pmod 3$$

ซึ่งสามารถจัดเป็นบลอกย่อย ๆ ได้ดังนี้

L = 0	L = 1	L = 2
000	020	010
110	100	120
220	210	200
021	011	001
101	121	111
211	201	221
012	002	102
122	112	212
202	222	022

การทดลองแบบ 3^3 แฟกตอเรียล อาจแยกปฏิกิริยา ABC ได้รูปแบบต่าง ๆ ดังต่อไปนี้คือ

$$ABC,\ L = x_1 + x_2 + x_3$$
$$AB^2C,\ L = x_1 + 2x_2 + x_3$$
$$ABC^2,\ L = x_1 + x_2 + 2x_3$$
$$AB^2C^2,\ L = x_1 + 2x_2 + 2x_3$$

ซึ่งนำไปใช้จัดบลอกได้ทั้งสิ้น เมื่อนำไปวิเคราะห์วาเรียนซ์ก็ได้ผลดังตาราง 14.6.3

ตาราง 14.6.3 แหล่งของความปรวนแปรจาก 3^3 แฟกตอเรียลคอนฟาวด์โดย AB^2C

Sources	df
Blocks or AB^2C	2
A	2
B	2
AB	4
C	2
AC	4
BC	4
Error or ABC, ABC^2, AB^2C^2	6
Total	26

จากตาราง 14.4.3 เห็นได้ว่าเราสามารถวิเคราะห์อิทธิพลหลักและปฏิกิริยา 2 แฟกเตอร์ได้ทั้งหมด

ง. การวิเคราะห์วาเรียนซ์คอนฟาวด์ 3^n แฟกตอเรียล

ตัวอย่างที่ 1 จากตัวอย่างในตาราง 14.6.2 ซึ่งเราสามารถวิเคราะห์เพื่อแยก SS(AB) ออกเป็น SS(AB^2) และ SS(AB) โดยใช้การจัดบลอกแบบ I และ J เรายังสามารถวิเคราะห์ SS(AB^2) และ SS(AB) โดยวิธีอื่น ดังตาราง 14.6.4 ในตารางนี้ เราบรรจุข้อมูลลงไป 2 ครั้งในทางขวามือ

ตาราง 14.6.4 จัดข้อมูลในแบบทแยง

	A			A		
B	0	1	2	0	1	2
0	4	5	8	4	5	8
1	5	15	8	5	15	8
2	12	15	5	12	15	15

จากตารางนี้ ทำการบวกข้อมูลในด้านทแยงขวา เริ่มจากข้อมูลแรกคือ 4 แล้วบวกลงไปทางขวา

$$4 + 15 + 5 = 24$$
$$5 + 8 + 12 = 25$$
$$8 + 5 + 15 = 28$$

เมื่อหา SS ก็ได้ผลดังนี้

$$\frac{24^2 + 25^2 + 28^2}{3(3)} - \frac{77^2}{27} = 0.97$$

ถ้าหากบวกในแถวทแยงซ้าย

$$8 + 15 + 12 = 35$$
$$5 + 5 + 5 \quad = 15$$
$$4 + 8 + 15 \quad = 27$$

เมื่อหา SS ก็ได้

$$\frac{35^2 + 15^2 + 27^2}{3(3)} - \frac{77^2}{27} = 22.51$$

ซึ่งเท่ากับ SS(AB^2)I และ SS(AB)J ทุกประการ

ตัวอย่างที่ 2 ในการทดสอบแฟกเตอร์ A และ B อย่างละ 3 ระดับ แบ่งแปลงทดลองออกเป็น 3 บลอก ทำการทดลอง 1 ซ้ำ ได้ผลดังนี้

บลอก 1	บลอก 2	บลอก 3
00 = 6	10 = 3	01 = 5
11 = 9	21 = 6	12 = 4
22 = 7	02 = 7	20 = 8
22	16	17

$$A_0 = 18, A_1 = 16, A_2 = 21$$
$$B_0 = 17, B_1 = 20, B_2 = 18$$

อาจดำเนินการวิเคราะห์ดังนี้

$$\text{Total SS(TSS)} = 6^2 + 9^2 + \ldots + 8^2 - \frac{55^2}{9} = 28.89$$

$$\text{Block SS}(AB^2) = \frac{22^2 + 16^2 + 17^2}{3} - \frac{55^2}{9} = 6.89$$

$$SS(A) = \frac{18^2 + 16^2 + 21^2}{3} - \frac{55^2}{9} = 4.20$$

$$SS(B) = \frac{17^2 + 20^2 + 18^2}{3} - \frac{55^2}{9} = 1.56$$

$$SS(AB) = 16.24$$

ในการทดลองนี้ไม่มีซ้ำ ดังนั้นใช้ SS(AB) เป็น SSE

ตัวอย่างที่ 3 ในการทดลองเปรียบเทียบแฟกเตอร์ A และ B อย่างละ 3 ระดับในรูป 3^2 แฟกตอเรียล โดยใช้การคอนฟาวด์ทั้งแบบ I และแบบ J ดังแสดงในตาราง 14.6.5 และจัดระเบียบเพื่อการวิเคราะห์ในตาราง 14.6.6

ตาราง 14.6.5 การทดสอบ 3^2 แฟกตอเรียลที่คอนฟาวด์ในแบบ I และ J

บลอก		แฟกตอเรียล			รวม
1	(00)	(12)	(21)		
	6	5	6	17	
2	(10)	(01)	(22)		
	4	7	10	21	
3	(02)	(11)	(20)		
	4	2	8	14	R1 = 52
4	(00)	(11)	(22)		
	7	3	10	20	
5	(02)	(10)	21		
	3	5	6	14	
6	(01)	(12)	(20)		
	4	3	7	14	R2 = 48
รวม					100

ตาราง 14.6.6 จัดข้อมูลจากตาราง 14.6.5 ในแบบทแยง

ซ้ำ	B	A_0	A_1	A_2	A_0	A_1	A_2
ซ้ำที่ 1	B_0	6	4	8	6	4	8
(AB) (J)	B_1	7	2	6	7	2	6
	B_2	4	5	10	4	5	10
ซ้ำที่ 2	B_0	7	5	7	7	5	7
(AB^2) (I)	B_1	4	3	6	4	3	6
	B_2	3	3	10	3	3	10

$I_1 = 6 + 2 + 10 = 18,\ I_2 = 8 + 7 + 5 = 20,\ I_3 = 4 + 6 + 4 = 14$

$J_1 = 7 + 6 + 3 = 16,\ J_2 = 5 + 4 + 10 = 19,\ J_3 = 7 + 3 + 3 = 13$

วิเคราะห์วาเรียนซ์ข้อมูลในตาราง 14.6.4

$$\text{Correction factor (CF)} = \frac{\sum X_{ijk}^2}{(3^n)(r)} = \frac{100^2}{18} = 555.56$$

$$\text{Total SS(TSS)} = 6^2 + 5^2 + \ldots + 7^2 - 555.56 = 92.44$$

$$\text{Block SS(SSB)} = \frac{17^2 + 21^2 + \ldots + 14^2}{3} - 555.56 = 17.10$$

$$\text{SS Rep (SSR)} = \frac{52^2 + 48^2}{3^2} - 555.56 = 0.88$$

จากตาราง 14.6.5

$$\text{SS(AB)(J)} = \frac{17^2 + 21^2 + 14^2}{3} - \frac{52^2}{9} = 8.22$$

$$\text{SS(AB}^2\text{)(I)} = \frac{20^2 + 14^2 + 14^2}{3} - \frac{48^2}{9} = 8.00$$

และจัดตาราง AB เพื่อวิเคราะห์อิทธิพล A, B และปฏิกิริยา AB ดังนี้

	A_0	A_1	A_2	
B_0	13	9	15	37
B_1	11	5	12	28
B_2	7	8	20	35
	31	22	47	

$$SS(A) = \frac{31^2 + 22^2 + 47^2}{(2)(3)} - 555.56 = 53.44$$

$$SS(B) = \frac{37^2 + 28^2 + 35^2}{(2)(3)} - 555.56 = 7.44$$

$$SS(AB^2)(\text{rep }1) = \frac{18^2 + 20^2 + 14^2}{3} - \frac{52^2}{9} = 6.23$$

$$SS(AB)(\text{rep }2) = \frac{16^2 + 19^2 + 13^2}{3} - \frac{48^2}{9} = 6.00$$

$$SS(AB) = SS(AB^2)(I) + SS(AB)(J) = 12.23$$

$$\begin{aligned} SSE &= TSS - SSB - SS(A) - SS(B) - SS(AB) \\ &= 92.44 - 17.10 - 53.44 - 7.44 - 12.23 = 2.23 \end{aligned}$$

แล้วนำผลการวิเคราะห์ลงตารางวิเคราะห์วาเรียนซ์ดังตาราง 14.6.7

ตาราง 14.6.6 ผลการวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 14.4.4

Sources	df	SS	MS
Blocks	5	17.10	
Rep	1	0.88	
AB (rep 1)	2	8.22	
AB^2(rep 2)	2	8.00	
A	2	53.44	26.72
B	2	7.44	3.72
AB	4	12.23	3.06
AB^2(rep 1)	2	6.23	
AB (rep 2)	2	6.00	
Error	4	2.33	0.58
Total	17	92.44	

ในการทดลองนี้เราสามารถวิเคราะห์ปฏิกิริยา AB (หรือ J) จากซ้ำที่ 2 และ AB^2(I) จากซ้ำที่ 1 เท่านั้น ส่วน MSE ในการทดสอบ A, B และ AB นั้น คือผลรวมรวมของ A x rep + B x rep อย่างละ 2 df ปฏิกิริยา AB^2 และ AB นั้นมักรวมกันเป็นปฏิกิริยา AB แล้วทดสอบในครั้งเดียว แต่ในการทดลองเพียง 1 ซ้ำ ปฏิกิริยานี้จะถูกคอนฟาวด์และไม่ต้องทดสอบ

14.7 แบบฝึกหัด

1. ถ้าทำการทดสอบแฟกตอเรียลที่เกิดจาก 4 แฟกเตอร์ คือ A, B, C และ D จงแก้ปัญหาดังต่อไปนี้

(1) จงใช้ ABD เป็นตัวแบ่งแยกเป็น 2 บลอก โดยใช้วิธีการต่าง ๆ มา 3 วิธี

(2) ถ้าบลอกหนึ่งมีแฟกตอเรียลเป็น acd a ab c abcd bc d bd อะไรเป็นตัวแบ่งแยก และอีกบลอกหนึ่งมีชุดแฟกตอเรียลอะไรบ้าง

(3) จะหากลุ่มบวกและกลุ่มลบเมื่อใช้ ABCD, ABC, และ BCD เป็นตัวแบ่งแยก

(4) ถ้าให้ ABCD, ABC และ ABD คอนฟาวด์ในซ้ำที่ 1, 2 และ 3 ตามลำดับ จงแสดงกลุ่มของแฟกตอเรียลในแต่ละซ้ำ

2. ในการทดลองเปรียบเทียบผลของปุ๋ย N, P, K โดยให้ NPK คอนฟาวด์ในซ้ำที่ 1 และ NP คอนฟาวด์ในซ้ำที่ 2 ได้ผลการทดลองดังนี้

ซ้ำที่ 1 NPK คอนฟาวด์		ซ้ำที่ 2 NP คอนฟาวด์	
(1) = 1	N = 4	(1) = 3	N = 5
NP = 6	P = 3	NP = 7	P = 4
NK = 6	K = 3	K = 4	NK = 5
PK = 5	NPK = 10	NPK = 9	PK = 5

จงแสดงการวิเคราะห์วาเรียนซ์ และแสดงตารางวิเคราะห์วาเรียนซ์ โดยใช้วิธีปกติและวิธี Yates' algorithm

3. จากการทดลองแบบคอนฟาวด์บางส่วน ได้ผลการทดลองดังนี้

ซ้ำที่	I		II		III		IV	
บลอกที่	1	2	3	4	5	6	7	8
	ab = 24	b = 26	(1) = 28	ab = 27	abc = 30	b = 31	(1) = 32	a = 34
	abc = 30	a = 28	abc = 26	c = 31	a = 32	ab = 34	bc = 30	abc = 33
	(1) = 30	ab = 29	ac = 27	bc = 33	bc = 32	c = 35	ac = 33	b = 36
	c = 31	ac = 25	b = 30	a = 33	(1) = 32	ac = 34	ab = 29	c = 32
รวมบลอก	115	108	111	124	126	134	124	135
รวมทั้งหมด								977

จงทำการวิเคราะห์วาเรียนซ์

4. ในการทดสอบแฟกเตอร์ A, B, C, D และ E อย่างละ 2 ระดับ
   (1) ถ้าเราต้องการที่จะแบ่งการทดลองออกเป็น 4 บลอก เราเลือกตัวแบ่งแยกอย่างไรจึงนับว่าเหมาะสม อธิบาย
   (2) ถ้าใช้ ABC และ CDE เป็นตัวแบ่งแยกออกเป็น 4 บลอก จงแสดงแฟกตอเรียลในแต่ละบลอกจนครบ
5. ในการทดสอบแฟกตอเรียลชนิด 3^2 ถ้าต้องการแบ่งการทดลองออกเป็น 3 บลอก โดยให้ AB^2 เป็นตัวคอนฟาวด์ จงแสดงวิธีการแยกออกเป็นบลอกย่อย ๆ โดยใช้ modulus arithmetic
6. ถ้าทำการทดลอง โดยใช้แผนในข้อ 5 ได้ผลดังนี้

00 = 6	10 = 2	20 = 8
11 = 9	21 = 6	01 = 5
22 = 7	02 = 7	12 = 4

จงแสดงการวิเคราะห์วาเรียนซ์ และตารางวิเคราะห์วาเรียนซ์

คำในบท

(1) defining contrast, (2) confounding, (3) complete confounding, (4) partial confounding, (5) principal block, (6) linear combination, (7) modulus arithmetic, (8) congruent, (9) orthogonal latin square.

บทที่ 15
แฟกตอเรียล-การทดลองเพียงบางส่วน

15.1 คำนำ

ในการทดลองแบบแฟกตอเรียลนั้น เมื่อจำนวนแฟกเตอร์เพิ่มขึ้น จำนวนแฟกตอเรียลที่ต้องทดสอบก็เพิ่มเป็นทวีคูณ เช่น มี 5 แฟกเตอร์ จำนวนแฟกตอเรียลมี 32 ชุด ถ้ามี 6 แฟกเตอร์ จำนวนแฟกตอเรียลเพิ่มเป็น 64 ชุด คือ เพิ่มเป็น 2 เท่า พร้อมกันนั้นระดับของปฏิกิริยาระหว่างแฟกเตอร์ก็เพิ่มขึ้นด้วย ดังตาราง 15.1.1 แม้จะแก้ปัญหาโดยใช้วิธีการทดลองแบบคอนฟาวด์ก็ยังเป็นการทดลองขนาดใหญ่อยู่นั่นเอง ดังนั้นอาจแก้ปัญหาโดยการแบ่งมาทดลองเพียงบางส่วน การทดลองเช่นนี้เรียกว่า การทดลองเพียงบางส่วน[1] จัดเป็นการศึกษาเบื้องต้นสำหรับการทดลองที่มีหลายแฟกเตอร์ ซึ่งต้องการทราบว่าแฟกเตอร์ใดบ้างที่สำคัญ เมื่อพบแล้วก็นำมาศึกษาอย่างละเอียดในแผนการทดลองแบบอื่นต่อไป การทดลองเพียงบางส่วนนับว่ามีการใช้แพร่หลายในการวางแผนการผลิตสินค้าต่าง ๆ โดยเฉพาะอย่างยิ่งเมื่อคาดหมายว่าปฏิกิริยาระดับสูงไม่ค่อยสำคัญ

ตาราง 15.1.1 อิทธิพลหลักและปฏิกิริยาระดับต่าง ๆ ที่เกิดจากแฟกเตอร์ใน 2^n แฟกตอเรียล

n	2^n	อิทธิพล	ปฏิกิริยาระดับ/จำนวน					
		หลัก	1	2	3	4	5	6
5	32	5	10	10	5	1		
6	64	6	15	20	15	6	1	
7	128	7	21	35	35	21	7	1

ในตาราง 15.1.1 นั้นปฏิกิริยาระดับ 1, 2, 3 คือปฏิกิริยาที่มี 2, 3 และ 4 แฟกเตอร์ตามลำดับ

15.2 หลักวิธีการทดลองเพียงบางส่วน

หากเราเปิดไปดูตาราง 13.6.1 หรือ 14.2.1 จะเห็นได้ว่าในแต่ละอิทธิพลหรือปฏิกิริยาจะมีแฟกตอเรียล มีทั้งกลุ่มบวกและกลุ่มลบ เช่น แฟกตอเรียล ABC พบว่ามีทรีตเมนต์อยู่ 2 กลุ่ม คือ กลุ่มบวก (+) และกลุ่มลบ (−) ดังนี้

กลุ่มบวก (+)	a	b	c	abc
กลุ่มลบ (−)	(1)	ab	ac	bc

ซึ่งเรียกว่ากลุ่มนี้เกิดจากการที่ใช้ ABC เป็นตัวแบ่งแยก[(2)] ถ้าเรานำเพียงกลุ่มใดกลุ่มหนึ่งไปทดสอบ อาจได้คำตอบเกี่ยวกับอิทธิพล A, B, C และอื่น ๆ ที่ต้องการแต่คงไม่สมบูรณ์ มีอะไรบ้างที่หายไป และเราสามารถชดเชยหรือแก้ไขได้อย่างไร ถ้าเรานำกลุ่มบวกมาทดลองและจัดแยกกลุ่มบวกลบย่อยลงไปอีก ดังตาราง 15.2.1 ซึ่งเมื่อพิจารณาจากเครื่องหมายบวก-ลบ ในบรรทัดแรกของตาราง 15.2.1 พบว่าผลของ A ได้จาก (a + abc) – (b + c) คือ

ตาราง 15.2.1 แสดงปฏิกิริยาระหว่างแฟกเตอร์ในการทดลองเพียง 1/2 ซ้ำ (กลุ่มบวกของ ABC)

อิทธิพล	a	b	c	abc
A	+	–	–	+
B	–	+	–	+
AB	–	–	+	+
C	–	–	+	+
AC	–	+	–	+
BC	+	–	–	+
ABC	+	+	+	+

$$A = +a - b - c + abc$$

เมื่ออ่านตารางดังกล่าวลงไปยังบรรทัดล่าง ๆ พบว่า ผลของ A เหมือนกับผลของ BC คือ

$$BC = +a - b - c + abc$$

ซึ่งแสดงว่า ในการทดลองแบบครึ่งซ้ำนี้ ผลของแฟกตอเรียลชุดหนึ่งจะปนอยู่กับผลของแฟกตอเรียลอีกชุดหนึ่ง การที่แฟกตอเรียล 2 ชุด มีผลร่วมกันหรือปนกัน เรียกว่า เป็นเอเลียส[(3)] ซึ่งกันและกัน ในทำนองเดียวกันอาจแสดงได้ว่า B เป็นเอเลียสกับ AC และ C เป็นเอเลียสกับ AB ถ้านำกลุ่มลบไปทดลองก็จะได้ผลเหมือนกัน แต่มีเครื่องหมายต่างกันเท่านั้น จึงสรุปได้ดังนี้

กลุ่มบวก	กลุ่มลบ
A = BC	A = – BC
B = AC	B = – AC
C = AB	C = – AB

วิธีการหาเอเลียสที่รวดเร็วสำหรับ 2^n แฟกตอเรียลคือนำแฟกเตอร์มาคูณกับตัวแบ่งแยกแล้วใส่โมดูลัส[(4)]2 คือหาเศษจากการหารด้วย 2 ดังนี้

แฟกเตอร์	ตัวแบ่งแยก	ผลคูณ	เอเลียส
A	ABC	A^2BC	BC
B	ABC	AB^2C	AC
C	ABC	ABC^2	AB

คือกล่าวได้ง่าย ๆ ว่าแฟกเตอร์ใดที่ยกกำลัง 2 ก็จะหายไป เมื่อนำแฟกตอเรียลเพียงครึ่งซ้ำนี้ไปทดลองแล้วนำไปวิเคราะห์วาเรียนซ์ จะได้ผลดังตาราง 15.2.2

ตาราง 15.2.2 ผลการวิเคราะห์วาเรียนซ์ของ 1/2 ซ้ำ มี ABC เป็นตัวแบ่งแยก

Sources	df
A(BC)	1
B(AC)	1
C(AB)	1

ซึ่งเห็นว่าไม่สามารถแยกอิทธิพลหลัก (A, B และ C) ออกจากอิทธิพลของแฟกตอเรียลที่เป็นเอเลียสของมัน

15.3 การจัดทำครึ่งซ้ำจาก 2^n แฟกตอเรียล

การทดลองแบบครึ่งซ้ำเรียกว่าเป็นการทดลองแบบ 1/2 ซ้ำ[5] หรือในกรณีของ 2^n แฟกตอเรียลก็เรียกว่าเป็นการทดลองแบบ 2^{n-1} ก็ได้ ในการทดลองแบบครึ่งซ้ำมีขั้นตอนที่สำคัญ ๆ ดังนี้คือ

1. เลือกตัวแบ่งแยก ในการทดลองแบบครึ่งซ้ำตัวแบ่งแยกมักเลือกชุดแฟกตอเรียลที่สูงสุด เช่น ถ้ามี 3 แฟกเตอร์ ก็ใช้

$$I = ABC$$

I คือชุดแฟกตอเรียลที่ให้ค่าเป็นบวกทั้งหมด ซึ่งปรากฏในคอลัมน์สุดท้ายของตาราง 13.6.1 ถ้ามี 5 แฟกเตอร์ตัวแบ่งแยกก็เป็น

$$I = ABCDE$$

อย่างไรก็ดี ตัวแบ่งแยกอาจเลือกได้ตามความจำเป็น โดยเฉพาะอย่างยิ่งเมื่อจะลดขนาดของการทดลองลงอีก เช่น ลดจาก 1/2 เป็น 1/4 ซ้ำ เมื่อเลือกได้ตัวแบ่งแยกตัวใดแล้ว ก็จะเกิดเอเลียสตามมาเสมอ เช่น

แฟกตอเรียล	ตัวแบ่งแยก	เอเลียส
A	ABCDE	$A^2BCDE = BCDE$
B	ABCDE	$AB^2CDE = ACDE$
AB	ABCDE	$A^2B^2CDE = CDE$

ซึ่งเห็นได้ว่าในกรณีนี้อิทธิพลหลัก (A, B) มีปฏิกิริยา 4 แฟกเตอร์เป็นเอเลียส และปฏิกิริยา 2 แฟกเตอร์มีปฏิกิริยา 3 แฟกเตอร์เป็นเอเลียส

2. **จัดชุดของแฟกตอเรียลเพื่อทดลอง** เมื่อกำหนดตัวแบ่งแยกแล้ว ในขั้นต่อไปก็แบ่งชุดแฟกตอเรียลออกเป็น 2 ชุด โดยใช้หลักของการทำคอนฟาวด์นั่นเอง แล้วเลือกมาอย่างสุ่มเพื่อทำการทดลอง 1 ชุด ในการจัดชุดทดลองนั้น ถ้ามีแฟกเตอร์มาก ๆ ก็มีความยุ่งยาก เพื่อความสะดวกให้แยกวิธีการเป็นขั้นตอนดังนี้

(1) ในแฟกตอเรียลชุดหนึ่ง อาจแบ่งเป็น 2 กลุ่มได้ 2 วิธี

วิธีที่ 1 แยกโดยเครื่องหมายบวก-ลบ ดูตาราง 14.2.1 เช่น ซึ่งทดสอบ 3 แฟกเตอร์ ถ้าใช้ AB เป็นตัวแบ่งแยกก็ได้

กลุ่มบวก : (1) ab c abc

กลุ่มลบ : a b ac bc

การแยกวิธีนี้ เราต้องจัดทำตารางให้เห็นชัดเจน ดังตาราง 14.2.1 ถ้ามีหลายแฟกเตอร์ก็มีความยุ่งยาก ขอให้สังเกตว่าบลอกที่มีแฟกตอเรียล (1) เรียกว่าบลอกประ ธาน[6]

วิธีที่ 2 แยกโดยจัดกลุ่มแฟกเตอร์เป็นคู่และคี่ ทั้งนี้ตั้งต้นจากตัวแบ่งแยกที่กำหนดให้โดยบลอกประธาน จัดเป็นบลอกคู่ ตัวอย่างเช่น

แฟกตอเรียลที่มี 3 แฟกเตอร์ (A, B, C) แต่ละแฟกเตอร์มี 2 ระดับ ถ้าใช้ ABC เป็นตัวแบ่งแยกก็ได้

ชุดคู่คือ : (1), ab, ac, bc
คือนำตัวแบ่งแยก ABC มาจับกันเป็นคู่ ๆ ได้ 3 ชุด รวมกับ (1) ก็ได้ 4 แฟกตอเรียล

ชุดคี่คือ : a, b, c, abc

แฟกตอเรียลที่มี 5 แฟกเตอร์ (A, B, C, D และ E) แฟกเตอร์ละ 2 ระดับ ถ้าให้ ABCDE เป็นตัวแบ่งแยกก็แบ่งได้เป็น 2 ชุด ดังนี้

ชุดคู่คือ : (1), ab, ac, ad, ae, bc, bd, be, cd, ce, de, abcd, abce, abde, acde, bcde คือนำตัวแบ่งแยก มาจับชุดเป็นคู่ ๆ จนครบได้ 16 ชุด

ชุดคี่คือ : a, b, c, d, e, abc, abd, abe, acd, ace, ade, bcd, bce, bde, cde, abcde

จากแฟกตอเรียลชุดเดียวกัน แทนที่จะใช้แฟกตอเรียล ABCDE เป็นตัวแบ่งแยก ถ้าใช้แฟกตอเรียล 3 แฟกเตอร์ เช่น CDE เป็นตัวแบ่งแยกก็ได้ดังนี้

ชุดคู่คือ : (1), cd, ce, de, a, acd, ace, ade, b, bcd, bce, bde, ab, abcd, abce, abde คือเริ่มต้นจากแยกชุดคู่เสียก่อน เมื่อจัดจนครบก็นำเอาแฟกเตอร์ที่ยังขาด คือ ไม่มีในตัวแบ่งแยกไปคูณตลอด เช่น ในกรณีนี้นำ a ไปคูณ a x (1) = a, a x cd = acd, ...

ชุดคี่คือ : c, d, e, cde, ac, ad, ac, acde, bc, bd, be, bcde, abc, abd, abe, abcde
การสร้างตัวชุดคี่ก็ทำนองเดียวกันกับการสร้างชุดคู่

แล้วใช้กลุ่มเลขคู่หรือคี่ หรือกลุ่มบวกหรือลบ ไปทดลองต่อไป อย่างไรก็ดี ในการทดลองนั้นเราต้องศึกษาเอเลียสด้วยว่า ชุดแฟกเตอร์ใดมีชุดใดเป็นเอเลียสบ้าง เช่น ถ้าใช้ ABCDE เป็นตัวแบ่งแยก ถ้าใช้กลุ่มบวก (ชุดคี่) มีเอเลียสดังนี้

A = BCDE	AB = CDE
B = ACDE	AC = ADE
C = ABDE	AD = BCE
D = ADCE	-------------
E = ABCD	DE = ABC

ซึ่งหาโดยวิธีโมดูลัส 2 ดังที่กล่าวมาแล้ว เช่น A x ABCDE = A^2BCDE = BCDE

15.4 แฟกตอเรียล 2^n ทดลอง 1/4 ซ้ำ

ถ้าหากว่าลดขนาดการทดลองลงมาเหลือครึ่งซ้ำ ขนาดของการทดลองยังใหญ่เกินไป ก็สามารถลดลงมาอีกครึ่งหนึ่ง เราเรียกว่า เป็นการทดลองบางส่วนชนิด 1/4 ซ้ำ[7] หรือ 2^{n-2} แฟกตอเรียล ในการลดลงมาอีกครึ่งหนึ่งจากเดิมนี้ จะต้องเลือกตัวแบ่งแยกเพิ่มอีก 1 ตัว และถ้ามีตัวแบ่งแยก 2 ตัว ก็จะเกิดตัวที่สามโดยอัตโนมัติ ซึ่งเกิดจากผลคูณของตัวแบ่งแยก 2 ตัวที่เลือกไว้นั้นเอง ตัวอย่างเช่น ถ้ามีแฟกเตอร์ A, B, C, D และ E เราใช้ ABCD เป็นตัวแบ่งแยกเพื่อแบ่งออกเป็น 2 ชุด ๆ ละครึ่งซ้ำ และใช้ CDE แบ่งแยกย่อยละให้เหลือเพียง 1/4 ซ้ำ ก็ได้ตัวแบ่งเพิ่มอีก 1 ชุด ดังนี้

$$ABCD \times CDE = ABC^2D^2E = ABE$$

คือ ABE เป็นตัวแบ่งแยกตัวที่ 3 แฟกตอเรียล 2^5 เมื่อทำการทดลอง 1/4 ซ้ำ อาจเลือกตัวแบ่งแยกดังนี้

ABCD, BCE (ADE)
ABDE, BCE (ACD)
BCDE, ABC (ADE)

คือพยายามเลือกชุดแฟกตอเรียลที่มีความสำคัญน้อย และเป็นตัวแบ่งแยกที่ให้เอเลียสที่เป็นแฟกตอเรียลที่สำคัญน้อยเช่นกัน

เมื่อได้ตัวแบ่งแยกครบแล้ว ก็ดำเนินการแบ่งกลุ่มแฟกตอเรียลตามต้องการต่อไป เช่น ในการแบ่งกลุ่มแฟกตอเรียล 2^5 ออก 1/4 ซ้ำโดยใช้ ABCD และ BCE เป็นตัวแยก ในขั้นต้นเมื่อใช้ ABCD เป็นตัวแบ่งแยกได้ผลดังนี้

ชุดคู่(ประธาน) : (1), ab, ac, ad, bc, bd, cd, abcd, e, abe ace, ade, bce, bde, cde, abcde
ชุดคี่ : a, b, c, d, abc, abd, acd, bcd, ae, be, ce, de, abce, abde, acde, bcde

หลังจากนั้นก็เลือกมากลุ่มหนึ่งแล้วแบ่งย่อยลงเป็น 2 กลุ่ม โดยใช้ BCE เป็นตัวแบ่งแยก สมมุติว่าใช้กลุ่มบวก ก็นำมาแบ่งโดยใช้วิธีที่เหมาะสม เช่นใช้วิธีใส่เครื่องหมายบวก-ลบ ดังตาราง 15.4.1 ซึ่งหาได้ว่า

กลุ่มบวก คือ ab, ac, bd, cd, e, ade, bce, abcde
กลุ่มลบ คือ (1), ad, bc, abcd, abe, ace, bde, cde

เรานำไปใช้ทดลองต่อไป

ตาราง 15.4.1 การใช้ BCE เป็นตัวแบ่งแยก 2^5 จาก 1/2 เป็น 1/4 ซ้ำ

	(1)	ab	ac	ad	bc	bd	cd	abcd	e	abe	ace	ade	bce	bde	cde	abcde
B	−	+	−	−	+	+	−	+	−	+	−	−	+	+	−	+
C	−	−	+	−	+	−	+	+	−	−	+	−	+	−	+	+
E	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+
BCE	−	+	+	−	−	+	+	−	+	−	−	+	+	−	−	+

การแบ่งแยกออกจาก 1/2 เป็น 1/4 ซ้ำอาจใช้วิธีชุดคู่-คี่ก็ได้ ถ้าใช้ ABCD และ BCE เป็นตัวแบ่งแยกก็เริ่มต้นจากการแบ่งเป็น 1/2 ซ้ำ โดยใช้ BCE เป็นตัวแบ่งแยกได้ผลดังนี้

ชุดคู่ : (1), bc, be, ce, a, abc, abe, ace, d, bcd, bde, cde, ad, abcd, abde, acde

ชุดคี่ : b, c, e, bce, ab, ac, ae, abce, bd, cd, de, bcde, abd, acd, ade, abcde

ต่อจากนั้นก็นำกลุ่มคู่-คี่ ซึ่งได้จากตัวแบ่งแยก ABCD และ BCE มาจับคู่กัน แล้วแยกเฉพาะชุดแฟกเตอร์ที่เหมือนกันออกมา เช่น (ABCD) คู่กับ (BCE) คี่จะได้ชุด

ab, ac, bd, cd, e, ade, bce, abcde

ซึ่งเหมือนกับชุดที่หาได้จากตาราง 15.4.1 ทุกประการ เมื่อจับคู่อื่น ๆ ก็จะได้ชุดทดสอบขนาด 1/4 ซ้ำอีก 3 ชุด ซึ่งนำชุดใดชุดหนึ่งก็ได้มาทำการทดลองต่อไป

15.5 การวิเคราะห์ผลการทดลอง

เนื่องจากเป็นการทดลองแบบครึ่งซ้ำ หรือ 1/4 ซ้ำ ดังนั้น ต้องดัดแปลงวิธีการวิเคราะห์ข้อมูลตามความเหมาะสม ซึ่งใช้ 2 วิธีดังนี้

1. วิธีบวก-ลบของ Yates วิธีนี้ได้อธิบายมาแล้วในตอน 13.8.2 ในการวิเคราะห์ต้องจัดข้อมูลให้เข้าลำดับที่กำหนด

3 แฟกเตอร์ มีลำดับ 1-8 ดังนี้

(1) a b ab c ac bc abc ...(15-1)

4 แฟกเตอร์ มีลำดับ 1-16 ดังนี้

(1) a b ab c ac bc abc d ad bd abd cd acd bcd abcd

ถ้าเป็น 1/2 ซ้ำของ 2^4 แฟกตอเรียลที่มี ABCD เป็นตัวแบ่งแยกแต่ละกลุ่มพยายามจัดให้มีลำดับ

กลุ่มบวก (1) ab ac ad bc bd cd abcd

จากกลุ่มนี้อาจจัดเพื่อวิเคราะห์โดยวิธีของ Yates ก็จัดให้เหมือนแฟกตอเรียลชุด (15-1) ดังนี้

ลำดับเดิม (1) ab ac ad bc bd cd abcd

ลำดับใหม่ (1) a(d) b(d) ab c(d) ac ad abc(d)

ซึ่งลำดับนี้เหมือนกับชุด (15-1) ทุกประการ หลักในการจัดนี้ให้สังเกตว่าในการทดลองนี้เราแยก d ออกมาจากชุดแฟกตอเรียลโดยใส่ไว้ในวงเล็บ ทั้งนี้เพราะไม่อาจทดสอบแฟกเตอร์ที่ 4 คือ D เมื่อการทดลองเป็นเพียง 1/2 ซ้ำ ก็ทดสอบได้อย่างสูง 8 ชุด เมื่อแยก d ออกมา ก็จัดลำดับใหม่ ส่วนแฟกเตอร์ D ประมาณได้จากผลของเอเลียสของ D คือ D x ABCD = ABC ซึ่งปรากฏอยู่ในการทดลองแล้ว

จาก 2^4 แฟกตอเรียล ที่มี ABCD เป็นตัวแบ่งแยกกลุ่มลบได้แก่

a b c d abc abd acd bcd

เมื่อจัดลำดับใหม่

(1)d a b ab(d) c ac(d) bc(d) abc

ซึ่งได้ลำดับเหมือนสมการ (15-1) เช่นกัน

ในกรณีที่เป็นการทดลองแบบ 1/4 ซ้ำ ก็จัดลำดับของแฟกตอเรียลอย่างเดียวกันจึงสามารถวิเคราะห์ได้ เช่น จากชุดแฟกตอเรียลของการทดลอง 2^5 ที่มี ABCD และ BCE เป็นตัวแบ่งแยกสามารถจัดระเบียบได้ดังนี้

ลำดับเดิมจากกลุ่มบวก	ลำดับใหม่
ab	1(e)
ac	a(de)
bd	b(d)
cd	ab
e	c(d)
ade	ac
bce	bc(e)
abcde	abc(de)

เมื่อเรียงลำดับได้ดังนี้ แล้วก็สามารถวิเคราะห์โดยใช้วิธีของ Yastes ได้ต่อไป

ตัวอย่าง ในการเปรียบเทียบผลของปุ๋ย N, P, K, ปูนขาว L และธาตุอาหาร Mg ต่อขนาดเมล็ดของถั่วเขียว โดยการทดลองเพียง 1/4 ซ้ำ มี NPKL, PKMg และ NLMg เป็นตัวแบ่งแยกทำการทดลอง 3 บลอก ได้ผลดังแสดงในตาราง 15.5.1

จากตารางดังกล่าว ได้มีการเรียงลำดับของแฟกตอเรียล เพื่อให้สามารถวิเคราะห์ตามวิธีของ Yastes อย่างไรก็ดีการวิเคราะห์ส่วนอื่น ๆ ยกเว้นผลของแฟกตอเรียลใช้วิธีปกติอยู่นั่นเอง

ตาราง 15.5.1 ขนาดเมล็ดของถั่วเขียวที่ได้รับธาตุอาหารต่าง ๆ กัน (กรัม/100 เมล็ด)

แฟกตอเรียล	บลอก I	บลอก II	บลอก III	รวม	(1)	(2)	(3)	SS
1(Mg)	5.0	5.4	5.4	15.8	31.4	70.5	144.9	
N(LMg)	5.2	5.0	5.4	15.6	39.1	74.4	–2.7	0.30
P(L)	6.6	6.5	6.8	19.9	34.7	– 0.9	12.7	6.73
NP	6.2	6.5	6.5	19.2	39.7	– 1.8	1.1	0.05
K(L)	6.0	6.4	5.8	18.2	– 0.2	7.7	3.9	0.63
NK	5.8	5.7	5.5	16.5	– 0.7	5.0	– 0.9	0.03
PK(Mg)	6.4	6.7	6.4	19.9	– 1.7	– 0.5	– 2.7	0.30
NPK(LMg)	6.4	6.6	6.8	19.8	– 0.1	1.6	2.1	0.18
	48.0	48.3	48.6	144.9				8.23

วิเคราะห์วาเรียนซ์ข้อมูลในตาราง 15.5.1

$$\text{Correction factor (CF)} = \frac{(144.9)^2}{24} = 874.83$$

$$\text{Total SS(TSS)} = 5.0^2 + 5.2^2 + \ldots + 6.8^2 - 874.83 = 9.06$$

$$\text{Block SS(SSB)} = \frac{(48.0^2 + 48.3^2 + 48.6^2)}{8} - 874.83 = 0.03$$

$$\text{SSTr} = SS(N) + SS(P) + SS(NP) + SS(K) + SS(NK) + SS(PK) + SS(NPK)$$

จากตาราง 15.5.1

$$= 8.23$$

$$\text{SSE} = \text{TSS} - \text{SSB} - \text{SSTr}$$

$$= 9.06 - 0.03 - 8.23 = 0.80$$

แล้วนำผลลงตารางวิเคราะห์วาเรียนซ์ดังตาราง 15.5.2 ต่อไป ในคอลัมน์สุดท้ายของตาราง 15.5.1 นั้น แสดงถึงผลของแฟกตอเรียลต่าง ๆ ที่อาจรวมกันอยู่ วิธีการหาแฟกตอเรียลต่าง ๆ ดังกล่าวให้คูณอิทธิ-พลที่วิเคราะห์กับตัวแบ่งแยกทั้ง 3 ชุด เช่น N x NPKL = PKL, N x PKMg = NPKMg, N x NLMg = LMg ซึ่งแสดงไว้ในตาราง 15.5.2 ว่าอิทธิพลเอเลียสเหล่านี้ ปนอยู่กับอิทธิพล N

15.6 การทดลองบางส่วนใน 3^n แฟกเตอร์

จากการทำคอนฟาวด์ 3^2 และ 3^3 แฟกตอเรียล ในตอน 14.6 โดยใช้หลัก orthogonal latin square หรือวิธีฐานเลขคณิต เราก็อาจนำบลอกใดบลอกหนึ่งมาทำการทดลอง ดังนั้น การทดลองจะเป็นแบบ 1/3 ซ้ำ หรือลดต่อไปเป็น 1/9 ซ้ำ ตัวอย่างเช่นถ้าเราคอนฟาวด์

$$I = AB^2$$

$$L = x_1 + 2x_2$$

ก็ได้ 3 บลอก ดังนี้

L = 0	L = 1	L = 2
00	10	01
11	02	12
22	21	20

ถ้าเรานำบลอกนี้ไปทดลอง 1 บลอก เอเลียสของ A หาได้ดังนี้

$$A(AB^2) = A^2B^2 = A^4B^4 = AB$$

เพราะค่ายกกำลังของ A กำหนดไว้ว่าไม่เกิน 1 ดังนั้นนำ A^2B^2 มายกกำลังเป็น $A^4B^4 = AB \pmod 3$ ส่วนเอเลียสของ B

$$B(AB^2) = AB^3 = A$$

ดังนั้นพบว่า A = AB = B ทั้งนี้ ผลของ A, B, AB แยกจากกันไม่ได้

ถ้าทำการทดลอง 1/3 ซ้ำ จาก 3^3 แฟกตอเรียล ชุดของแฟกตอเรียลแตกย่อย ๆ ออกได้ตามหลักของ orthogonal latin square ดังนี้ : A, B, C, AB, AB^2 , AC, AC^2, BC, BC^2, ABC, AB^2C AB^2C^2 ถ้าเราใช้ ABC^2 เป็นตัวแบ่งแยกจะได้บลอกย่อย ๆ 3 บลอก ดังนี้

L_0 =	000	011	022	101	112	120	210	221	202
L_1 =	100	111	122	201	212	220	010	021	002
L_1 =	200	211	222	001	012	020	110	121	102

วิธีการหาเอเลียส นำชุดของแฟกเตอร์ (X) มาคูณกับตัวแบ่งแยก (I) แล้วหาโมดูลัส 3 ทั้งนี้ X และ I ค่าใดค่าหนึ่งยกกำลังสองดังนี้

$$A = A(ABC^2) = A^2BC^2 = A^4B^2C^4 = AB^2C$$

$$A = A(ABC^2)^2 = A^3B^2C^4 = A^6B^4C^8 = BC^2$$

$$B = B(ABC^2) = AB^2C^2$$
$$B = B(ABC^2)^2 = A^2B^3C^4 = A^4B^6C^8 = AC^2$$
$$C = C(ABC^2) = ABC^3 = AB$$
$$C = C(ABC^2)^2 = A^2B^2C^5 = A^4B^4C^{10} = ABC$$
$$AB^2 = AB^2(ABC^2) = A^2B^3C^2 = A^4B^6C^4 = AC$$
$$AB^2 = AB^2(ABC^2)^2 = A^3B^4C^4 = BC$$

การวิเคราะห์วาเรียนซ์ได้ผลดังตาราง 15.6.1

ตาราง 15.6.1 การวิเคราะห์วาเรียซ์ 1/3 ซ้ำใน 3^2 แฟกตอเรียล

Sources	df
A (หรือ BC^2 หรือ AB^2C)	2
B (หรือ AC^2 หรือ AB^2C^2)	2
C (หรือ AB หรือ ABC)	2
AC (หรือ AB^2 หรือ BC)	2

ตัวอย่าง ในการทดสอบปุ๋ย N, P และ K อย่างละ 3 ระดับคือ ไม่ใส่ปุ๋ย ใส่ปานกลาง และสูง จัดแบบแฟกตอเรียล 3^3 แต่ทดลองเพียง 1/3 ซ้ำ โดยใช้ $L = x_1 + 2x_2 + x_3$ ซึ่งพบว่าซ้ำที่ทดสอบมีชุดแฟกตอเรียลดังนี้

100 210 111 221 012 122 020 001 202

ผลของการทดลองแสดงในตาราง 15.6.2

ตาราง 15.6.2 การทดสอบปุ๋ย 3 ชนิด ชนิดละ 3 ระดับและทดลองเพียง 1 ซ้ำ

		N								
		0			1			2		
P		0	1	2	0	1	2	0	1	2
K	0			336	224				190	
	1	316				288				356
	2		342				410	314		
รวม										2776

ซึ่งสามารถจัดระเบียบได้ดังนี้

001 = 316	100 = 224	202 = 314
012 = 342	111 = 288	210 = 190
020 = 336	122 = 410	221 = 356

วิธีการวิเคราะห์

ผลของอิทธิพลหลัก

$$
\begin{array}{llllll}
N & : & N_0 & = & 316 + 342 + 336 & = 994 \\
 & & N_1 & = & 224 + 288 + 410 & = 922 \\
 & & N_2 & = & 314 + 190 + 356 & = 860 \\
P & : & P_0 & = & 316 + 224 + 314 & = 854 \\
 & & P_1 & = & 342 + 288 + 190 & = 820 \\
 & & P_2 & = & 336 + 410 + 356 & = 1{,}102 \\
K & : & K_0 & = & 336 + 224 + 190 & = 750 \\
 & & K_1 & = & 316 + 288 + 356 & = 960 \\
 & & K_2 & = & 342 + 410 + 314 & = 1{,}066
\end{array}
$$

$$\text{Correction factor (CF)} = \frac{(2{,}776)^2}{9} = 856{,}241.78$$

$$\text{Total SS (TSS)} = 336^2 + 224^2 + \ldots + 314^2 - 856{,}241.78 = 36{,}126.22$$

$$SS(N) = \frac{(994^2 + 922^2 + 860^2)}{3} - 856{,}241.78 = 2{,}998.22$$

$$SS(P) = \frac{(854^2 + 820^2 + 1{,}102^2)}{3} - 856{,}241.78 = 15{,}798.22$$

$$SS(K) = \frac{(750^2 + 960^2 + 1{,}066^2)}{3} - 856{,}241.78 = 17{,}243.55$$

$$\text{PK หรือ Error} = TSS - SS(N) - SS(P) - SS(K)$$

$$= 86.23$$

ตาราง 15.6.3 ผลการวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 15.6.2

Sources	df	SS	MS
N (หรือ NPK^2 หรือ PK^2)	2	2,998.22	1,499.11**
P (หรือ NK หรือ NPK)	2	15,798.22	7,899.11**
K (หรือ NP^2 หรือ NP^2K^2)	2	17,243.55	8,621.76**
NP (หรือ NK^2 หรือ PK) (Error)	2	86.23	43.11
Total	8		

15.7 แบบฝึกหัด

1. ในการทดลองเปรียบเทียบแฟกตอเรียลที่มี 4 แฟกเตอร์ แฟกเตอร์ละ 2 ระดับ
   (1) ถ้าให้ ABC เป็นตัวแบ่งแยก จงแสดงบลอกประธาน บลอกบวก บลอกลบ บลอกคู่ และบลอกคี่ และจงแสดงเอเลียส อิทธิพลหลัก และปฏิกิริยาระหว่างแฟกเตอร์ทุกชุด
   (2) ถ้าให้ ABD และ BCD เป็นตัวแบ่งแยก จะได้ผลอย่างไรบ้าง
2. ในการทดลองเปรียบเทียบแฟกตอเรียลที่มี 5 แฟกเตอร์ แฟกเตอร์ละ 2 ระดับ
   (1) ถ้าใช้ ABC และ CDE เป็นตัวแบ่งแยก เพื่อทำการทดลองแบบ 1/4 ซ้ำ จงแสดงบลอกขนาด 1/4 ซ้ำ ที่เป็นบลอกบวกและบลอกลบ
   (2) จากข้อ (1) จงจัดลำดับของแฟกตอเรียลให้พร้อมที่จะวิเคราะห์โดยวิธี Yates' algorithm
   (3) จากข้อ (1) จงแสดงองค์ประกอบในตารางวิเคราะห์วาเรียนซ์ให้มากที่สุด
   (4) ถ้าใช้ AB, ABC, ABCD เป็นตัวแบ่งแยก ตัวแบ่งแยกตัวใดให้ผลตามแผนกำหนด 5 (resolution 5)

3. จากการเปรียบเทียบเปอร์เซ็นต์โปรตีนในหญ้ากินนี พบว่าขึ้นอยู่กับปัจจัยดังต่อไปนี้ คือ A การตัดในฤดูแล้งหรือฤดูฝน, B คือการใส่ปุ๋ยในโตรเจน หรือไม่ใส่ C = การใส่ปุ๋ยฟอสฟอรัส หรือไม่ใส่ และ D การตัดสูงหรือตัดเตี้ย

   (1) ถ้าเราสามารถวิเคราะห์เปอร์เซ็นต์โปรตีนได้เพียง 8 ตัวอย่าง จงแสดงว่าเราควรวิเคราะห์ชุดใดบ้าง ทั้งนี้ให้แฟกเตอร์เหล่านี้มี 2 ระดับ เช่น A มีระดับ a_0, a_1
   (2) ถ้าแฟกตอเรียลต่อไปนี้มีเปอร์เซ็นต์โปรตีน ดังนี้คือ

(1)	= 11.8	ad	= 13.5
ab	= 13.2	bd	= 15.0
ac	= 13.8	cd	= 14.5
bc	= 14.3	abcd	= 13.5

จงแสดงการวิเคราะห์โดยวิธีใช้ Yates' algorithm และแสดงตารางวิเคราะห์วาเรียนซ์

4. ในการทดลองแบบ 1/4 ซ้ำนั้น เราเลือกใช้ชุดย่อยชุดใดเพื่อการทดลองก็ได้ ถ้าเราทดสอบแฟกตอเรียลที่เกิดจาก 5 แฟกเตอร์คือ A, B, C, D และ E

   (1) ถ้าชุดย่อยชุดหนึ่งเป็น b, e, ac, ad, bcd, cde, abce, abde จงแสดงชุดย่อยอื่น ๆ อีก 3 ชุด
   (2) ถ้านำแต่ละชุดมาวิเคราะห์ โดยใช้วิธี Yates' algorithm จงแสดงลำดับของแฟกตอเรียลที่ถูกต้อง

5. ในการทดลองเปรียบเทียบธาตุปุ๋ยและธาตุอาหาร 6 ชนิด คือ N, P, K, Mg, B, Zn ชนิดละ 2 ระดับ โดยใช้ 32 แฟกตอเรียลและแบ่งการทดลองออกเป็น 8 บลอก ดังนี้

บลอกที่ 1		2		3		4	
NPBZn	40	NP	39	PKBZn	31	KBMgZn	33
NPMgZn	33	PKBMg	33	NPMgZn	42	PMg	33
NK	36	MgZn	34	(1)	31	NPKB	39
BNg	36	NKBZn	36	NKBMg	37	NZn	36
รวม	145		142		141		144

บลอกที่ 5		6		7		8	
KB	30	NKMgZn	42	PB	28	NB	33
NMg	38	NPBMg	46	KL	30	NPKZn	37
NPKBMgZn	44	BZn	32	NBMgZn	38	KMg	31
PZn	32	PK	33	NPKMg	40	PBMgZn	32
รวม	144		153		136		133

ก. จงวิเคราะห์วาเรียนซ์โดยใช้วิธีบวกลบของ Yates และวิธีปกติ

ข. ในการวิเคราะห์วาเรียนซ์ ถ้าพบว่าธาตุใดให้ความแตกต่างทางสถิติ จงหาค่าเฉลี่ยเนื่องจากการใส่หรือไม่ใส่ธาตุชนิดนั้น

คำในบท

(1) fractional experiment, (2) defining contrast, (3) alias (4) modulus (5) half replication (6) principal block (7) one-fourth replication

บทที่ 16
แผนการทดลองแบบมีแปลงย่อย

16.1 คำนำ

ในการทดลองโดยใช้แผนการทดลองแบบ CRD, RCB หรือลาตินสแควร์ดังที่กล่าวมาแล้ว ในแผนการทดลองเหล่านี้ แต่ละแปลงย่อยภายในแต่ละซ้ำหรือแต่ละบลอก เราอาจแบ่งออกเป็นแปลงย่อย ๆ เพื่อบรรจุและศึกษาปัญหาที่สองไปพร้อม ๆ กัน โดยเฉพาะอย่างยิ่งเมื่อปัญหาแรกมักต้องใช้หน่วยทดลองขนาดใหญ่ เช่น ในการทดลองเกี่ยวกับพืช การทดลองเปรียบเทียบวิธีการเตรียมดินอาจทำการเปรียบเทียบระยะปลูกไปพร้อม ๆ กัน สมมุติว่ามีการเตรียมดิน 3 วิธีคือ a_1, a_2, a_3 และใช้ระยะปลูก 4 ระยะคือ b_1, b_2, b_3 และ b_4 การเตรียมดินมักต้องใช้พื้นที่ขนาดใหญ่เพราะต้องใช้เครื่องจักร แต่ระยะปลูกใช้แปลงขนาดเล็กก็ได้ เราจึงใส่วิธีการเตรียมเข้าไปในปัญหาแรก การทดลองเช่นนี้เรียกว่าการทดลองแบบสปลิตพลอต หรือพูดง่าย ๆ ว่าเป็นการทดลองแบบมีแปลงย่อย เป็นการทดลองที่ศึกษาหลายปัญหาร่วมกันเช่นเดียวกับการทดลองแบบแฟกตอเรียลนั่นเอง

การทดลองแบบมีแปลงย่อยนั้นมีหลายชนิดและระดับดังนี้

1. แผนการทดลองแบบสปลิตพลอต (split- plot design) เป็นแผนการทดลองที่มีแปลงย่อยระดับเดียว คือ มีการแบ่งหน่วยทดลองของปัญหาแรกเพื่อบรรจุปัญหาที่สอง
2. แผนการทดลองแบบสปลิต-สปลิตพลอต (split-split plot design) เป็นแผนการทดลองที่มีแปลงย่อย 2 ระดับ คือแบ่งแปลงย่อย (split plot) เพื่อรับปัญหาที่สาม
3. การทดลองแบบสตริปพลอต (strip-plot design) เป็นการทดลองที่นำปัญหา 2 ปัญหามาทดลองในแนวตัดขวางกัน คือในแนวตั้งและแนวนอน

16.2 แผนการทดลองแบบสปลิตพลอต

หลักการ

ในการทดลองแบบสปลิตพลอตนั้น เราแยกปัญหาหรือแฟกเตอร์ที่ศึกษาออกเป็น 2 ชุด คือ ชุด A และ B ชุด A เป็นทรีตเมนต์ที่อยู่ในแปลงใหญ่ หรือหน่วยทดลองใหญ่ ชุด B เป็นทรีตเมนต์ที่อยู่แปลงย่อย ในการเลือกว่าชุดใดควรอยู่แปลงใหญ่หรือแปลงย่อย ควรพิจารณาดังนี้คือ ความสำคัญของปัญหา ปัญหาที่สำคัญกว่าควรจัดอยู่ในแปลงย่อย นอกจากนั้นศึกษาถึงความเหมาะสมว่าปัญหาใดเหมาะจะทดลองในแปลงใหญ่หรือแปลงย่อย โดยถือหลักความสะดวกและความเหมาะสมในการปฏิบัติเป็นสำคัญ ต่อไปนี้ปัญหาที่อยู่ในแปลงใหญ่เรียกว่า เมนพลอต(1) แปลงใหญ่เรียกว่า โวลพลอต(2) ส่วนปัญหาในแปลงย่อยเรียกว่า ซับพลอต(3) ดังนั้นเห็นได้ว่า ถ้าปัญหาต่าง ๆ มีความ

สำคัญเท่ากันเราก็ไม่ควรใช้การทดลองแบบสปลิตพลอต ปัญหา B ซึ่งอยู่ในแปลงย่อยมีซ้ำมากกว่าจะมีความเที่ยงตรงมากกว่า ในทางตรงกันข้าม ปัญหา A ในเมนพลอต มีซ้ำน้อย และมีพื้นที่ขนาดใหญ่ หรือมีโอกาสที่ได้รับหน่วยทดลองที่ไม่เที่ยงตรงได้มาก ผลการทดสอบจึงเที่ยงตรงน้อยกว่า

การจัดแปลงทดลอง

ในแผนการทดลองแบบสปลิตพลอต อาจจัดเมนพลอตแบบแผนการทดลอง CRD, RCB หรือลาตินสแควร์ก็ได้ สมมุติว่าในเมนพลอตมีการเปรียบเทียบแฟกเตอร์ A 3 ระดับคือ a_1, a_2, a_3 ในซับพลอตจัดแฟกเตอร์ B 4 ระดับคือ b_1, b_2, b_3, b_4 อาจจัดการทดลองดังนี้

การทดลองแบบ CRD สมมุติว่าทดลอง 3 ซ้ำ

(1) สุ่มเมนพลอตแต่ละระดับลงใน 3 แปลง ซึ่งมีทั้งสิ้น 3 ซ้ำ x 3 เมนพลอต = 9 แปลง

$a_1^{(1)}$	$a_2^{(2)}$	$a_1^{(3)}$	$a_3^{(4)}$	$a_3^{(5)}$	$a_2^{(6)}$	$a_2^{(7)}$	$a_3^{(8)}$	$a_1^{(9)}$

ซึ่งเห็นได้ว่าแฟกเตอร์ A แต่ละระดับจะได้รับการสุ่มลงหน่วยทดลองใดก็ได้

(2) สุ่มซับพลอตลงในแปลงใหญ่ สุ่มลงครบทุกแปลง จำนวน 9 แปลง

a_1	a_2	a_1	a_3	a_3	a_2	a_2	a_3	a_1
b_4	b_3	b_1	b_3	b_1	b_2	b_4	b_1	b_3
b_2	b_2	b_3	b_1	b_3	b_4	b_1	b_4	b_1
b_1	b_4	b_2	b_4	b_1	b_3	b_3	B_3	b_2
b_3	b_1	b_4	b_2	b_2	b_3	b_2	b_2	b_4

ทดลองแบบ RCB จำนวน 3 บลอก

ในการทดลองแบบ RCB จัดซ้ำเป็นบลอกคือ 3 บลอก แต่ละบลอกมี 3 แปลงเพื่อใส่แฟกเตอร์ A ระดับ a_1, a_2 และ a_3

(1) สุ่มเมนพลอตเช่นเดียวกับแผนทดลอง RCB

บลอก I

a_3	a_1	a_2

บลอก II

a_2	a_3	a_1

บลอก III

a_1	a_2	a_3

(2) สุ่มซับพลอตลงในเมนพลอตครบทุกแปลง

I

a_3	a_1	a_2
b_4	b_2	b_3
b_2	b_1	b_2
b_1	b_4	b_1
b_3	b_3	b_4

II

a_2	a_3	a_1
b_1	B_4	b_4
b_2	B_3	b_1
b_4	b_2	b_2
b_3	b_1	b_3

III

a_1	a_2	a_3
b_2	b_1	b_2
b_1	b_2	b_4
b_3	b_3	b_1
b_4	b_4	b_3

ทดลองแบบลาตินสแควร์

ในการจัดแบบลาตินสแควร์นี้ เมื่อเมนพลอตมี 3 ระดับ ก็มีเมนพลอตทั้งสิ้น $3^2 = 9$ แปลง และสามารถสุ่มเมนพลอตตามวิธีการของแผนการทดลองดังกล่าวดังนี้

$a_1^{(1)}$	$a_2^{(2)}$	$a_3^{(3)}$
$a_2^{(4)}$	$a_3^{(5)}$	$a_1^{(6)}$
$a_3^{(7)}$	$a_1^{(8)}$	$a_2^{(9)}$

ต่อจากนั้นก็สุ่มแฟกเตอร์ B จำนวน 4 ระดับคือ b_1, b_2, b_3 และ b_4 ลงในแต่ละแปลงต่อไป

แหล่งของความแปรปรวนและ df

จากแผนการทดลองทั้ง 3 ชนิด อาจแยกแหล่งของความปรวนแปรและ degrees of freedom ดังแสดงในตาราง 16.2.1

ตาราง 16.2.1 แหล่งของความแปรปรวนและ df ของแผนการทดลองแบบสปลิตพลอต

CRD		**RCB**		**Latin Squares**	
Sources	**df**	**Sources**	**df**	**Sources**	**df**
Whole plots	ar - 1	Whole plots	an - 1	Whole plots	an - 1
		Blocks	(n - 1)	Rows	r - 1
				Columns	c - 1
A	a - 1	A	a - 1	A	a - 1
Error (a) (E_a)	a(r - 1)	Error(a)(E_a)	(a - 1)(n - 1)	Error(a)(E_a)	(a - 1)(a - 2)
Sub- plots	ar(b - 1)	Sub-plots	an(b - 1)	Sub-plots	an(b - 1)
B	b - 1	B	b - 1	B	b - 1
AB	(a - 1)(b - 1)	AB	(a - 1)(b - 1)	AB	(a - 1)(b - 1)
Error (b) (E_b)	a(r - 1)(b - 1)	Error(b)(E_b)	a(n - 1)(b - 1)	Error(a)(E_b)	a(a - 1)(b - 1)
Total	abr - 1	Total	abn - 1	Total	$a^2b - 1$

แบบจำลองทางคณิตศาสตร์ และ EMS

แบบจำลองทางคณิตศาสตร์ของการทดลองแบบสปลิตพลอตที่ใช้แผนการทดลองต่าง ๆ มีความคล้ายคลึงกัน แตกต่างกันเฉพาะในเรื่องการใช้ตัวแปรในรูปของบลอก แถว และคอลัมน์เท่านั้น ถ้าเป็นการทดลองแบบ RCB ก็แสดงแบบจำลองได้ดังนี้

$$X_{ijk} = \mu + B_i + \alpha_j + (B\alpha)_{ij} + \beta_k + (\alpha\beta)_{jk} + (B\beta)_{ik} + (B\alpha\beta)_{ijk}$$

ทั้งนี้

μ = ค่าเฉลี่ยของประชากร

B = อิทธิพลของบลอก

α = ผลของปัญหาในเมนพลอต

β = ผลของปัญหาในซับพลอต

$\alpha\beta$ = ปฏิกิริยาระหว่างปัญหาในเมนพลอตและซับพลอต

ปฏิกิริยาระหว่างบลอกและเมนพลอต ($B\alpha$) จัดเป็นความคลาดเคลื่อนในเมนพลอต (E_a), ปฏิกิริยาระหว่างบลอก-เมนพลอต-ซับพลอต ($B\alpha\beta$) จัดเป็นความคลาดเคลื่อนในซับพลอต (E_b) ส่วนปฏิกิริยาระหว่างบลอกและซับพลอต ถ้ามีก็รวมอยู่ในความคลาดเคลื่อนในซับพลอต ดังนั้นอาจเขียนแบบจำลองใหม่ว่า

$$X_{ij} = \mu + B_i + \alpha_j + \delta_{ij} + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk} \quad \text{...(16-1)}$$

ทั้งนี้กำหนดให้ δ = ความคลาดเคลื่อนในเมนพลอต และ ε เป็นความคลาดเคลื่อนในซับพลอต

กำหนดให้

$i = 1, 2, \ldots, n$ (n = จำนวนบลอก)

$j = 1, 2, \ldots, a$ (a = จำนวนระดับของแฟกเตอร์ A)

$k = 1, 2, \ldots, b$ (b = จำนวนระดับของแฟกเตอร์ B)

โดยมีค่าคาดหมายมีนสแควร์ (EMS) ดังแสดงในตาราง 16.2.2

ตาราง 16.2.2 แสดง **EMS ของการทดลองแบบสปลิตพลอต เมื่อบลอกเป็นปัจจัยสุ่ม**

Sources	Main และ Sub-plot fixed	Main และ Sub-plot random
Blocks	$\sigma_e^2 + ab\sigma_B^2$	$\sigma_e^2 + ab\sigma_B^2$
A	$\sigma_e^2 + b\sigma_{\alpha B}^2 + nbK_a^2$	$\sigma_e^2 + b\sigma_{\alpha B}^2 + nb\sigma_a^2$
Error (a)	$\sigma_e^2 + b\sigma_{\alpha B}^2$	$\sigma_e^2 + b\sigma_{\alpha B}^2$
B	$\sigma_b^2 + naK_b^2$	$\sigma_b^2 + n\sigma_{ab}^2 + an\sigma_b^2$
AB	$\sigma_b^2 + nK_{ab}^2$	$\sigma_b^2 + n\sigma_{ab}^2$
Error (b)	σ_b^2	σ_b^2

หมายเหตุ: **n, a, b = จำนวนบลอก ระดับแฟกเตอร์ A และ B ตามลำดับ**

B, α, β = ผลของบลอก แฟกเตอร์ A และ B ตามลำดับ

$\sigma_b^2 = \sigma^2 + \sigma_{B\alpha\beta}^2$ ซึ่ง σ^2 ประมาณไม่ได้

ตัวอย่าง

ทำการทดลองเปรียบเทียบน้ำหนักต้นแห้งของถั่วเขียว 3 พันธุ์คือ ก, ข และ ค ภายใต้การควบคุมแสงให้มีความเข้มต่าง ๆ กัน 5 อัตรา คือ 100, 90, 80, 70 และ 60 เปอร์เซ็นต์ ทำการทดลอง 4 บลอกได้ผลดังแสดงในตาราง 16.2.3

นำข้อมูลในตารางดังกล่าวไปจัดระเบียบดังตาราง 16.2.4 เพื่อความสะดวกในการวิเคราะห์วาเรียนซ์

วิธีการวิเคราะห์

$$\mathrm{CF} = \frac{(\sum X_{ijk})^2}{nab} = \frac{(569)^2}{(3)(5)(4)} = 5{,}396.02$$

$$\mathrm{TSS} = \sum X_{ijk}^2 - \mathrm{CF}$$

$$= 14^2 + 11^2 + \ldots\ldots + 6^2 - 5{,}396.02 = 148.98$$

ต่อไปก็วิเคราะห์ข้อมูลในเมนพลอตและซับพลอตเป็นลำดับ

ขั้นที่ 1 วิเคราะห์เมนพลอต ให้ดูในตาราง 16.2.3 ดังนี้

$$\mathrm{SS(Main\ plot)} = \frac{\sum X_{ij(k)}^2}{b} - \mathrm{CF} = \frac{\sum RA^2}{b} - \mathrm{CF}$$

$$= \frac{50^2 + 44^2 + \ldots + 44^2}{5} - 5{,}396.02 = 35.38$$

$$\mathrm{Block\ SS(SSB)} = \frac{\sum X_{i(jk)}^2}{ab} - \mathrm{CF} = \frac{\sum R^2}{ab} - \mathrm{CF}$$

$$= \frac{(141^2 + 144^2 + \ldots + 143^2)}{(3)(5)} - 5{,}396.02 = 0.85$$

$$\mathrm{SS(A)} = \frac{\sum X_{j(ik)}^2}{bn} - \mathrm{CF} = \frac{\sum A^2}{bn} - \mathrm{CF}$$

$$= \frac{190^2 + 204^2 + 175^2}{(5)(4)} - 5{,}396.08 = 21.03$$

$$\mathrm{SS\ Error(a)} = \mathrm{SS(Main\ plot)} - \mathrm{SSB} - \mathrm{SS(A)}$$

$$= 35.38 - 0.85 - 21.03$$

$$= 13.50$$

ขั้นที่ 2 วิเคราะห์ซับพลอต โดยใช้ผลในตาราง 16.2.4

$$SS(B) = \frac{\sum X^2_{(ij)k}}{an} - CF = \frac{\sum B^2}{an} - CF$$

$$= \frac{(133^2 + 112^2 + ... + 107^2)}{(3)(4)} - 5{,}396.02 = 41.57$$

$$SS(AB) = \frac{\sum X^2_{(i)jk}}{n} - CF - SS(A) - SS(B)$$

$$= \frac{\sum (AB)^2}{n} - CF - SS(A) - SS(B)$$

$$= \frac{(47^2 + 40^2 + ... + 35^2)}{4} - 5{,}396.02 - 21.03 - 41.57 = 13.63$$

SS(Error b) = TSS – SS อื่น ๆ ทั้งหมด

= TSS – Main plot SS – SS(B) – SS(AB)

= 148.98 – 35.38 – 41.52 – 13.63

= 58.40

ขั้นที่ 3 นำผลลงตารางวิเคราะห์วาเรียนซ์ ดังตาราง 16.3.5 แล้วหาค่า mean square โดยหารค่า sum of squares ด้วย df เช่น

$$\text{MS Blocks} = \frac{\text{Block SS}}{n-1} = \frac{0.85}{3} = 0.28$$

$$\text{MS (Varieties)} = \frac{SS(A)}{a-1} = \frac{21.03}{2} = 10.52$$

ในการทดสอบ F-test ของเมนพลอต ใช้ MSE_a เป็นตัวหาร อย่างไรก็ดี จากการพิจารณาดู EMS ของบลอกในตาราง 16.2.2 เห็นได้ว่าเราไม่อาจนำ MSE_a มาทดสอบบลอก อย่างไรก็ดี การใช้บลอกก็เป็นการช่วยลดความคลาดเคลื่อนจากการทดลอง

การทดสอบซับพลอตใช้ MSE_b เป็นตัวหาร ถ้าปัญหาเป็นแบบคงที่ ก็ใช้ MSE_b ทดสอบทั้ง MS(A) และ MS(AB) ถ้าปัญหาเป็นแบบสุ่มก็ให้ทดสอบเป็นขั้นตอน คือทดสอบ MS(AB) เสียก่อน โดยใช้ $F = MS(AB)/MSE_b$ ถ้าไม่มีนัยสำคัญก็ใช้ MSE_b ทดสอบ MS(A) ต่อไป แต่ถ้ามีนัยสำคัญก็ใช้ MS(AB) ทดสอบ MS(A) ต่อไป จากการพิจารณาในสมการ (16-1) เห็นได้ว่า ในการทดสอบสปลิต-พลอตนั้น ไม่สามารถประมาณ σ^2 ได้โดยตรงแต่ MSE_b ที่หาได้เป็นการรวมกันของปฏิกิริยาระหว่างบลอกหรือซ้ำ x ผลของแฟกเตอร์ B $(B\beta)$ และปฏิกิริยาระหว่างบลอกหรือซ้ำกับปฏิกิริยาระหว่างแฟกเตอร์ A และ B $(B \times \alpha\beta = B\alpha\beta)$ แต่อนุโลมให้รวมกันเพื่อใช้เป็นค่าทดสอบ ดังนั้น EMS ของ B, AB และ Error (b) ในตาราง 16.2.2 เป็นไปโดยอนุโลมเท่านั้น และจงสังเกตค่า df ของ Error (b)

ตาราง 16.2.3 ผลการทดลองแบบสปลิตพลอต ศึกษาน้ำหนักต้นแห้งของถั่วเขียว 3 พันธุ์ที่ได้รับแสงในความเข้มต่าง ๆ กัน

พันธุ์ถั่วเขียว	ปริมาณแสง	บลอก				รวม
		I	II	III	IV	
	%	กรัม/ต้น				
ก	100	14	11	10	12	47
	90	11	8	10	11	40
	80	6	8	9	10	33
	70	10	8	9	8	35
	60	9	9	8	9	35
	รวม (ก)	**50**	**44**	**46**	**50**	**190**
ข	100	10	11	11	12	44
	90	9	10	9	10	38
	80	9	11	9	10	39
	70	12	10	11	9	42
	60	12	10	10	9	41
	รวม (ข)	**52**	**52**	**50**	**50**	**204**
ค	100	7	11	12	12	42
	90	8	9	9	8	34
	80	7	9	8	9	33
	70	8	10	8	9	35
	60	9	9	7	6	31
	รวม (ค)	**39**	**48**	**44**	**44**	**175**
รวมบลอก		141	144	141	143	569

ตาราง 16.2.4 ตารางเมนพลอต x ซับพลอต (ตาราง AB)

พันธุ์	100	90	80	70	60	รวม (A)
ก	47	40	33	35	35	190
ข	44	38	39	42	41	204
ค	42	34	33	35	31	175
รวม (B)	133	112	105	112	107	569

df error (b) = a(n – 1)(b – 1) = 36

= (n – 1)(b – 1) = Block x Factor B = 12

+ (n – 1) (a – 1)(b – 1) = Block × Factor A × Factor B = 24

ซึ่งแสดงให้เห็นว่า MSE_b คือผลรวมของ MS(Block x B) และ MS(Block × AB)

ตาราง 16.2.5 ผลการวิเคราะห์วาเรียนซ์ของข้อมูลในตาราง 16.3.1

Sources	df	SS	MS	F
Whole plots	(11)			
Blocks	3	0.85	0.28	
Varieties (A)	2	21.03	10.52	4.67
Error (a)	6	13.50	2.25	
Sub-plots	(48)			
Lights(B)	4	41.57	10.39	6.41**
AB	8	13.63	1.70	1.05
Error (b)	36	58.40	1.62	
Total	59	148.98		

CV(A) = 15.80%, CV(B) = 13.40%

จากผลการวิเคราะห์วาเรียนซ์ในตาราง 16.2.5 สามารถทดสอบปัญหาต่าง ๆ ดังนี้

$$F(A) = \frac{MS(A)}{MS(E_a)} = \frac{10.52}{2.25} = 4.67^{ns}[\text{vs } F_{.05}(2, 6) = 5.14]$$

$$F(AB) = \frac{MS(AB)}{MS(E_b)} = \frac{1.70}{1.62} = 1.05^{ns}[\text{vs } F_{.05}(8, 36) = 2.21]$$

$$F(B) = \frac{MS(A)}{MS(E_b)} = \frac{10.39}{1.62} = 6.41^{**}[\text{vs } F_{.01}(4, 36) = 3.89]$$

ในแผนการทดลองแบบสปลิตพลอต สามารถคำนวณสัมประสิทธิ์ของความปรวนแปรได้ดังนี้

(1) CV(A) คือ CV สำหรับเมนพลอต (พันธุ์)

$$CV(A) = \frac{\sqrt{MSE_a}}{\overline{X}...} \times 100 = \frac{\sqrt{2.25}}{9.48} \times 100 = 15.80\%$$

(2) CV(B) คือ CV สำหรับซับพลอต (ความเข้มของแสง)

$$CV(B) = \frac{\sqrt{MSE_b}}{\overline{X}...} \times 100 = \frac{\sqrt{1.62}}{9.48} \times 100 = 13.40\%$$

สัมประสิทธิ์ของความปรวนแปรจะชี้ให้เห็นถึงระดับความเที่ยงตรงของการทดลอง ตามปกติ CV(B) มักจะต่ำกว่า CV(A) เพราะ B เป็นปัญหาที่ทดสอบได้เที่ยงตรงกว่า มีจำนวนแปลงมากกว่าปัญหา A แต่อาจพบว่า CV(B) มากกว่า CV(A) ก็ได้ อนึ่ง CV แต่ละลักษณะที่สังเกตไม่จำเป็นต้องเท่ากัน เพราะ CV จะเปลี่ยนแปลงไปตามธรรมชาติของลักษณะที่สังเกตเสมอ และจะมากหรือน้อยขึ้นอยู่กับลักษณะนั้น และขึ้นอยู่กับความเที่ยงตรงในการทดลองด้วย

16.3 การทดสอบค่าเฉลี่ยในแผนการทดลองแบบสปลิตพลอต

ความคลาดเคลื่อนมาตรฐาน

ความคลาดเคลื่อนมาตรฐานของการทดสอบค่าเฉลี่ยต่าง ๆ ในการทดลองแบบสปลิตพลอตของตัวอย่างในตอน 16.2 แสดงไว้ในตาราง 16.3.1 เห็นได้ว่า MSE_a คือ MSE ของการทดลองแบบ RCB ทั่วไป จึงสามารถใช้ทดสอบความแตกต่างระหว่างค่าเฉลี่ยของเมนพลอตได้โดยตรง การทดสอบปัญหาในซับพลอตก็ใช้ MSE_b ซึ่งเป็นวาเรียนซ์ของซับพลอต อย่างไรก็ดีการทดสอบปัญหาที่คาบเกี่ยวระหว่างเมนพลอตและซับพลอตต้องคำนวณค่าเกี่ยวข้องขึ้นมาต่างหาก ดังนั้นถ้าจะใช้ LSD ก็หา $s_{\bar{d}}$ และค่า LSD ดังนี้

(1) ความแตกต่างระหว่างเมนพลอตหรือพันธุ์ (a_1 vs a_2)

$$s_{\bar{d}} = \sqrt{2(E_a)/nb} = \sqrt{2(2.25)/(4)(5)} = 0.47$$

$$LSD = t_\alpha s_{\bar{d}} = (2.447)(0.47) = 1.15 \text{ กรัม/ต้น}$$

(เปิดตาราง t ที่ระดับความแตกต่าง 0.05, df 6)

(2) ความแตกต่างระหว่างซับพลอตหรือความเข้มของแสง (b_1 vs b_2)

$$s_{\bar{d}} = \sqrt{2(E_b)/na} = \sqrt{2(1.62)/(4)(3)} = 0.52$$

$$LSD = t_\alpha s_{\bar{d}} = (2.029)(0.52) = 1.06 \text{ กรัม/ต้น}$$

(เปิดตาราง t ที่ระดับความแตกต่าง 0.05, df 36)

(3) ความแตกต่างระหว่างความเข้มของแสงในพันธุ์เดียวกัน

$$s_{\bar{d}} = \sqrt{2(E_b)/n} = \sqrt{2(1.62)/4} = 0.90$$

$$LSD = t_\alpha s_{\bar{d}} = (2.029)(0.90) = 1.83 \text{ กรัม/ต้น}$$

(4) ความแตกต่างระหว่างพันธุ์ในความเข้มของแสงเดียวกันหรือความเข้มของแสงต่างกัน (a_1b_1 vs a_2b_1 หรือ a_1b_2 vs a_2b_1)

$$s_{\bar{d}} = \sqrt{2[(b-1)E_b + E_a]/nb} = \sqrt{2[(5-1)1.62 + 2.25]/(4)(5)} = 0.93$$

ค่า LSD สำหรับกรณีนี้ต้องคำนวณหาค่า t′ ดังนี้

$$t' = \frac{(b-1)E_b t_b + E_a t_a}{(b-1)E_b + E_a} = \frac{(4)(1.62)(2.029)+(2.25)(2.447)}{(4)(1.62)+2.25} = 2.14$$

ดังนั้น

$$LSD = t'_{\alpha} s_{\bar{d}} = (2.14)(0.93) = 1.99 \text{ กรัม/ต้น}$$

แล้วนำค่า LSD เหล่านี้ไปเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยของเมนพลอต และซับพลอตต่าง ๆ ดังแสดงในตาราง 16.3.2 ต่อไป

ตาราง 16.3.1 ความคลาดเคลื่อนมาตรฐานที่ใช้ในการเปรียบเทียบระหว่างค่าเฉลี่ยต่าง ๆ [1]

ชนิดของการเปรียบเทียบ	ตัวอย่าง	s_d^2	
1. ระหว่างค่าเฉลี่ยของแฟกเตอร์ A (เมนพลอต)	$a_1 - a_2$	$2E_a/nb$	(16-2)
2. ระหว่างค่าเฉลี่ยของแฟกเตอร์ B (ซับพลอต)	$b_1 - b_2$	$2E_b/na$	(16-3)
3. ระหว่างค่าเฉลี่ยของแฟกเตอร์ B ในแฟกเตอร์ A ระดับเดียวกัน	$a_1b_1 - a_1b_2$	$2E_b/n$	(16-4)
4. ระหว่างค่าเฉลี่ยของแฟกเตอร์ A (ในเมนพลอต)			
1. ในแฟกเตอร์ B ระดับเดียวกัน	$a_1b_1 - a_2b_2$	$2[(b-1)\ E_b + E_a]/nb$	(16-5)
2. ในแฟกเตอร์ B คนละระดับ	$a_1b_2 - a_2b_1$		

[1]**n, a, b = จำนวนบลอก, เมนพลอต และซับพลอต ตามลำดับ , $E_a = MSE_a$, $E_b = MSE_b$**

ตาราง 16.3.2 ค่าเฉลี่ยต่าง ๆ จากข้อมูลในตาราง 16.3.1

พันธุ์	ความเข้มของแสง (%)					เฉลี่ย
	100	90	80	70	60	
			มม./วัน			
ก	11.75	10.00	8.25	8.75	8.75	9.50
ข	11.00	9.50	9.75	10.50	10.25	10.25
ค	10.50	8.50	8.25	8.75	7.75	7.75
เฉลี่ย	10.08	9.33	8.375	9.33	8.92	

การเปรียบเทียบแบบแนวโน้ม [4]

สำหรับทรีตเมนต์ที่มีการกระจายเป็นช่วงแน่นอน และระดับระหว่างช่วงเท่ากัน โดยมีการเพิ่มและการลดเป็นขั้น ๆ เช่น การใส่ปุ๋ย 0, 20, 40, 60 ฯลฯ กก./ไร่ ผู้ทดลองจึงอาจสนใจว่าผลผลิตของพืชมีความสัมพันธ์กับอัตราปุ๋ยในรูปใด เช่น อาจเป็นแบบเส้นตรง หรือเป็นแบบสลับซับซ้อน ดังแสดงในรูป 16.3.1 ถ้าการตอบสนองของทรีตเมนต์มีความสลับซับซ้อน ก็อาจอธิบายโดยใช้โพลี-โนเมียล[5] ซึ่งใช้หลักของรีเกรซชัน สมการของโพลีโนเมียล คือ

$$Y = b_0 + b_1X + b_2X^2 + b_3X^3 + \ldots b_nX^n \qquad \ldots(16\text{-}6)$$

จากสมการ (16-6) อาจทดสอบได้ว่าการสนองตอบของพืชเป็นแบบต่าง ๆ ดังนี้

แบบเส้นตรง (Linear)	เมื่อ	$Y = b_0 + b_1X$
แบบเส้นโค้ง (Quadratic)	เมื่อ	$Y = b_0 + b_1X + b_2X^2$
แบบคลื่น (Cubic)	เมื่อ	$Y = b_0 + b_1X + b_2X^2 + b_3X^3$

ความสัมพันธ์แบบเส้นตรง (linear)

ความสัมพันธ์แบบเส้นโค้ง

ความสัมพันธ์แบบคลื่น

รูป 16.3.1 รูปแบบของความสัมพันธ์ระหว่างผลและปัจจัยที่ใส่ลงไปในการทดลอง ทั้งในทางบวก (+) และลบ (-)

การเปรียบเทียบแต่ละชุดมี 1 df วิธีการวิเคราะห์โพลีโนเมียลเหมือนกับที่กล่าวมาแล้วในตอน 13.9 นั่นเอง ทั้งนี้โดยการเลือกใช้สัมประสิทธิ์ที่เหมาะสม ในกรณีของการเปรียบเทียบผลของช่วงแสงต่อน้ำหนักแห้งของต้นกล้าถั่วเขียว ก็อาจแสดงวิธีการดังตาราง 16.3.3

ตาราง 16.3.3 ผลรวมของทรีตเมนต์และสัมประสิทธิ์โพลีโนเมียล (สัมประสิทธิ์โพลีโนเมียล ได้จากตาราง ผ.13)

	ผลรวมของทรีตเมนต์และสัมประสิทธิ์						
แสง (%)	100	90	80	70	60	ตัวตั้ง	ตัวหาร
ผลรวม	133	112	105	112	107		
Linear	− 2	− 1	0	+ 1	+ 2	2,704	120
Quadratic	+ 2	− 1	− 2	− 1	+ 2	2,116	168
Cubic	− 1	+ 2	0	− 2	+ 1	676	120

สามารถนำค่าจากตาราง 16.3.3 มาคำนวณ sum of squares ต่าง ๆ ได้ ดังนี้

$$SS(L_j) = \frac{(\sum c_i T_i)^2}{na \sum c_i^2}$$

$$SS(Linear) = \frac{[(-2)(133)+(-1)(112)+\ldots+(2)(107)]^2}{(4)(3)[(-2)^2+(-1)^2+\ldots+2^2]} = \frac{(-52)^2}{120} = 22.53$$

$$SS(Quadratic) = \frac{[(2)(133)+(-1)(112)+\ldots+(2)(107)]^2}{(4)(3)[2^2+(-1)^2+\ldots+2^2]} = \frac{46^2}{168} = 12.59$$

$$SS(Cubic) = \frac{[(-1)(133)+(2)(112)+\ldots+(1)(107)]^2}{(4)(3)[(-1)^2+2^2+\ldots+1^2]} = \frac{676}{120} = 5.63$$

การวิเคราะห์อาจใช้ระดับที่สูงกว่านี้ แต่ส่วนมากไม่เกินระดับ Cubic ดังนั้น SS ที่ไม่ได้แยกจึงจัดเป็นส่วนที่เหลือ คือ

$$Residual = SS(B) - SS(Linear) - SS(Quadratic) - SS(Cubic)$$

แล้วนำผลการวิเคราะห์ลงตาราง 16.3.4 ต่อไป

ตาราง 16.3.4 การวิเคราะห์ข้อมูลในตาราง 16.3.2 เพื่อแยกซับพลอตโดยวิธีโปลีโนเมียร์

Sources	df	SS	MS	F
Main plots				
Blocks	3	0.85	0.28	
Varieties (A)	2	21.03	10.51	4.67
Error a (E_a)	6	13.50	2.25	
Sub-plot				
Lights (B)	4	41.57	10.15	6.38**
Linear	1	22.53	22.53	13.91**
Quadratic	1	12.59	12.59	7.77**
Cubic	1	5.63	5.63	3.48
Residual	1	0.83	0.80	0.49
AB	8	13.63	1.70	1.05
Error (b)	36	58.40	1.62	
Total	59	148.40		

16.4 ค่าสูญหายและข้อดีข้อเสียในการทดลองแบบสปลิตพลอต

ในการทดลองโดยใช้แผนการทดลองนี้ อาจมีค่าสูญหายเกิดขึ้น ถ้าสังเกตให้ดีจะพบว่า ในแต่ละเมนพลอต การจัดระเบียบของหน่วยทดลองคล้ายกับแผนการทดลองแบบ RCB ดังนั้นในการคำนวณค่าสูญหายใช้สมการเดียวกัน ในกรณีที่มีค่าสูญหาย 1 ค่า คำนวณโดยใช้สมการ ดังนี้

$$\hat{X} = \frac{nB + bT - G}{(n-1)(b-1)}$$

เมื่อ B = ผลรวมของบลอกในเมนพลอตที่มีค่าสูญหาย

T = ผลรวมของซับพลอตที่มีค่าสูญหาย

G = ผลรวมของทั้งเมนพลอตที่มีค่าสูญหาย

n = จำนวนบลอก

b = จำนวนระดับของซับพลอต

ตาราง 16.4.1 ผลการทดลองแบบสปลิตพลอตที่มีค่าสูญหาย

เมนพลอต	ซับพลอต	I	II	III	รวม	
a_1	b_1	4	6	8	18	
	b_2	3	4	6	13	
a_2	b_1	5	8	6	19	28(G)
	b_2	4	X	5	9(T)	

จากตาราง อาจคำนวณค่าสูญหาย X ดังนี้

$$\hat{X} = \frac{3(8)+2(9)-28}{(3-1)(2-1)} = 7$$

แล้วนำ $\hat{X}$ ที่คำนวณได้ไปใส่ลงในช่องที่มีค่าสูญหาย และวิเคราะห์ตามปกติต่อไป อย่างไรก็ดี df ของ Total และของ E_b ลดลงอย่างละ 1 ส่วนเบี่ยงเบนมาตรฐานของความแตกต่าง ($s_{\bar{d}}$) สำหรับชุดเปรียบเทียบต่าง ๆ แสดงไว้ในตาราง 16.3.6

ตาราง 16.4.2 **ความคลาดเคลื่อนมาตรฐานของความแตกต่างของแผนการทดลองแบบสปลิตพลอต เมื่อค่าสูญหาย 1 ค่า**

	ชุดเปรียบเทียบ	$s_{\bar{d}}$	
1.	ระหว่างค่าเฉลี่ยของเมนพลอต (a_1 vs a_2)	$\sqrt{\frac{2(E_a + fE_b)}{nb}}$	...(16-7)
2.	ระหว่างค่าเฉลี่ยของซับพลอต (b_1 vs b_2)	$\sqrt{\frac{2E_b\left[1+\frac{fb}{a}\right]}{na}}$	...(16-8)
3.	ระหว่างค่าเฉลี่ยของซัพพลอตในเมนพลอตเดียวกัน (a_1b_1 vs a_1b_2)	$\sqrt{\frac{2E_b\left(1+\frac{fb}{a}\right)}{n}}$	...(16-9)

เมื่อมีค่าสูญหายค่าเดียว $f = 1/[2(n-1)(b-1)]$

ข้อดีข้อเสียของการทดลองแบบสปลิตพลอต

ข้อดี

1. สามารถศึกษา 2 ปัญหาได้พร้อมกัน ทำให้ทราบปฏิกิริยาระหว่างปัญหา สามารถใช้ประโยชน์เมนพลอต ประหยัดเวลา และได้รับความเที่ยงตรงในการศึกษาปัญหาในซับพลอตได้ดีกว่าแผนการทดลองแบบ RCB

2. สามารถดัดแปลงการทดลองได้หลายรูปแบบตามความเหมาะสมและตามชนิดของปัญหา เช่น อาจใช้แผนการทดลองแบบต่าง ๆ ในเมนพลอต

ข้อเสีย

1. เป็นการทดลองที่มีความสลับซับซ้อน ยุ่งยากในการทดลอง วิเคราะห์ และแปลผลของข้อมูล
2. คำตอบเกี่ยวกับปัญหาในเมนพลอต มีความเที่ยงตรงน้อยกว่าในการทดลองแบบ RCB และน้อยกว่าปัญหาในซับพลอต
3. การวิเคราะห์จะยุ่งยากยิ่งขึ้นเมื่อมีค่าสูญหาย

16.5 สปลิตพลอตที่จัดซับพลอตแบบแฟกตอเรียล

ในการทดลองแบบสปลิตพลอต แต่ละซับพลอตอาจจะเกิดจากการจัดทรีตเมนต์แบบแฟกตอเรียล เช่น การใช้ทดลองเปรียบเทียบพันธุ์พืช 3 พันธุ์ (A_1, A_2, A_3) ใช้ปุ๋ย 2 ระดับ (b_0 , b_1) และระยะปลูก 2 ระยะ (c_0, c_1) โดยจัดปุ๋ยและระยะปลูกแบบแฟกตอเรียล และทดลอง 2 ซ้ำ ดังรูป 16.5.1 เมื่อนำไปวิเคราะห์วาเรียนซ์ได้ผลดังตาราง 16.5.1

ตัวอย่าง ได้ทำการทดลองเพื่อเปรียบเทียบแฟกเตอร์ A, B และ C โดยใช้แฟกเตอร์ A จำนวน 3 ระดับ อยู่ในเมนพลอต แฟกเตอร์ B และ C จำนวน 2 และ 3 ระดับ ตามลำดับ และจัดชุดแบบแฟกตอเรียลอยู่ในซับพลอต ทำการทดลอง 4 ซ้ำ ดังแสดงในตาราง 16.5.2

วิธีการวิเคราะห์

นำข้อมูลในตาราง 16.5.2 มาทำการวิเคราะห์วาเรียนซ์ดังนี้

$$\text{CF} = \frac{(\sum X_{ijkl})^2}{nabc} = \frac{(491)^2}{72} = 3{,}348.34$$

$$\text{Total SS(TSS)} = \sum X_{ijkl}^2 - \text{CF}$$

$$= 4^2 + 6^2 + \ldots + 4^2 - 3{,}348.34 = 392.66$$

I					II				
A_1	b_1c_0	b_0c_1	b_0c_0	b_1c_1	b_1c_0	b_0c_1	b_0c_0	b_1c_1	A_2
A_2	b_1c_1	b_1c_0	b_0c_1	b_0c_0	b_0c_0	b_1c_1	b_1c_0	b_0c_0	A_1
A_3	b_1c_0	b_0c_1	b_0c_0	b_1c_1	b_1c_1	b_1c_0	b_0c_0	b_0c_1	A_3

รูป 16.5.1 การทดลองแบบสปลิตพลอตที่จัดซับพลอตแบบแฟกตอเรียล

ตาราง 16.5.1 การวิเคราะห์วาเรียนซ์ในการทดลองแบบสปลิตพลอตที่จัดซับพลอตแบบแฟกตอเรียล

Sources	df
Main plot	
Blocks	$n-1 = 1$
A	$a-1 = 2$
Error(a)	$(n-1)(a-1) = 2$
Sub-plots	
B	$b-1 = 1$
C	$c-1 = 1$
BC	$(b-1)(c-1) = 1$
A × sub-plot	
AB	$(a-1)(b-1) = 2$
AC	$(a-1)(c-1) = 2$
ABC	$(a-1)(b-1)(c-1) = 2$
Error(b)	$a(n-1)(bc-1) = 9$
Total	$nabc-1 = 23$

ขั้นที่ 1 วิเคราะห์เมนพลอต

$$\text{Main plot SS} = \frac{\sum A_i^2}{bc} - CF$$

$$= \frac{45^2 + 41^2 + ... + 36^2}{6} - 3{,}348.34 = 44.49$$

$$\text{Block SS (SSB)} = \frac{\sum R^2}{abc} - CF$$

$$= \frac{117^2 + 115^2 + 127^2 + 132^2}{36} - 3.348.34 = 10.93$$

$$\text{SS (A)} = \frac{\sum A^2}{nbc} - CF$$

$$= \frac{181^2 + 164^2 + 146^2}{24} - 3{,}348.34 = 25.53$$

$$\text{Error (a) SS} = \text{Main plot SS} - \text{SSB} - \text{SS(A)}$$

$$= 44.49 - 10.93 - 25.54 = 8.03$$

ตาราง 16.5.2 ข้อมูลจากการทดลองแบบสปลิต-พลอตที่จัดซับพลอตแบบแฟกตอเรียล

		I	II	III	IV	Total
A_1	B_1C_1	4	6	8	6	24
	B_1C_2	7	5	8	8	28
	B_1C_3	10	8	8	9	35
	B_2C_1	6	4	3	4	17
	B_2C_2	8	10	9	10	37
	B_2C_3	10	8	10	12	40
	รวม	45	41	46	49	181
A_2	B_1C_1	4	6	4	4	18
	B_1C_2	7	10	10	8	35
	B_1C_3	7	6	5	8	26
	B_2C_1	8	4	10	10	32
	B_2C_2	5	8	5	10	28
	B_2C_3	5	6	7	7	25
	รวม	36	40	41	47	164
A_3	B_1C_1	10	8	10	10	38
	B_1C_2	4	6	4	4	18
	B_1C_3	3	4	6	4	17
	B_2C_1	6	4	8	10	28
	B_2C_2	4	7	8	4	23
	B_2C_3	9	5	4	4	22
	รวม	36	34	40	36	146
รวม		117	115	127	132	491

ตาราง 16.5.3 ตารางปฏิกิริยาระหว่างค่าต่าง ๆ จากตาราง 16.5.2

ก. ตาราง $A \times B$

	A_1	A_2	A_3	
B_1	87	79	73	239
B_2	94	85	73	252

ข. ตาราง $A \times C$

	A_1	A_2	A_3	
C_1	41	55	66	157
C_2	65	63	41	169
C_3	75	51	39	165

ค. ตาราง $B \times C$

	C_1	C_2	C_3	
B_1	80	81	78	239
B_2	77	88	87	252
	157	169	165	491

ขั้นที่ 2 วิเคราะห์ซับพลอต

ในการวิเคราะห์ซับพลอตนั้นต้องจัดทำตารางดังตาราง 16.5.3

$$SS(B) = \frac{\sum B^2}{nac} - CF$$

$$= \frac{239^2 + 252^2}{36} - 3{,}348.34 = 2.35$$

$$SS(C) = \frac{\sum C^2}{nab} - CF$$

$$= \frac{157^2 + 169^2 + 165^2}{24} - 3{,}348.34 = 3.11$$

$$SS(BC) = \frac{\sum (BC)^2}{na} - CF - SS(B) - SS(C)$$

$$= \frac{80^2 + 81^2 + \ldots + 87^2}{12} - 3.348.34 - 2.35 - 3.11 = 3.44$$

$$SS(AB) = \frac{\sum (AB)^2}{nc} - CF - SS(A) - SS(B)$$

$$= \frac{87^2 + 79^2 + \ldots + 73^3}{12} - 3{,}348.34 - 25.53 - 2.35 = 1.19$$

$$\mathrm{SS(AC)} = \frac{\sum(\mathrm{AC})^2}{\mathrm{nb}} - \mathrm{CF} - \mathrm{SS(A)} - \mathrm{SS(C)}$$

$$= \frac{41^2 + 50^2 + \ldots + 3a^2}{8} - 3{,}348.34 - 25.53 - 3.11 = 142.89$$

$$\mathrm{SS(ABC)} = \frac{\sum(\mathrm{ABC})^2}{\mathrm{n}} - \mathrm{CF} - \mathrm{SS(A)} - \mathrm{SS(B)} - \mathrm{SS(C)} - \mathrm{SS(AB)} - \mathrm{SS(AC)} - \mathrm{SS(BC)}$$

$$= \frac{24^2 + 28^2 + \ldots + 22^2}{4} - 3{,}348.34 - 25.33 - 2.35$$

$$-3.11 - 1.19 - 142.59 - 3.44 = 61.89$$

$$\text{Error (b) SS} = \text{TSS} - \text{Main-plot SS} - \mathrm{SS(B)} - \mathrm{SS(C)} - \mathrm{SS(BC)} - \mathrm{SS(AB)}$$

$$- \mathrm{SS\ (AC)} - \mathrm{SS\ (ABC)}$$

$$= 392.66 - 44.49 - 2.35 - 3.11 - 3.44 - 1.19 - 142.89 - 61.89$$

$$= 133.30$$

ขั้นที่ 3 นำผลการวิเคราะห์ลงตารางวิเคราะห์วาเรียนซ์ดังตาราง 16.5.4 ต่อไป

ตาราง 16.5.4 ผลการวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 16.5.1 ที่ทดลองโดยจัดสปลิตพลอตแบบแฟกตอเรียล

Sources	df	SS	MS
Main plot			
Blocks	3	10.93	3.64
A	2	25.53	12.77
Error (a)	6	8.03	1.34
Sub-plot			
B	1	2.35	2.35
C	2	3.11	1.56
BC	2	3.44	1.72
A × sub-plot			
AB	2	1.19	0.60
AC	4	142.89	35.72
ABC	4	61.89	15.47
Error (b)	45	133.30	2.96
Total	71	392.66	

16.6 แผนการทดลองแบบสปลิต-สปลิตพลอต

แผนการทดลองแบบสปลิต-สปลิตพลอต[6] เป็นการขยายการทดลองเพิ่มเติมจากการทดลองแบบสปลิตพลอต คือเมื่อศึกษา 3 ปัญหาที่เกี่ยวเนื่องกัน เช่น พันธุ์พืช-ปุ๋ย-และระยะปลูก อาหารสัตว์-พันธุ์สัตว์-อายุสัตว์ อายุของรถยนต์-ขนาดของรถยนต์-ความเร็ว ฯลฯ เหล่านี้ อาจนำมาศึกษาพร้อมกันในการทดลองเดียว ทำให้ประหยัดเวลา การลงทุน และให้คำตอบเป็นชุด และทราบปฏิกิริยาระหว่างปัญหา การทดลองชนิดนี้เกิดจากการแบ่งซับพลอตให้ย่อยลงไปอีก เพื่อใส่ปัญหาที่ 3 เข้าไปภายใต้ ปัญหาที่สอง หน่วยทดลองที่แบ่งย่อยลงไปนี้เรียกว่า ซับ-ซับพลอต ในแผนการทดลองแบบสปลิต- สปลิตพลอต ขนาดของแปลงหรือหน่วยทดลองจะมีขนาดโตและเล็กลดหลั่นลงมาจากเมนพลอต เป็นซับพลอตและซับ-ซับพลอต[7] หรือเรียกว่าหน่วยใหญ่ หน่วยย่อย และหน่วยย่อย–ย่อย เนื่องจากหน่วยเหล่านี้ทำให้ความสม่ำเสมอในหน่วยทดลองไม่เท่ากัน คือในหน่วยใหญ่สม่ำเสมอน้อยกว่า และจำนวนซ้ำก็ไม่เท่ากัน คือหน่วยใหญ่มีจำนวนซ้ำน้อยกว่า จึงทำให้ความเที่ยงตรงในการทดลองไม่เท่ากัน คือความเที่ยงตรงในซับ-ซับพลอต สูงกว่าในซับพลอต และในเมนพลอตตามลำดับ

การจัดแปลงทดลองและการสุ่ม

สมมุติว่าเราทำการทดลองเปรียบพันธุ์พืช 3 พันธุ์ ระยะปลูก 3 ระยะ ปุ๋ย 2 ระดับ ให้ปัญหาเหล่านี้อยู่ในเมนพลอต ซับพลอตและซับ-ซับพลอตตามลำดับ ใช้แผนการทดลองแบบ RCB 3 ซ้ำ ก็อาจกำหนดชื่อและระดับตัวแปรดังนี้

เมนพลอต A ระดับ a_1, a_2, a_3

ซับพลอต B ระดับ b_1, b_2, b_3

ซับ-ซับพลอต C ระดับ c_1, c_2

แล้วทำการสุ่มปัญหาลงในเมนพลอต ซับพลอต และซับ-ซับพลอต แต่ละระดับเหมือนการทดลองแบบ RCB นั่นเองแล้วจะได้ผลดังรูป 16.6.1

การแยกแหล่งของความแปรปรวน

ในการทดลองที่แฟกเตอร์ A, B, C มี 3, 3 และ 2 ระดับ และกระทำ 3 บลอก หรือ 3 ซ้ำ สามารถแยกแหล่งของความปรวนแปรและ degree of freedom ดังตาราง 16.6.1 และแสดง EMS ดังตาราง 16.6.2

ก. สุ่มเพื่อวางเมนพลอต

a_3	a_2	a_1
a_2	a_1	a_3
a_1	a_3	a_2
I	II	III

ข. สุ่มเพื่อวางซับพลอตลงในแต่ละเมนพลอต

a_3	b_3	b_2	b_1	b_1	b_3	b_2	b_3	b_2	b_1	a_1
a_2	b_1	b_2	b_3	b_2	b_3	b_1	b_1	b_2	b_3	a_3
a_1	b_2	b_3	b_2	b_3	b_2	b_1	b_2	b_3	b_1	a_2

ค. สุ่มเพื่อวางซับ-ซับพลอต ลงในซับพลอต

$a_3b_3c_1$	$a_3b_2c_2$	$a_3b_1c_1$						
$a_3b_3c_2$	$a_3b_2c_1$	$a_3b_1c_2$						
$a_2b_1c_1$	$a_2b_2c_2$	$a_2b_3c_1$						
$a_2b_1c_2$	$a_2b_2c_1$	$a_2b_3c_2$						
$a_1b_2c_2$	$a_1b_3c_1$	$a_1b_2c_2$						
$a_1b_2c_1$	$a_1b_3c_2$	$a_1b_2c_1$						
	I			II			II	

รูป 16.6.1 การวางแผนการทดลองแบบสปลิต-สปลิตพลอต

ตาราง 16.6.1 แหล่งของความแปรปรวนในแผนการทดลองแบบสปลิต-สปลิตพลอต

Sources	**df**
Main plot	8
Blocks	n - 1 = 2
A	a - 1 = 2
Error (a) (E_a)	(n – 1)(a - 1) = 4
Sub-plots	18
B	b - 1 = 2
AB	(a – 1)(b - 1) = 4
Error (b) (E_b)	a(n - 1)(b - 1) = 12
Sub-sub plots	27
C	c - 1 = 1
AC	(a – 1)(c - 1) = 2
BC	(b – 1)(c - 1) = 2
ABC	(a - 1) (b – 1)(c - 1) = 4
Error $(c)E_c$	ab (n - 1)(c - 1) = 18
Total	= 53

ที่แท้จริงค่า df ของการทดลองแบบสปลิตพลอต หรือ สปลิต – สปลิตพลอต คือปฏิกิริยาระหว่างซ้ำหรือบลอกกับเมนพลอตและซับพลอต ดังนี้

$$
\begin{aligned}
\text{Error (b)} &= (n-1)(b-1) &&= 4\\
&\quad + (n-1)(a-1)(b-1) &&= \underline{8}\\
& && \;\;\; 12\\
\text{Error (c)} &= (n-1)(c-1) &&= 2\\
&\quad + (n-1)(b-1)(c-1) &&= 4\\
&\quad + (n-1)(a-1)(c-1) &&= 4\\
&\quad + (n-1)(a-1)(b-1)(c-1) &&= \underline{8}\\
& && \;\;\; 18
\end{aligned}
$$

ซึ่งอาจหาค่านี้โดยทางลัดจากสมการที่แสดงไว้ในตาราง 16.6.1

แบบจำลองทางคณิตศาสตร์

แบบจำลองทางคณิตศาสตร์ของการทดลองแบบสปลิต – สปลิตพลอตมีดังนี้

$$\begin{aligned} X_{ijkl} = {} & \mu + B_i + \alpha_j + (B\alpha)_{ij} + \beta_k + (B\beta)_{ik} + (\alpha\beta)_{jk} + (B\alpha\beta)_{ijk} \\ & + \gamma_\ell + (B\gamma)_{il} + (\alpha\gamma)_{jl} + (B\alpha\gamma)_{ijl} + (\beta\gamma)_{kl} + (B\beta\gamma)_{ikl} \\ & + (\alpha\beta\gamma)_{jkl} + (B\alpha\beta\gamma)_{ijkl} \qquad \text{...(16-10)} \end{aligned}$$

ซึ่งเป็นแบบจำลองเต็มรูปแบบของการทดลองแบบสปลิต-สปลิตพลอต แต่อาจลดขนาดของแบบจำลองลงมา ดังนี้

$$\begin{aligned} X_{ijkl} = {} & \mu + B_i + \alpha_j + \delta_{ij} + \beta_k + (\alpha\beta)_{jk} + \rho_{ijk} + \gamma_l + (\alpha\gamma)_{jl} \\ & + (\beta\gamma)_{kl} + (\alpha\beta\gamma)_{jkl} + \varepsilon_{ijkl} \qquad \text{...(16-11)} \end{aligned}$$

และกำหนดให้

- μ = ค่าเฉลี่ยของประชากร
- B = ผลของบลอกหรือซ้ำ
- α = ผลของเมนพลอตหรือแฟกเตอร์ A
- β = ผลของซับพลอตหรือแฟกเตอร์ B
- γ = ผลของซับ-ซับพลอตหรือแฟกเตอร์ C
- δ = ความคลาดเคลื่อนในระดับเมนพลอต หรือ error (a)
- ρ = ความคลาดเคลื่อนในระดับซับพลอต หรือ error (b)
- ε = ความคลาดเคลื่อนในระดับซับ-ซับพลอต หรือ error (c)
- i = 1, 2, ..., n (n = จำนวนบลอกหรือซ้ำ)
- j = 1, 2, ..., a (a = จำนวนระดับของเมนพลอต (A)
- k = 1, 2, ... , b (b = จำนวนระดับของซับพลอต (B)
- l = 1, 2, ... , c (c = จำนวนระดับของซับ-ซับพลอต (C)

จากสมการ (16-10) $(B\alpha)_{ij} = \delta_{ij}$, $(B\beta)_{ik} + (B\alpha\beta)_{ijk} = \rho_{ijk}$, $(B\gamma)_{i\ell} + (B\alpha\gamma)_{ij\ell} + (B\beta\gamma)_{ik\ell} + (B\alpha\beta\gamma) = \varepsilon_{ijk\ell}$ ซึ่งแสดงในสมการ (16-11)

ตัวอย่าง ในการเปรียบเทียบแฟกเตอร์ A, B และ C อย่างละ 3, 3 และ 2 ระดับ โดยให้แฟกเตอร์ A เป็นเมนพลอต ส่วน B และ C เป็นซับพลอตและซับ-ซับพลอตตามลำดับ โดยทำการทดลองจำนวน 4 บลอก ดังตาราง 16.6.3

ตาราง 16.6.3 ข้อมูลจากแผนการทดลองแบบสปลิต-สปลิตพลอต

แฟกเตอร์			พลอก				ผลรวม
A	B	C	I	II	III	IV	
A_1	B_1	C_1	6	7	4	5	22
		C_2	6	10	6	8	30
	B_2	C_1	5	9	6	5	25
		C_2	8	9	6	8	31
	B_3	C_1	6	7	6	5	24
		C_2	9	12	7	10	38
		รวม	40	54	35	41	170
A_2	B_1	C_1	4	5	4	5	18
		C_2	6	8	7	6	27
	B_2	C_1	7	10	6	7	30
		C_2	8	10	8	8	34
	B_3	C_1	6	8	5	7	26
		C_2	9	12	10	8	39
		รวม	40	53	40	41	174
A_3	B_1	C_1	6	10	7	12	35
		C_2	7	8	9	5	29
	B_2	C_1	8	9	10	10	37
		C_2	8	10	12	9	39
	B_3	C_1	10	12	9	7	38
		C_2	11	12	10	8	41
		รวม	50	61	57	51	219
รวม			130	168	132	133	563

ตาราง 16.6.4 ตารางปฏิกิริยาระหว่างแฟกเตอร์ต่าง ๆ

ก. ตาราง R × A

แฟกเตอร์	I	II	III	IV	รวม
A_1	40	54	35	41	170
A_2	40	53	40	41	174
A_3	50	61	57	51	219
รวม	130	168	132	133	563

ข. ตาราง A × B

แฟกเตอร์	B_1	B_2	B_3	รวม
A_1	52	56	62	170
A_2	45	64	65	174
A_3	64	76	79	219
รวม	161	196	206	563

ค. ตาราง A × C

แฟกเตอร์	A_1	A_2	A_3	รวม
C_1	71	74	110	255
C_2	99	100	109	308
รวม	170	174	219	563

ง. ตาราง B × C

แฟกเตอร์	B_1	B_2	B_3	รวม
C_1	75	92	88	255
C_2	86	104	118	308
รวม	161	196	206	563

จ. ตาราง R × AB

แฟกเตอร์		I	II	III	IV	รวม
A_1	B_1	12	17	10	13	52
	B_2	13	18	12	13	56
	B_3	15	19	13	15	62
A_2	B_1	10	13	11	11	45
	B_2	15	20	14	15	64
	B_3	15	20	15	15	65
A_3	B_1	13	18	16	17	64
	B_2	16	19	22	19	76
	B_3	21	24	19	15	79

การวิเคราะห์วาเรียนซ์ดำเนินการดังนี้

ขั้นที่ 1. วิเคราะห์ความปรวนแปรทั้งหมด

$$CF = \frac{\sum X_{ijkl}^2}{nabc} = \frac{(563)^2}{(4)(3)(3)(2)} = 4{,}402.35$$

$$TSS = \sum X_{ijkl}^2 - CF$$

$$= 6^2 + 7^2 + \ldots + 8^2 - 4{,}402.25 = 332.65$$

ขั้นที่ 2. วิเคราะห์เมนพลอต

$$\text{Main plot SS} = \frac{\sum RA^2}{bc} - CF$$

$$= \frac{40^2 + 54^2 + \ldots 51^2}{6} - 4{,}402.35 = 128.15$$

$$\text{Block SS(SSB)} = \frac{\sum R^2}{abc} - CF$$

$$= \frac{130^2 + 168^2 + 132^2 + 133^2}{18} - 4{,}402.35 = 55.26$$

$$SS(A) = \frac{\sum A^2}{nbc} - CF$$

$$= \frac{170^2 + 174^2 + 219^2}{24} - 4{,}402.35 = 61.69$$

$$SS(E_a) = \text{Main plot SS} - SSB - SS(A)$$

$$= 128.15 - 55.26 - 61.69 = 11.19$$

ขั้นที่ 3 วิเคราะห์ซับพลอต

$$\text{Sub-plot SS} = \frac{\sum (RAB)^2}{c} - CF$$

$$= \frac{12^2 + 17^2 + \ldots + 15^2}{2} - 4{,}402.35 = 209.15$$

$$SS(B) = \frac{\sum B^2}{nac} - CF$$

$$= \frac{161^2 + 196^2 + 206^2}{24} - 4{,}402.35 = 46.53$$

$$SS(AB) = \frac{\sum (AB)^2}{nc} - CF - SS(A) - SS(B)$$

$$= \frac{52^2 + 56^2 + ... + 79^2}{8} - 4{,}402.35 - 61.69 - 46.53 = 7.31$$

$$SS(E_b) = \text{Sub-plot SS} - \text{Main plot SS} - SS(B) - SS(AB)$$

$$= 209.15 - 128.15 - 43.53 - 7.31 = 27.17$$

ขั้นที่ 4 วิเคราะห์ซับ - ซับพลอต

$$SS(C) = \frac{\sum C^2}{nab} - CF$$

$$= \frac{255^2 + 308^2}{36} - 4{,}402.35 = 39.01$$

$$SS(AC) = \frac{\sum (AC)^2}{nb} - CF - SS(A) - SS(C)$$

$$= \frac{71^2 + 99^2 + ... + 109^2}{12} - 4{,}402.35 - 61.69 - 39.01 = 21.86$$

$$SS(BC) = \frac{\sum (BC)^2}{na} - CF - SS(B) - SS(C)$$

$$= \frac{75^2 + 86^2 + ... + 118^2}{12} - 4{,}402.35 - 46.53 - 39.01 = 9.53$$

$$SS(E_c) = \frac{\sum (ABC)^2}{n} - CF - SS(A) - SS(B) - SS(C) - SS(AB)$$

$$- SS(AC) - SS(BC)$$

$$= \frac{(22^2 + 30^2 + ... + 41^2)}{4} - 4{,}402.35 - 61.69 - 46.53 - 39.01$$

$$- 7.31 - 21.86 - 9.53 = 5.97$$

$$SS(E_c) = TSS - \text{ค่า SS ทั้งหมด} - SS(E_b)$$

$$= TSS - SSB - SS(A) - SS(E_a) - SS(B) - SS(AB) - SS(E_b)$$

$$- SS(C) - SS(AC) - SS(BC) - SS(ABC)$$

$$= 332.65 - 55.26 - 61.69 - 11.19 - 46.53 - 7.31 - 27.17$$

$$-39.01 - 21.86 - 9.53 - 5.97 = 47.13$$

ขั้นที่ 5 นำผลการวิเคราะห์ลงตารางวิเคราะห์วาเรียนซ์ดังตาราง 16.6.5 เมื่อนำผลลงตารางวิเคราะห์วาเรียนซ์แล้วก็ดำเนินการคำนวณค่า mean square ตามปกติ เช่น

$$MSB = \frac{SSB}{n-1} = \frac{55.26}{3} = 18.43$$

และคำนวนค่า mean square อื่น ๆ ในทำนองเดียวกัน

ตาราง 16.6.5 ตารางวิเคราะห์วาเรียนซ์ของการทดลองแบบสปลิต-สปลิตพลอต

Sources	df	SS	MS	F
Blocks	3	55.26	18.42	
A	2	61.69	30.85	16.53**
Error (a)	6	11.19	1.87	
B	2	46.53	23.26	15.41**
AB	4	7.31	1.83	
Error (b)	18	27.17	1.51	
C	1	39.01	39.01	22.35**
AC	2	21.86	10.93	6.25**
BC	2	9.53	4.76	
ABC	4	5.97	1.49	
Error (c)	27	47.13	1.75	
Total	71	332.65		

CV (A) = 17.50%, CV (B) = 15.70%, CV(C) = 16.90%

ผลจากการทดสอบโดยวิธี F-test โดยใช้ MSE ที่เหมาะสมพบว่าแฟกเตอร์ A, B และ C และปฏิกิริยา AC มีความแตกต่างในทางสถิติระดับ 0.01

ขั้นที่ 6 เปรียบเทียบค่าเฉลี่ย

การเปรียบเทียบค่าเฉลี่ยอาจใช้วิธี LSD หรือ DMRT ในการใช้วิธี LSD อาจคำนวณค่าความคลาดเคลื่อนมาตรฐานของความแตกต่างดังสมการที่แสดงในตาราง 16.6.6 และสามารถหา LSD บางค่าดังนี้

1. เปรียบเทียบระหว่าง เมนพลอต (a_1 vs a_2)

$$LSD = (t_{.01,6})(s_d) = (3.707)(\sqrt{(2)(1.87)/24} = 1.46$$

2. เปรียบเทียบระหว่างซับพลอต (b_1 Vs b_2)

$$LSD = (t_{.01,18})(s_{\bar{d}}) = (2.878)(\sqrt{(2)(1.51)/24} = 1.02$$

3. เปรียบเทียบระหว่างซับ-ซับพลอต (c_1 vs c_2)

$$\text{LSD} = (t.01,27)\,(s_{\bar{d}}) = (2.771)\,(\sqrt{(2)\,(1.75)/36} = 0.86$$

การเปรียบเทียบค่าเฉลี่ยบางค่าแสดงไว้ในตาราง 16.6.7

ตาราง 16.6.6 ความคลาดเคลื่อนมาตรฐานของความแตกต่างระหว่างค่าเฉลี่ย จากการทดลองแบบสปลิต-สปลิตพลอต

เลขที่	ชุดเปรียบเทียบ	$s_{\bar{d}}$	ค่า t
1.	ค่าเฉลี่ยของเมนพลอต (a_0 vs a_1)	$\sqrt{2E_a/nbc}$	
2.	ค่าเฉลี่ยของซับพลอต (b_o vs b_1)	$\sqrt{2E_b/nac}$	
3.	ค่าเฉลี่ยของซับ-ซับพลอต (c_o vs c_1)	$\sqrt{2E_c/nab}$	
4.	ค่าเฉลี่ยของซับพลอตในเมนพลอตระดับเดียวกัน (a_ob_o vs a_ob_1)	$\sqrt{2E_b/nc}$	
5.	ค่าเฉลี่ยของซับ-ซับพลอตในเมนพลอตระดับเดียวกัน (a_oc_o vs a_oc_1)	$\sqrt{2E_c/nb}$	
6.	ค่าเฉลี่ยของซับ-ซับพลอตในซับพลอตระดับเดียวกัน (b_oc_o vs b_oc_1)	$\sqrt{2E_c/na}$	
7.	ค่าเฉลี่ยของซับ-ซับพลอตในเมนพลอตและซับพลอตระดับเดียวกัน ($a_ob_oc_o$ vs $a_ob_oc_1$)	$\sqrt{2E_c/n}$	
8.	ค่าเฉลี่ยในเมนพลอตในซับพลอตระดับเดียวกัน (a_ob_o vs a_1b_o)	$\sqrt{\frac{2[(b-1)\,E_b + E_a]}{nbc}}$	$\frac{(b-1)\,E_bt_b + E_at_a}{(b-1)\,E_b + E_a}$
9.	ค่าเฉลี่ยในเมนพลอตในซับ-ซับพลอตระดับเดียวกัน (a_oc_o vs a_1c_o)	$\sqrt{\frac{2[(c-1)\,E_c + E_a)]}{nbc}}$	$\frac{(c-1)\,E_ct_c + E_at_a}{(c-1)\,E_c + E_a}$
10.	ค่าเฉลี่ยในซับพลอตในซับ-ซับพลอตระดับเดียวกัน (b_oc_o vs b_1c_0)	$\sqrt{\frac{2[(c-1)\,E_c + E_b)}{nac}}$	$\frac{(c-1)\,E_ct_c + E_bt_b}{(c-1)\,E_c + E_b}$

ตาราง 16.6.7 เปรียบค่าเฉลี่ยของข้อมูลในตาราง 16.6.3

ก. ในเมนพลอตและซัพพลอต

แฟกเตอร์	A_1	A_2	A_3	เฉลี่ย
B_1	6.50	5.63	8.00	6.71
B_2	7.00	8.00	9.50	8.17
B_3	7.75	8.12	9.87	8.58
	7.08	7.25	9.12	

ชุดเปรียบเทียบ	LSD (.01)
ระหว่าง A ในแต่ละ B	2.04
ระหว่าง B ในแต่ละ A	1.76
ระหว่าง A	1.46
ระหว่าง B	1.02

ข. ในเมนพลอตและซัพ-ซัพพลอต

แฟกเตอร์	A_1	A_2	A_3	เฉลี่ย
C_1	5.92	6.17	9.17	7.08
C_2	8.25	8.33	9.08	8.56
เฉลี่ย	7.08	7.25	9.12	

ชุดเปรียบเทียบ	LSD (.01)
ระหว่าง C ในแต่ละ A	1.49
ระหว่าง A ในแต่ละ C	1.79
ระหว่าง C	0.86

ค. ในซัพพลอตและซัพ-ซัพพลอต

แฟกเตอร์	B_1	B_2	B_3	เฉลี่ย
C_1	6.25	7.67	7.33	7.08
C_2	7.17	8.67	9.83	8.55

ชุดเปรียบเทียบ	LSD (.01)
ระหว่าง C ในแต่ละ B	1.49
ระหว่าง B ในแต่ละ C	1.46

16.7 แผนการทดลองแบบสตริปพลอต

ในการทดลองที่ศึกษาปัญหาหรือแฟกเตอร์ 2 ชนิดพร้อมกัน แทนที่จะทดลองแบบสปลิตพลอตที่มีแปลงใหญ่และแปลงย่อย ก็อาจจัดเป็นแบบแปลงใหญ่ตัดขวางกัน คือให้ปัญหาหนึ่งอยู่ในแนวตั้งอีกปัญหาหนึ่งอยู่ในแนวนอน การทดลองเช่นนี้เรียกว่าแผนการทดลองแบบสตริปพลอตหรือสปลิตบลอก[8] เหมาะสำหรับการศึกษาปฏิกิริยาระหว่างปัญหามากกว่าที่ต้องการศึกษาอิทธิพลหลัก เพราะอิทธิพลหลักทั้ง 2 ชนิด อยู่ในแปลงใหญ่ ให้ความถูกต้องน้อยกว่าอยู่ในแปลงย่อยของการทดลองแบบสปลิตพลอตปกติ

การจัดแปลงทดลองและการสุ่ม

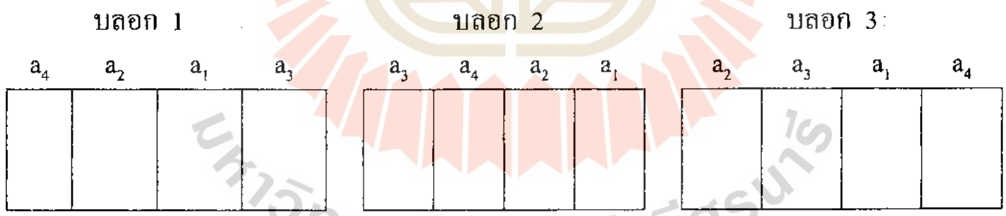
สมมุติว่าเราต้องการเปรียบพันธุ์พืช 4 พันธุ์ คือ a_1, a_2, a_3 และ a_4 โดยใช้ระยะปลูก 3 ระยะ คือ b_1, b_2 และ b_3 โดยทำการทดลอง 3 บลอกหรือ 3 ซ้ำ ก็มีวิธีการจัดแปลงทดลองและการสุ่มดังนี้

1. การสุ่มพันธุ์พืชในแนวตั้ง ถ้าแบ่งแปลงทดลองออกเป็น 3 บลอก แต่ละบลอกแบ่งเป็น 4 แปลง เพื่อปลูกพืชพันธุ์ต่าง ๆ อย่างสุ่ม ก็จะได้ดังรูป 16.7.1 ก

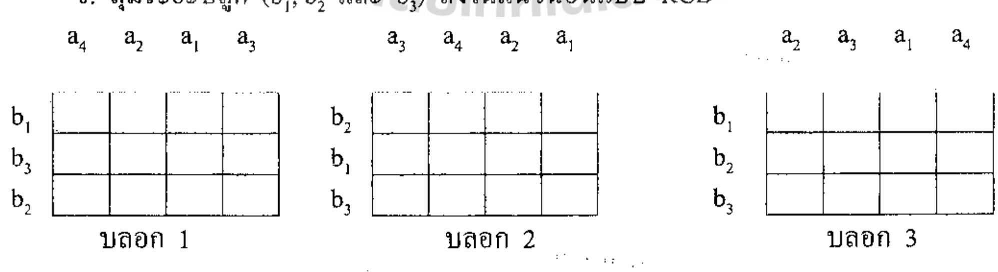
2. การสุ่มระยะปลูกในแนวนอน เมื่อสุ่มพันธุ์ลงในแนวตั้งแล้ว แบ่งแต่ละบลอกออกเป็น 3 ส่วน ในแนวนอนให้ยาวตลอดทั้งบลอก แล้วสุ่มระยะปลูกลงในแนวนอน ดังรูป 16.7.1 ข.

การจัดแปลงทดลองดังรูป เป็นการจัดแฟกเตอร์ A คือพันธุ์ในแนวตั้ง และแฟกเตอร์ B คือระยะปลูกในแนวนอนในการทดลองนี้เราจัดแฟกเตอร์ B แบบ RCB แต่อาจจัดเป็นแบบลาตินสแควร์ก็ได้ เช่นในการสุ่มของแฟกเตอร์ B ให้จัดระดับของ B ในแปลงนอนแต่ละบลอกไม่ซ้ำกัน ในกรณีเช่นนี้ จำนวนซ้ำ ต้องเท่ากับจำนวนระดับของแฟกเตอร์ B

ก. สุ่มพันธุ์พืช (a_1, a_2, a_3 และ a_4) ลงในแนวตั้ง



ข. สุ่มระยะปลูก (b_1, b_2 และ b_3) ลงในแนวนอนแบบ RCB



รูป 16.7.1 แสดงแผนการทดลองแบบสตริปพลอต

แบบจำลองทางคณิตศาสตร์

แผนการทดลองแบบสตริปพลอตมีแบบจำลองทางคณิตศาสตร์ดังนี้

$$X_{ijk} = \mu + B_i + \alpha_j + \delta_{ij} + \beta_k + \gamma_{ik} + (\alpha\beta)_{jk} + \varepsilon_{ijk}$$

และกำหนดให้

μ = ค่าเฉลี่ยของประชากร

B = ผลของบลอก

α = ผลของแฟกเตอร์ A

β = ผลของแฟกเตอร์ B

δ = ความคลาดเคลื่อนที่เกี่ยวข้องกับแฟกเตอร์ A

γ = ความคลาดเคลื่อนที่เกี่ยวข้องกับแฟกเตอร์ B

ε = ความคลาดเคลื่อนที่เกี่ยวข้องกับแฟกเตอร์ AB

$i = 1, 2, 3, \ldots n; \quad j = 1, 2, 3, \ldots, a; \quad k = 1, 2, 3, \ldots, b$

การแยกความปรวนแปร

แผนการทดลองแบบสตริปพลอตอาจแยกความปรวนแปรและ df ได้ดังตาราง 16.7.1 ซึ่งเห็นได้ว่ามีการแบ่งแหล่งของความปรวนแปรออกเป็น 3 ชุด คือ A, B และ AB แต่ละชุดมีความคลาดเคลื่อนของตัวเอง

ตาราง 16.7.1 แหล่งของความปรวนแปรในแผนการทดลองแบบสตริปพลอต

Sources	df	EMS
Blocks	n – 1 = 3	
Varieties (A)	a – 1 = 3	$\sigma_e^2 + b\sigma_k^2 + nb\sigma_a^2$
Error(a)	(n – 1)(a – 1) = 9	$\sigma_e^2 + b\sigma_k^2$
Spacings (B)	b – 1 = 2	$\sigma_e^2 + a\sigma^2 + na\sigma_b^2$
Error(b)	(n – 1)(b – 1) = 6	$\sigma_e^2 + b\sigma_y^2$
Interaction (AB)	(a – 1)(b – 1) = 6	$\sigma_e^2 + nb\sigma_{ab}^2$
Error(c)	(n – 1)(a – 1)(b – 1) = 18	σ_e^2
Total	nab – 1 = 47	

ตัวอย่าง ในการทดลองเปรียบเทียบแฟกเตอร์แฟกเตอร์ A 3 ระดับแฟกเตอร์ B 2 ระดับ โดยใช้แผนการทดลองแบบสตริปพลอต โดยให้แฟกเตอร์ A อยู่ในแนวตั้ง และแฟกเตอร์ B อยู่ในแนวนอน ดังตาราง 16.7.2

ตาราง 16.7.2 ผลการเปรียบเทียบระดับต่าง ๆ ของแฟกเตอร์ A และ B โดยใช้แผนการทดลองแบบสตริปพลอต

แฟกเตอร์		I	II	III	IV	รวม
A_1	B_1	4	6	2	4	16
	B_2	6	4	6	4	20
	รวม	10	10	8	8	36
A_2	B_1	3	7	3	8	21
	B_2	7	10	8	12	37
	รวม	10	17	11	20	58
A_3	B_1	5	6	7	8	26
	รวม	13	16	17	18	64
รวม		33	43	36	46	158

นำข้อมูลในตาราง 16.7.3 มาทำการวิเคราะห์ ดังนี้

ขั้นที่ 1 วิเคราะห์ความปรวนแปรทั้งหมด

$$CF = \frac{(\sum X_{ijk})^2}{nab} = \frac{(150)^2}{24} = 1,040.17$$

$$TSS = \sum X_{ijk}^2 - CF$$

$$= 4^2 + 6^2 + ... + 10^2 - 1,040.17 = 161.83$$

$$= \frac{\sum R^2}{ab} - CF$$

$$= \frac{33^2 + 43^2 + 36^2 + 46^2}{6} - 1,070.17 = 18.17$$

ตาราง 16.7.3 ตารางปฏิกิริยาจากข้อมูลในตาราง 16.7.2

ก. ตาราง R × A

แฟกเตอร์	I	II	III	IV	รวม
A_1	10	10	8	8	36
A_2	10	17	11	20	58
A_3	13	16	17	18	64

ข. ตาราง R × B

แฟกเตอร์	I	II	III	IV	รวม
B_1	12	19	12	20	63
B_2	21	24	24	26	95

ค. ตาราง A × B

แฟกเตอร์	A_1	A_2	A_3	รวม
B_1	16	21	26	63
B_2	20	37	38	95
	36	58	64	158

ขั้นที่ 2 วิเคราะห์แฟกเตอร์ A (แนวตั้ง)

$$SS(RA) = \frac{\sum (RA)^2}{b} - CF$$

$$= \frac{10^2 + 10^2 + ... + 18^2}{2} - 1,040.17 = 97.83$$

$$SS(A) = \frac{\sum A^2}{nb} - CF$$

$$= \frac{36^2 + 58^2 + 64^2}{8} - 1,040.17 = 54.33$$

$$Error(a) = SS(RA) - SSB - SS(A)$$

$$= 97.83 - 18.17 - 54.33 = 25.33$$

ขั้นที่ 3 วิเคราะห์แฟกเตอร์ B

$$SS(RB) = \frac{\sum (RB)^2}{a} - CF$$

$$= \frac{12^2 + 19^2 + ... + 26^2}{3} - 1,040.17 = 65.83$$

$$SS(B) = \frac{\sum B^2}{na} - CF$$

$$= \frac{63^2 + 95^2}{12} - 1{,}040.17 = 42.66$$

$$\text{Error(b)SS} = \text{SS(RB)} - \text{SSB} - \text{SS(B)}$$

$$= 65.83 - 18.17 - 42.67 = 5.00$$

ขั้นที่ 4 วิเคราะห์ปฏิกิริยา AB

$$\text{SS(AB)} = \frac{\sum (\text{AB})^2}{n} - \text{CF} - \text{SS(A)} - \text{SS(B)}$$

$$= \frac{16^2 + 20^2 + \ldots + 38^2}{4} - 1{,}040.17 - 54.33 - 42.66 = 9.34$$

$$\text{Error(c)SS} = \text{TSS} - \text{SS} \text{ อื่น ๆ ทั้งหมด}$$

$$= \text{TSS} - \text{SSB} - \text{SS(A)} - \text{SS}(E_a) - \text{SS(B)} - \text{SS}(E_b) - \text{SS(AB)}$$

$$= 161.83 - 18.17 - 54.33 - 25.33 - 42.66 - 5.00 - 9.34 = 7.00$$

ขั้นที่ 5 นำผลการวิเคราะห์ลงตารางวิเคราะห์วาเรียนซ์ดังตาราง 16.7.4 แล้วคำนวณ MS และค่า F ตามปกติ

ตาราง 16.7.4 ผลการวิเคราะห์วาเรียนซ์ของข้อมูลในตาราง 16.7.2

Sources	df	SS	MS	F
Blocks	3	18.17	6.05	
A	2	54.33	27.17	6.43*
Error (a)	6	25.33	4.22	
B	1	42.66	42.66	25.54*
Error (b)	3	5.00	1.67	
AB	2	9.34	4.67	3.94^{ns}
Error (c)	6	7.00	1.17	
Total	23	161.83		

จากตาราง 16.7.4 การทดสอบ A, B และ AB กระทำโดยใช้ MSE ที่เกี่ยวข้องโดยตรงดังนี้

$$F(A) = \frac{\text{MSA}}{\text{MSE}_a} = \frac{27.17}{4.22} = 6.43$$

$$F(B) = \frac{MSB}{MSE_b} = \frac{42.66}{1.67} = 25.54$$

$$F(AB) = \frac{MS(AB)}{MSE_c} = \frac{4.67}{1.17} = 3.44$$

แล้วเปิดตาราง F ต่อไป อย่างไรดี การทดสอบแฟกเตอร์ B ไม่อาจกระทำได้เพราะ Error (c) มี df ต่ำเกินไป เมื่อจะเปรียบเทียบค่าเฉลี่ยก็ใช้ค่าความคลาดเคลื่อนมาตรฐาน ดังแสดงในตาราง 16.7.5 เช่นค่า LSD ของความแตกต่างระหว่าง A_1 และ A_2

$$LSD = (t_{.05,6})(\sqrt{2(MSE_2 / nb)}) = (2.447)(\sqrt{2(4.22)/8}) = 2.54$$

ตาราง 16.7.5 ความคลาดเคลื่อนมาตราฐานของความแตกต่างในแผนการทดลองแบบสตริปพลอต

เลขที่	ชุดเปรียบเทียบ	$s_{\bar{d}}$
1	ระหว่างแฟกเตอร์ A	$\sqrt{\frac{2E_a}{nb}}$
2	ระหว่างแฟกเตอร์ B	$\sqrt{\frac{2E_b}{na}}$
3	แฟกเตอร์ A ในแฟกเตอร์ B ระดับเดียวกัน	$\sqrt{\frac{2[(b-1)E_c + E_a]}{nb}}$
4	แฟกเตอร์ B ในแฟกเตอร์ A ระดับเดียวกัน	$\sqrt{\frac{2[(a-1)E_c + E_b]}{na}}$

16.8 แผนการทดลองแบบแบ่งกลุ่มภายในบลอก

ในการทดลองที่ทรีตเมนต์มีความแตกต่างกัน เช่น ในการเปรียบเทียบพันธุ์ถั่วเหลือง บางพันธุ์ต้นสูง บางพันธุ์ต้นเตี้ย การเปรียบเทียบระดับโปรตีนในอาหารสุกร มีทั้งโปรตีนจากพืชและโปรตีนจากสัตว์ ในการทดลองเช่นนี้อาจจะจัดเป็นบลอกย่อย ๆ โดยให้ปัญหาทดลองเหมือนกันหรือคล้ายกันอยู่ในกลุ่มเดียวกัน การเปรียบเทียบจะเกิดภายในบลอกย่อย ๆ แผนการทดลองเช่นนี้เรียกว่า การทดลองแบบแบ่งกลุ่มภายในบลอก[9]

การจัดแปลงทดลอง

สมมุติว่าทำการทดลองเปรียบเทียบพันธุ์ถั่วเหลือง 9 พันธุ์ โดยทำการทดลอง 3 บลอก ถั่วเหลือง 9 พันธุ์มีต้นสูง 3 พันธุ์ ปานกลาง 3 พันธุ์ และเตี้ย 3 พันธุ์ คือจัดเป็น 3 กลุ่ม ดังนี้

ถั่วเหลืองพันธุ์	A	B	C	D	E	F	G	H	I
ความสูง	สูง	สูง	สูง	กลาง	กลาง	กลาง	เตี้ย	เตี้ย	เตี้ย
กลุ่ม		T			M			S	

ดังนั้นในแต่ละบลอกปลูก 9 แปลง แบ่งเป็น 3 กลุ่ม คือ T, M และ S ก็ทำการสุ่มกลุ่มลงไปในบลอก ต่อจากนั้นก็สุ่มพันธุ์ลงไปในแต่ละกลุ่มดังรูป 16.8.1

ก. การสุ่มกลุ่มของทรีตเมนต์ภายในบลอกเป็นกลุ่ม

I	II	III
T	S	M
S	M	T
M	T	S

ข. การสุ่มทรีตเมนต์ภายในกลุ่มแต่ละกลุ่ม

TB	SI	MD
TA	SG	ME
TC	SH	MF
SG	MF	TA
SI	MD	TC
SH	ME	TB
ME	TC	SH
MF	TB	SG
MD	TA	SI

รูป 16.8.1 รูปแปลงทดลองเปรียบเทียบพันธุ์ถั่วเหลืองที่มีความสูงของต้นต่างกัน

ตัวอย่าง ในการเปรียบเทียบพันธุ์ถั่วลิสง 12 พันธุ์ ทำการทดลอง 3 บลอก แบ่งพืชออก 3 กลุ่ม คือกลุ่มเวอร์จิเนีย วาเลนเซีย และสแปนนิช ผลการทดลองแสดงดังตาราง 16.8.1

แบบจำลองทางคณิตศาสตร์

$$X_{ijk} \quad = \quad \mu + B_i + G_j + \delta_{ij} + T_k + \varepsilon_{ijk}$$

ทั้งนี้กำหนดให้ B, G, T เป็นผลของบลอก ของกลุ่มทรีตเมนต์ และของทรีตเมนต์ แต่ละกลุ่ม δ และ ε เป็นความคลาดเคลื่อนที่เกิดระหว่างกลุ่ม และภายในกลุ่มตามลำดับ

J 4	K 2	L 3	I 3
D 7	A 6	C 2	B 4
H 7	F 4	G 7	E 3

I

A 7	B 6	D 8	C 3
F 5	G 8	H 10	E 4
L 5	J 6	K 4	I 4

II

G 8	F 7	H 10	E 4
J 5	K 3	I 3	L 4
B 7	C 3	D 8	A 8

III

กลุ่มที่ 1 : A, B, C, D กลุ่มที่ 2 : E, F, G, H, กลุ่มที่ 3 : I, J, K, L

รูปที่ 16.8.2 แปลงทดลองเปรียบเทียบถั่วลิสง 12 พันธุ์แบ่งออกเป็น 3 กลุ่ม

ตาราง 16.8.1 ผลจากการเปรียบเทียบถั่วลิสง 12 พันธุ์ 3 กลุ่ม จำนวน 3 บลอก

กลุ่ม	พันธุ์	I	II	III	รวม
G_1	A	6	7	8	21
	B	4	6	7	17
	C	2	3	3	8
	D	7	8	8	23
	รวม	19	24	26	69
G_2	E	3	4	4	11
	F	4	5	7	16
	G	7	8	8	23
	H	7	10	10	27
	รวม	21	27	29	77
G_3	I	3	4	3	10
	J	4	6	5	15
	K	2	4	3	9
	L	3	5	4	12
	รวม	12	19	15	46
	รวม	52	70	70	192

ทั้งนี้

i = 1, 2,...., n เมื่อ n คือจำนวนบลอก
j = 1, 2,...., G เมื่อ G คือจำนวนกลุ่ม
k = 1, 2,...., g เมื่อ g คือจำนวนสมาชิกในแต่ละกลุ่ม

ตาราง 16.8.2 ตารางวิเคราะห์เรียนซ์ของการทดลองแบบแยกกลุ่มภายในบลอก

Sources	df
Blocks	$n-1$
Groups	$G-1$
Error (a)	$(n-1)(G-1)$
Within Group	
Group 1	g_1-1
Group 2	g_2-1
Group 3	g_3-1
Error (b)	$G(n-1)(g-1)$
Total	$nGg-1$

วิเคราะห์วาเรียนซ์

วิเคราะห์ผลในแปลงใหญ่

$$\text{CF} = \frac{(\sum X_{ijk})^2}{nt} = \frac{(192)^2}{(3)(12)} = 1024.00$$

$$\text{TSS} = X_{ijk}^2 - \text{CF}$$

$$= 6^2 + 7^2 + \ldots + 4^2 - 1{,}024.00 = 172.00$$

ขั้นที่ 1 วิเคราะห์ปัญหาในกลุ่ม

$$\text{Whole plot SS} = \frac{\sum G_{ij}^2}{g} - \text{CF}$$

$$= \frac{19^2 + 24^2 + \ldots 15^2}{4} - 1{,}024.00 = 64.50$$

$$\text{Block SS (SSB)} = \frac{\sum B^2}{Gg} - \text{CF}$$

$$= \frac{(52^2 + 70^2 + 70^2)}{12} - 1{,}024.00 = 10.00$$

$$\text{Group SS(SSG)} = \frac{\sum G_j^2}{ng} - CF$$

$$= \frac{69^2 + 77^2 + 46^2}{12} - 1{,}024.00 = 43.17$$

$$\text{Error(a)SS} = \text{Whole plot SS} - \text{SSB} - \text{SSG}$$

$$= 64.50 - 10.00 - 43.17 = 11.33$$

ขั้นที่ 2 วิเคราะห์ในแปลงเล็ก

$$\text{Treatment Group SS} = \frac{\sum g_k^2}{n} - CF_g$$

$$\text{Group 1 SS} = \frac{21^2 + 17^2 + 8^2 + 23^2}{3} - \frac{69^2}{12} = 44.25$$

$$\text{Group 2 SS} = \frac{11^2 + 16^2 + 23^2 + 27^2}{3} - \frac{77^2}{12} = 50.92$$

$$\text{Group 3 SS} = \frac{10^2 + 15^2 + 9^2 + 12^2}{3} - \frac{46^2}{12} = 7.00$$

Error(b) SS = TSS – SS อื่นๆ ทั้งหมด

$$= 172.00 - (10.00 + 43.17 + 11.33 + 44.25 + 50.92 + 7.00)$$

$$= 5.33$$

ขั้นที่ 3 หาค่ามีนสแควร์และทดสอบโดยวิธี F-test

นำค่า SS ลงตารางดังตาราง 16.8.3 แล้วหาค่ามีนสแควร์ต่าง ๆ และทดสอบความแตกต่างต่อไป การทดสอบระหว่างกลุ่มและบลอกใช้ MSE_a เป็นตัวหาร ส่วนการทดสอบพันธุ์ภายในกลุ่มใช้ MSE_b เป็นตัวหาร แล้วเปิดตาราง F ด้วยวิธีการที่อธิบายมาแล้ว

ตาราง 16.8.3 ผลการวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 16.8.1

Sources	df	SS	MS	F
Blocks	2	10.00	5.00	1.77
Groups	2	43.13	21.56	7.62*
Error (a)	4	11.33	2.83	
Group 1	3	44.25	14.75	49.17**
Group 2	3	50.92	16.67	55.57**
Group 3	3	7.00	2.33	7.77**
Error (b)	18	5.33	0.30	
Total	35	172.00		

หาค่าสัมประสิทธิ์ของการปรวนแปร

$$CV(a) = \frac{\sqrt{E_a}}{\text{Mean}} \times 100 = \frac{\sqrt{2.83}}{5.33} \times 100 = 31.84\%$$

$$CV(b) = \frac{\sqrt{E_b}}{\text{Mean}} \times 100 = \frac{\sqrt{0.30}}{5.33} \times 100 = 10.27\%$$

16.9 แบบฝึกหัด

1. ในการทดลองโดยใช้แผนการทดลองแบบสปลิต-พลอต สปลิต-สปลิต-พลอต และสตริป-พลอตนั้น ให้พิจารณาการใช้แปลงใหญ่แปลงย่อยตามความเหมาะสม การทดลองดังต่อไปนี้ ควรใช้แผนการทดลองชนิดใด และควรใช้ปัญหาอะไรเป็นแปลงใหญ่ แปลงย่อย เพราะอะไร

ก. ทดสอบผลของระดับการไถลึกระดับต่าง ๆ ต่อพันธุ์พืช

ข. การเปรียบเทียบผลของปุ๋ย NPK และธาตุอาหารโบรอน

ค. การทดลองการตัดหญ้าอาหารสัตว์ในระดับความสูงต่าง ๆ กัน

ง. การทดสอบการสึกหรอของยางรถยนต์ที่ผลิตโดย 3 บริษัทในสภาพน้ำหนักบรรทุกต่างกัน

2. ในการทดลองเปรียบเทียบผลของระยะปลูกระหว่างแถว 3 ระยะ และระยะระหว่างหลุม 3 ระยะ ต่อผลผลิตของถั่วลิสงพันธุ์ไทนาน 9 โดยใช้แบบการทดลองแบบสปลิตพลอตให้ระยะระหว่างแถวเป็นแปลงใหญ่ และระยะระหว่างหลุมเป็นแปลงย่อย ทำการทดลอง 3 บลอกให้ผลดังนี้

	บลอก 1			บลอก 2			บลอก 3		
แถว(ซม.)	30	50	70	30	50	70	30	50	70
หลุม (ซม.)									
10	225	219	215	249	243	278	244	275	[illegible]
20	230	228	237	267	264	287	282	277	298
30	236	240	267	260	272	291	284	268	275

ก. จงแสดงแบบจำลองทางคณิตศาสตร์ของการทดลองนี้

ข. จงวิเคราะห์วาเรียนซ์

ค. จงคำนวณค่า LSD เพื่อเปรียบเทียบระหว่างระยะระหว่างแถว ระหว่างหลุม

ง. ในบลอกที่ 3 ถ้าแปลงย่อยที่ X(20,50) สูญหาย จงคำนวณค่าดังกล่าว

3. จากการเปรียบเทียบผลผลิตของข้าวโพดลูกผสม ทำการปลูกใน 3 วันปลูก โดยใช้ระยะปลูกแตกต่างกัน 5 ระยะปลูก ได้ผลผลิต (ตัน/เฮกตาร์)

		บลอก			
วันปลูก	ระยะปลูก (ซม.)	I	II	III	IV
มิถุนายน	50 x 50	0.9	1.5	5.9	3.4
	50 x 25	1.1	5.1	7.2	1.8
	100 x 25	1.5	2.9	5.0	1.7
	100 x 12.5	1.6	4.4	7.0	4.5
กรกฎาคม	50 x 50	5.4	7.1	6.5	4.2
	50 x 25	4.8	6.6	6.1	5.2
	100 x 25	5.6	5.1	5.0	4.5
	100 x 12.5	6.2	6.8	5.0	4.6
สิงหาคม	50 x 50	5.7	4.6	5.1	6.2
	50 x 25	5.0	5.9	6.4	4.9
	100 x 25	4.8	4.4	5.2	5.5
	100 x 12.5	5.3	4.6	5.1	5.6

จงวิเคราะห์วาเรียนซ์ข้อมูลนี้ โดยใช้วิธีการแผนการทดลองแบบสปลิตพลอต

ก. คำนวณหาสัมประสิทธิ์ของความแปรปรวนของวันปลูก และอัตราปลูก

ข. คำนวณ LSD ของวันปลูก อัตราปลูก ของอัตราปลูกสองชนิดในวันปลูกเดียวกัน

ค. จงอธิบายผลการทดลอง ถ้าหากว่าพันธุ์และวันปลูกมีความแตกต่างถึงระดับนัยสำคัญทางสถิติ

4. จงวิเคราะห์ข้อมูลในข้อ 2. โดยใช้วิธีการของแผนการทดลองแบบสตริปพลอต

5. ในการทดลองเปรียบเทียบผลของความชื้น 2 ระดับ อัตราปลูก 3 ระดับ และปุ๋ย 3 ระดับ ต่อขนาดของเมล็ดของถั่วเขียวพันธุ์ชัยนาท 36 ซึ่งได้ผลการทดลองดังนี้ (น้ำ : 0 = ไม่ให้น้ำ, 1= ให้น้ำเมื่อขาดน้ำฝนเกิน 7 วัน , อัตราปลูก : 1= 50 x 5 ซม. , 2 = 50 x 10 ซม., 3=50x15 ซม. , ปุ๋ย : 1 = NPK 20 กก./ไร่ 2 = NPK 30 กก./ไร่ , 3 = NPK 40 กก./ไร่)

			บลอก			
น้ำ (A)	อัตราปลูก (B)	ปุ๋ย (C)	1	2	3	4
0	1	1	6.0	5.5	5.5	5.7
		2	6.3	5.3	5.9	5.2
		3	7.1	6.3	5.9	5.9
	2	1	6.1	6.5	7.5	5.3
		2	5.9	6.5	6.9	5.7
		3	6.1	7.1	6.1	5.8
	3	1	5.4	4.9	5.5	5.7
		2	6.1	5.4	5.2	5.9
		3	6.2	4.9	6.3	5.5
1	1	1	5.3	6.8	6.0	6.2
		2	5.8	7.3	6.9	7.6
		3	6.7	7.0	7.6	7.6
	2	1	7.5	6.6	6.0	7.3
		2	7.3	6.3	7.9	7.5
		3	7.9	7.5	7.5	7.4
	3	1	7.2	7.3	7.0	6.8
		2	7.5	7.5	7.3	7.3
		3	7.8	7.6	7.5	7.6

ก. จงแสดงแบบจำลองทางคณิตศาสตร์

ข. จงวิเคราะห์วาเรียนซ์และทดสอบทางสถิติ

ค. จงวิเคราะห์ว่า ระยะปลูกและปุ๋ยให้ผลแบบ linear และ quadratic

ง. จงคำนวณหาค่า LSD ในการเปรียบเทียบดังต่อไปนี้ : ระหว่าง a_0 และ a_1, b_1 และ b_2, c_1 และ c_2, a_0b_1 และ a_1b_1

คำในบท

(1) main plot, (2) whole plot, (3) sub-plot, (4) trend comparison, (5) polynomial, (6) split-split plot design, (7) sub-sub plot, (8) strip-plot or split-block design (9) group balanced block design.

บทที่ 17
การทดลองซ้ำหลายครั้ง

17.1 คำนำ

ในการทดสอบเพื่อหาคำตอบในทรีตเมนต์หรือปัญหาบางอย่าง อาจต้องต้องทำการทดลองซ้ำหลายครั้งจึงสามารถเลือกทรีตเมนต์ที่ดีหรือสรุปผลได้ถูกต้องและแน่นอน ในการที่ทางการแพทย์จะแนะนำยาชนิดหนึ่งเพื่อรักษาโรค ต้องทำการทดสอบกับคนไข้เป็นจำนวนมากในหลายประเทศ ในทางเกษตร การที่จะแนะนำสูตรปุ๋ย พันธุ์พืช พันธุ์สัตว์ ฯลฯ เพื่อให้เกษตรกรนำไปใช้ประโยชน์อย่างกว้างขวาง ก็ต้องมีการทดลองซ้ำหลายครั้ง การพัฒนาพันธุ์พืชบางพันธุ์ของกรมวิชาการเกษตรพบว่าต้องทำการทดลองซ้ำ ๆ เป็นเวลาหลายปี หลายท้องที่รวมแล้วไม่น้อยกว่า 30 ครั้ง เหตุที่ต้องมีการทดลองซ้ำ ๆ เช่นนี้ เพราะเราต้องการให้สภาพแวดล้อม หรือแหล่งทดลองอันหลากหลาย เข้ามามีส่วนช่วยในการคัดเลือก เช่นนี้ทำให้ผลการทดสอบถูกต้องและเชื่อถือได้

ในสภาวะที่กล่าวมาแล้วดังข้างบน การที่จะให้ได้ข้อสรุปในปัญหาหนึ่ง ๆ ในการที่จะเลือกเทคโนโลยีหนึ่ง ๆ หรือการคัดเลือกพันธุ์หรือพืชพันธุ์ใด ต้องมีการทดลองซ้ำ ๆ หลายครั้ง ผลที่ปรากฏออกมามีหลายรูปแบบ เช่น ในการทดสอบพันธุ์พืช พันธุ์หนึ่งอาจดีกว่าอีกพันธุ์หนึ่งในทุก ๆ การทดลอง หรืออาจพบว่าในท้องที่หนึ่งพันธุ์ ก ให้ผลผลิตสูงกว่า ข แต่ในอีกท้องที่ผลเป็นแบบตรงกันข้าม ดังนั้น การทดลองซ้ำ ๆ คือการทดสอบการปรับตัวของพันธุ์พืชนั่นเอง

17.2 การทดลองซ้ำและการวิเคราะห์

ในการทดลองทางเกษตร โดยเฉพาะอย่างยิ่งเกี่ยวกับพืช ย่อมมีความจำเป็นต้องทำการทดลองหลายครั้ง เพราะเราต้องทดสอบพืช หรือปุ๋ย หรืออย่างอื่นเกี่ยวกับพืช กระทำในสภาพแวดล้อมที่ปรวนแปร การที่ให้ได้คำตอบที่ถูกต้อง ต้องทดลองซ้ำมาก ๆ แล้วใช้ผลเฉลี่ยในการสรุปข้อมูล ในการทดลองอาจมีสภาพแวดล้อมได้ 2 ชนิด คือ สถานที่และเวลา สำหรับสถานที่เราพอกำหนดได้ ทราบประวัติได้ ฯลฯ ดังนั้นมักให้เป็นปัจจัยคงที่ ส่วนเวลาเรากำหนดไม่ได้ เราไม่รู้ว่าปีต่อไปสภาพอากาศจะเป็นอย่างไร จึงมักกำหนดให้เป็นปัจจัยสุ่ม จึงอาจสรุปได้ดังนี้

สถานที่ทดลอง สถานที่ทดลองมักเป็นตัวแทนของสภาพแวดล้อมที่จะปลูกพืชหรือใช้ทรีตเมนต์นั้น ถ้าสถานที่มีความแตกต่างกันมากเท่าใดก็ควรใช้สถานที่มากขึ้นเท่านั้น อาจแยกสถานที่ออกเป็นกลุ่ม ๆ ตามความเหมือน

เวลาในการทดลอง ในการทดลองในแต่ละสถานที่อาจทำการทดลองหลายครั้ง หรือหลายฤดู เช่น ต้นฤดูฝน ปลายฤดูฝน และฤดูแล้งหรือทดลองหลายปี เวลาในการทดลองเหล่านี้เราไม่อาจกำหนดได้ว่าแต่ละครั้งฝนตกมากหรือตกน้อย ดังนั้นต้องให้มีจำนวนเวลามากพอที่จะเป็นตัวแทนของสภาพแวดล้อมในแต่ละท้องที่

การซ้ำทดลองทั้งในเรื่องสถานที่และเวลาหรือทั้งสองอย่าง มีวัตถุประสงค์เพื่อผลการทดลองมาวิเคราะห์ร่วมกัน และหาคำตอบคำตอบเดียว ต้องใช้วิธีการทดลองที่เหมือนกัน คือ ใช้ทรีตเมนต์ชุดเดียวกัน ใช้แผนการทดลองเดียวกัน ขนาดของแปลงปลูกเท่ากัน จำนวนซ้ำเทคนิคการปลูก การใช้ปุ๋ย สารเคมี ดูแลรักษา ฯลฯ เหมือน ๆ กัน ให้เหมือนว่าเป็นการทดลองเดียวกัน อย่างไรก็ดีแต่ละการทดลองต้องทำการสุ่มวางทรีตเมนต์ใหม่ทุกครั้ง

การวิเคราะห์ร่วม

การวิเคราะห์ร่วม[1] คือการที่นำการทดลองในเรื่องเดียวกันมาวิเคราะห์ร่วมกัน คล้ายกับการทดลองแบบใช้แปลงย่อยที่กล่าวมาแล้วในบทที่ 16 นั่นเอง ข้อกำหนดที่สำคัญในการวิเคราะห์ร่วม คือ

(1) เป็นการทดลองในชุดเดียวกัน คือใช้ทรีตเมนต์ แผนการทดลอง และวิธีการปฏิบัติเหมือนกัน

(2) วาเรียนซ์ของการทดลองต่าง ๆ เท่ากัน[2] สามารถทำการทดสอบว่าวาเรียนซ์ของการทดลองต่าง ๆ เท่ากันโดยใช้วิธีการทดสอบของ Bartlett (1937) ดังได้อธิบายไว้ในตอน 7.4 หรือถ้ามีเพียง 2 การทดลองก็อาจใช้ค่า F ดังแสดงในสมการ (4-6) แล้วเปิดตาราง F ที่ df ของตัวตั้งและตัวหาร ความแตกต่างของวาเรียนซ์อาจพิจารณาได้ง่าย ๆ คือ ถ้าหากว่าวาเรียนซ์ที่สูงกว่ามากกว่าวาเรียนซ์ต่ำไม่เกิน 3 เท่า ก็ถือว่าไม่แตกต่างกันในทางสถิติ

(3) ในการทดสอบทรีตเมนต์โดยใช้วาเรียนซ์ของปฏิกิริยาสำหรับทรีตเมนต์และสภาพแวดล้อม $(F = MSTr / MSTr \times Place)$ ต้องมีการกำหนดว่าปฏิกิริยาของทรีตเมนต์กับสภาพแวดล้อมต้องเท่า ๆ กัน มิใช่ว่าบางทรีตเมนต์มีความแตกต่างในสภาพแวดล้อมต่าง ๆ มาก แต่บางทรีตเมนต์แตกต่างกันน้อย

17.3 การทดลองหลายครั้งแนวเดียว

การทดลองหลายครั้งแนวเดียวหมายถึงการทดลองซ้ำ ๆ แต่เป็นการซ้ำในสภาพแวดล้อมชนิดเดียวกัน เช่น การเปรียบเทียบพันธุ์ในหลายท้องที่ในปีเดียวกัน หรือเปรียบเทียบพันธุ์หรือในท้องที่เดียวกันแต่หลายฤดูหรือหลายปี ในการทดลองนั้นใช้แผนการทดลองอะไรก็ได้ เช่น ใช้ CRD, RCB, ลาตินสแควร์ หรือสปลิตพลอต แต่ต้องใช้แผนการทดลองเดียวกันทุกครั้ง ใช้วิธีการทดลองเหมือนกัน ถ้าเป็นการทดลองทางพืชก็มีขนาดของแปลง ระยะปลูก การดูแลรักษาเหมือน ๆ กัน เพราะการปฏิบัติที่ต่างกัน อาจทำให้ผลของทรีตเมนต์แตกต่างออกไป การทดลองแบบนี้มีวัตถุประสงค์จะดู

ความดีเด่นของทรีตเมนท์ และปฏิกิริยาระหว่างทรีตเมนต์และสภาพแวดล้อม แบบจำลองทางคณิตศาสตร์ของการทดลองหลายครั้งมีดังนี้

$$X_{ijk} = \mu + E_i + B_{(i)j} + T_k + (ET)_{ik} + \varepsilon_{ijk} \qquad ...(17\text{-}1)$$

กำหนดให้

μ ค่าเฉลี่ยของประชากร

E,T ผลของสภาพแวดล้อมและทรีตเมนต์

B ผลของบลอกในสภาพแวดล้อมที่ i

(ET) คือปฏิกิริยาระหว่างสภาพแวดล้อมและทรีตเมนต์

$i = 1, 2, \ldots, e$ เมื่อ e เป็นจำนวนสภาพแวดล้อม

$j = 1, 2, \ldots, n$ เมื่อ n เป็นจำนวนบลอก

$k = 1, 2, \ldots, t$ เมื่อ t เป็นจำนวนทรีตเมนต์

ซึ่งสามารถแสดงแหล่งของความปรวนแปรและ EMS ได้ดังตาราง 17.3.1

ตาราง 17.3.1 แหล่งของความปรวนแปรและ EMS ของการทดลองหลายครั้งในแนวเดียว (Model II)

Sources	df	
Environments(E)	$e-1$	$\sigma^2 + t\sigma^2_{b/e} + nt\sigma^2_e$
Block within env.	$e(n-1)$	$\sigma^2 + t\sigma^2_{b/e}$
Treatments	$t-1$	$\sigma^2 + n\sigma^2_{et} + ne\sigma^2_t$
Env. × Treatment	$(e-1)(t-1)$	$\sigma^2 + n\sigma^2_{et}$
Pooled error	$e(n-1)(t-1)$	σ^2

ก. การทดลองในสถานที่ต่างกัน

ตัวอย่าง ในการทดลองเปรียบเทียบพันธุ์พืช 5 พันธุ์ ทดลองใน 3 ท้องที่ แต่ละการทดลองใช้แผนการทดลองแบบ RCB จำนวน 3 ซ้ำ ผลการทดลองแสดงไว้ในตาราง 17.3.2

ขั้นตอนในการวิเคราะห์

1. วิเคราะห์แต่ละการทดลอง แต่ละการทดลองทำการวิเคราะห์วาเรียนซ์ดังวิธีที่ใช้ในบทที่ 10 ผลการวิเคราะห์แสดงไว้ในตาราง 17.3.3

2. ทำการทดสอบว่าวาเรียนซ์เท่ากัน ทำการทดสอบโดยใช้วิธีการของ Bartlett ดังแสดงในตอน 7.4 เพื่อแสดงว่าวาเรียนซ์หรือค่ามีนสแควร์ error ของ 3 การทดลองนี้เท่ากัน ซึ่งผลการทดสอบต้องพบว่าวาเรียนซ์เท่ากัน จึงดำเนินการวิเคราะห์ทั้ง 3 การทดลองร่วมกันในครั้งเดียว เราเรียกสั้น ๆ ว่า**การวิเคราะห์ร่วม**[1]

3. การวิเคราะห์ร่วม เมื่อผลการทดสอบพบว่าวาเรียนซ์ของการทดลองต่าง ๆ เท่ากัน หรือถ้าหากไม่ทำการทดสอบก็สามารถตั้งเป็นข้อกำหนดว่าเท่ากันแล้ว ก็สามารถดำเนินการวิเคราะห์ร่วมเพื่อทดสอบพันธุ์พืชในทุกท้องที่ปลูกในคราวเดียวกัน โดยใช้แบบจำลองทางคณิตศาสตร์ (17-1)

วิธีการวิเคราะห์

$$\text{Correction factor (CF)} = \frac{(\sum X_{ijk})^2}{lnt} = \frac{(498)^2}{(3)(3)(5)} = 5{,}511.20$$

ทั้งนี้กำหนดให้

$i = 1, 2, \ldots, l$ (l = จำนวนท้องที่ทดลอง)

$j = 1, 2, \ldots, n$ (n = จำนวนบลอก)

$k = 1, 2, \ldots, t$ (t = จำนวนพันธุ์พืช)

$$\text{Total SS(TSS)} = \sum X_{ijk}^2 - CF$$
$$= 6^2 + 4^2 + \ldots + 9^2 - 5{,}511.20 = 380.80$$

ขั้นที่ 1 วิเคราะห์ท้องที่ทดลอง

$$SS(L) = \frac{\sum L^2}{nt} - CF$$
$$= \frac{135^2 + 208^2 + 155^2}{(3)(5)} - 5{,}511.20 = 189.73$$

ตาราง 17.3.2 ผลการเปรียบเทียบพันธุ์พืช 5 พันธุ์ใน 3 ท้องที่ โดยใช้แผนการทดลองแบบ RCB จำนวน 3 บลอก

พันธุ์	ท้องที่ 1			รวม	ท้องที่ 2			รวม	ท้องที่ 3			รวม	รวมยอด
	I	II	III		I	II	III		I	II	III		
ก	6	4	9	19	15	13	13	41	11	13	9	33	93
ข	14	11	12	37	13	16	18	47	10	11	14	35	119
ค	10	8	9	27	15	13	14	42	8	10	8	26	95
ง	11	8	10	29	12	13	12	37	8	9	9	26	92
จ	6	8	9	23	12	14	15	41	12	14	9	35	99
	47	39	49	135	67	69	72	208	49	57	49	155	498

ตาราง 17.3.3 ผลการวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 17.3.2

Sources	df	ท้องที่ 1		ท้องที่ 2		ท้องที่ 3	
		SS	MS	SS	MS	SS	MS
Blocks	2	11.20	5.60	2.53	1.27	8.53	4.26
Varieties	4	61.33	15.33	17.06	4.26	28.67	7.17
Error	8	17.47	2.18	20.14	2.52	24.13	3.01
Total	14	90.00		39.73		61.33	

$$SS\,(B/L) = \frac{\sum B_{(i)j}^2}{lnt} - CF - SS(Location)$$

$$= \frac{(47^2+39^2+49^2)+(67^2+69^2+72^2)+(49^2+57^2+49^2)}{5} - 5{,}511.20 - 189.73$$

$$= 22.26$$

ซึ่งอาจหาได้โดยนำ $SSB(L_1) + SSB(L_2) + SSB(L_3)$ จากตาราง 17.3.3 ซึ่งเท่ากับ $11.20 + 2.53 + 8.53 = 22.26$

ขั้นที่ 2 วิเคราะห์พันธุ์และปฏิกิริยา

$$SS(Varieties) = \frac{\sum T_k^2}{ln} - CF$$

$$= \frac{93^2+119^2+\ldots+99^2}{(3)(3)} - 5{,}511.20$$

$$= 55.47$$

$$SS(Locations \times Varieties) = SS(LV) = \frac{\sum (LV)^2}{n} - CF - SS(L) - SS(V)$$

$$= \frac{(19^2+\ldots+23^2)+(41^2+\ldots+41^2)+(33^2+\ldots+35^2)}{3} - 5{,}511.20$$

$$= 189.73 - 55.47$$

$$= 51.60$$

$$SS\,(Pooled\ error) = TSS - SS(B/L) - SS(L) - SS(V) - SS(LV)$$

$$= 380.80 - 189.73 - 22.26 - 55.47 - 51.60$$

$$= 61.74$$

ซึ่งอาจหาได้จากโดยการรวม $SSE(L_1) + SSE(L_2) + SSE(L_3)$ จากตาราง 17.3.3 ซึ่งเท่ากับ $17.47 + 20.14 + 24.13 = 61.74$

ขั้นที่ 3 นำผลการวิเคราะห์ลงตารางวิเคราะห์วาเรียนซ์ ดังตาราง 17.3.4 แล้วคำนวณค่ามีนสแควร์ต่าง ๆ ตามปกติ คือแต่ละค่าได้จาก SS หารด้วย df ในบรรทัดเดียวกัน และดำเนินการทดสอบผลของท้องที่และพันธุ์ การทดสอบโดยวิธี F-test การทดสอบท้องที่นั้น ใช้มีนสแควร์รวมของบลอกเป็นตัวหารดังนี้

$$F(\text{Locations}) = \frac{94.87}{3.71} = 25.57^{**}$$

ตาราง 17.3.4 ผลการวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 17.3.2

Sources	df	SS	MS	F
Locations (L)	2	189.73	94.87	25.57**
Blocks within location	6	22.26	3.71	
Varieties (V)	4	55.47	13.87	5.13**
L x V	8	51.60	6.45	2.51*
Pooled error	24	61.74	2.57	
Total	44	380.80		

ซึ่งแสดงว่าผลผลิตจากแต่ละท้องที่แตกต่างกันในทางสถิติ ในการทดสอบผลของปฏิกิริยาระหว่างพันธุ์และสถานที่ทดลอง ใช้ MS error รวมเป็นตัวหาร ดังนี้

$$F(\text{Varieties x Locations}) = \frac{6.45}{2.57} = 2.51^{**}$$

ในการทดสอบนี้มีสมมุติฐานว่า ความแตกต่างระหว่างพันธ์เหมือนกันในทุกท้องที่ เนื่องจากปฏิกิริยาระหว่างพันธุ์และสถานที่มีความสำคัญ ถ้าหากว่าปัญหาเป็นปัจจัยสุ่ม การทดสอบพันธุ์ใช้ MS(LV) เป็นตัวหารดังนี้

$$F(\text{Varieties}) = \frac{13.87}{6.45} = 2.15^{*}$$

แต่ถ้าเป็นปัจจัยคงที่

$$F(\text{Varieties}) = \frac{13.87}{2.57} = 5.13^{**}$$

ในการทดสอบสมมุติฐานว่าพันธุ์ไม่แตกต่างกันนี้ ตามข้อกำหนดว่าเป็นปัจจัยสุ่ม ในบางครั้งคิดว่าปฏิกิริยาอาจไม่มีอยู่จริง ในกรณีเช่นนั้นอาจนำค่า sum of squares ของปฏิกิริยาระหว่างพันธุ์กับท้องที่และของ error มาบวกกัน หามีนสแควร์แล้วใช้เป็นตัวหาร

ตาราง 17.3.5 ค่าเฉลี่ยของพันธุ์ในท้องที่ต่าง ๆ

พันธุ์	ท้องที่ 1	ท้องที่ 2	ท้องที่ 3	เฉลี่ย
ก	6.33	13.67	11.00	10.33
ข	12.33	15.67	11.67	13.22
ค	9.00	14.00	8.67	10.57
ง	9.67	12.33	8.67	10.22
จ	7.67	13.67	11.67	10.90

ในการทดลองนี้พบว่าปฏิกิริยาระหว่างพันธุ์และท้องที่มีความสำคัญ ถ้าหากพิจารณาจากค่าเฉลี่ยในตาราง 17.3.5 พบว่าไม่มีปฏิกิริยาระหว่างพันธุ์กับท้องที่ที่ 1 และ 2 เพราะทุกพันธุ์ให้ผลไปในทางเดียวกัน คือท้องที่ที่ 2 สูงกว่าท้องที่ที่ 1 แต่มีปฏิกิริยาระหว่างพันธุ์กับท้องที่ที่ 3 คือพันธุ์ ก. และ จ. ให้ผลผลิตสูงขึ้น แต่พันธุ์ ข, ค และ ง ให้ผลผลิตลดลง

ข. การทดลองในเวลาต่างกัน

การทดลองบางอย่างถ้ากระทำให้เวลาต่างกัน อาจจะได้ชุดคำตอบที่แตกต่างออกไปก็ได้ เช่น ในการเปรียบเทียบพันธุ์พืช สภาพดินฟ้าอากาศในแต่ละฤดูหรือแต่ละปีมักไม่เหมือนกัน ไม่ว่าเรื่องของอุณหภูมิ ความชื้น ปริมาณแสง ฯลฯ เหล่านี้สามารถทำให้การแสดงออกของพันธุ์แต่ละพันธุ์แตกต่างออกไป ดังนั้นในการปรับปรุงพันธุ์พืชต้องมีการทดลองซ้ำ ๆ ถ้าต้องการปลูกพืชพันธุ์นั้นหลายฤดูในแต่ละปี ก็ต้องทดสอบปีละหลายครั้ง ในฤดูที่จะมีการปลูกจริง แล้วคัดเลือกพันธุ์ที่ให้ผลผลิตสูงสุด หรือพันธุ์ที่ปรับตัวได้ดีกว่าไว้เป็นพันธุ์แนะนำต่อไป การทดสอบหลาย ๆ ปีก็เป็นสิ่งหลีกเลี่ยงไม่ได้ ทั้งนี้เพราะปีคือกลุ่มของสภาพอากาศที่ใช้ในการปลูกพืช แต่ละกลุ่มอาจมีความแตกต่างกัน แต่เป็นได้ว่าเป็นกลุ่มของสภาพแวดล้อมที่ได้มาโดยวิธีการสุ่ม การทดลองซ้ำในท้องที่เดียวกัน มีขั้นตอนการวิเคราะห์วาเรียนซ์เหมือนกับการทดลองซ้ำในหลายท้องที่เช่นที่กล่าวมาแล้วนั่นเอง

17.4 การทดลองซ้ำในแบบสองแนว

สภาพแวดล้อมในการทดลองอาจมีความแตกต่างกันแบบสองแนว เช่น อาจเรียกว่าแนวนอนและแนวตั้ง หรือแนวกว้างและแนวลึก ในการทดลองเกี่ยวกับพืชเป็นความแตกต่างในเรื่องสถานที่และเวลา เช่น ในการทดสอบพันธุ์พืชแต่ละพันธุ์อาจทดสอบ p ท้องที่แต่ละปีทดลอง y ปี

จำนวนการทดลองทั้งสิ้นจึงมี py การทดลอง แต่ละการทดลองมักใช้วิธีการปฏิบัติที่เหมือน ๆ กัน แต่มีการสุ่มใหม่ทุกครั้ง ดังนั้นถือว่าเป็นการทดลองที่เป็นอิสระแก่กัน

แบบจำลองทางคณิตศาสตร์

$$X_{ijkl} = \mu + P_i + Y_j + (PY)_{ij} + B_{(ij)_k} + T_l + (PT)_{il} + (YT)_{jl} + (PYT)_{ijl} + \varepsilon_{ijkl} \quad \ldots(17\text{-}2)$$

เมื่อให้ μ = ค่าเฉลี่ยของประชากร

P = คือสถานที่ทดลอง

Y = ปีที่ทำการทดลอง

PY = ปฏิกิริยาระหว่างสถานที่และปีที่ทำการทดลอง

B = ซ้ำหรือบลอก

T = ผลของทรีตเมนต์

ε = ความคลาดเคลื่อน

PY, PT, YT, PYT คือปฏิกิริยาระหว่างอิทธิพลต่าง ๆ ดังกล่าวข้างบน

i = 1, 2,..., p เมื่อ p คือจำนวนสถานที่ทดลอง

j = 1, 2,..., y เมื่อ y คือจำนวนปีที่ทำการทดลอง

k = 1, 2,..., n เมื่อ n คือจำนวนซ้ำในแต่ละการทดลอง

l = 1, 2,..., t เมื่อ t คือ จำนวนทรีตเมนต์

ในการทดลองเช่นนี้ คาดหมายว่าสถานที่และปีที่ทำการทดลองเป็นตัวแทนของสภาพแวดล้อมของแหล่งและเวลาที่จะปลูกพืชพันธุ์นั้น ๆ ดังนั้นสถานที่และปีจึงให้เป็นปัจจัยสุ่ม ส่วนพันธุ์พืชอาจเป็นปัจจัยคงที่

ตัวอย่าง ในการเปรียบเทียบพันธุ์พืช 5 พันธุ์ ทำการทดลอง 2 ปี ปีละ 3 ท้องที่ แต่ละการทดลองใช้แผนการทดลองแบบ RCB จำนวน 3 ซ้ำ ผลการทดลองแสดงไว้ในตาราง 17.4.1.ก

ในการทดลองซ้ำหลายทิศทางเช่นนี้ ก็ใช้ข้อกำหนดและสมมุติฐาน เช่นเดียวกับการทดลองเดี่ยว ๆ การดำเนินการวิเคราะห์ที่สมบูรณ์ต้องวิเคราะห์ข้อมูลแต่ละการทดลองเดี่ยว ๆ เพื่อทดสอบว่าวาเรียนซ์ของแต่ละการทดลองเท่ากัน ดังตาราง 17.4.1.ข ซึ่งทดสอบโดยวิธีของ Barttlet หรืออาจตั้งเป็นข้อกำหนดโดยไม่ต้องทดสอบ ก่อนการวิเคราะห์วาเรียนซ์ทุกการทดลองร่วมกันต้องจัดทำตารางปฏิกิริยาระหว่างท้องที่ x ปี, พันธุ์ x ท้องที่ และพันธุ์ x ปี ดังตาราง 17.4.2

วิเคราะห์วาเรียนซ์

$$CF = \frac{(\sum X_{ijkl})^2}{npyt} = \frac{(956)^2}{90} = 10{,}154.84$$

$$TSS = \sum X_{ijkl} - CF = 6^2 + 4^2 + \ldots + 9^2 - 10{,}154.84 = 679.16$$

ขั้นที่ 1 วิเคราะห์สภาพแวดล้อม

$$SS(P) = \frac{(\sum P^2)}{nyt} - CF = \frac{287^2 + 375^2 + 298^2}{30} - 10{,}154.84 = 158.82$$

$$SS(Y) = \frac{(\sum Y^2)}{npt} - CF = \frac{498^2 + 456^2}{45} - 10{,}154.84 = 17.78$$

$$SS(PY) = \frac{(\sum PY^2)}{nt} - CF - SS(P) - SS(Y)$$

$$= \frac{135^2 + 152^2 + ... + 137^2}{15} - 10{,}154.84 - 158.82 - 17.78 = 61.09$$

$$SS(Blocks/Place/Year) = SS(B/P/Y) = \frac{\sum(B/P/Y)^2}{t} - CF$$

$$= \frac{47^2 + 39^2 + ... + 45^2}{5} - 10{,}154.84 = 280.36$$

$$Error(a)SS = SS(B/P/Y) - SS(P) - SS(Y) - SS(PY) = 280.36 - 158.82 - 17.78 - 61.09 = 42.67$$

ขั้นที่ 2 วิเคราะห์พันธุ์และปฏิกิริยาต่าง ๆ

$$SS(T) = \frac{\sum T^2}{npy} - CF$$

$$= \frac{165^2 + ... + 176^2}{18} - 10{,}154.84 = 142.83$$

$$SS(TP) = \frac{\sum(TP)^2}{ny} - CF - SS(P) - SS(T)$$

$$= \frac{45^2 + 73^2 + ... + 62^2}{6} - 10{,}154.84 - 158.82 - 142.83 = 65.51$$

$$S(TY) = \frac{\sum(TY)^2}{ny} - CF - SS(Y) - SS(T)$$

$$= \frac{93^2 + 119^2 + ... + 77^2}{9} - 10{,}154.84 - 17.78 - 142.83 = 55.22$$

$$SS(TPY) = \frac{\sum(TPY)^2}{n} - CF - SS(P) - SS(Y) - SS(T) - SS(PY) - SS(PT) - SS(YT)$$

$$= \frac{19^2 + 37^2 + \cdots + 31^2 + 27^2}{3}$$

$$- 10{,}154.84 - 158.82 - 17.78 - 142.83 - 61.09 - 55.22 - 65.51$$

$$= 27.24$$

SS(error) = TSS – SS อื่น ๆ ทั้งหมด

= TSS – [SS(P) + SS(Y) + SS(PY) + SS(error a) + SS(T) + SS(TP) + SS(TY) + SS(TPY)]

= 679.16 – (158.82 + 17.78 + 61.09 + 42.67 + 142.83 + 65.51 + 55.24 + 27.24)

= 107.98

ขั้นที่ 3 นำค่า SS ต่าง ๆ ลงตารางวิเคราะห์วาเรียนซ์ดังตาราง 17.4.3 และทำการทดสอบต่อไป

การทดสอบโดยใช้ค่า F จากตาราง 17.4.3 ต้องมีการคัดเลือกวาเรียนซ์ตัวหารให้ถูกต้อง โดยเฉพาะอย่างยิ่งเมื่อสถานที่และปีเป็นปัจจัยสุ่ม ซึ่งทำให้เกิดค่า EMS ดังแสดงในตาราง 17.4.4 ส่วนทรีตเมนต์หรือพันธุ์พืชให้เป็นปัจจัยสุ่มหรือคงที่ก็ได้ การทดสอบค่าต่าง ๆ แสดงเป็นลำดับดังนี้

1. ทดสอบปฏิกิริยาระหว่าง 3 ปัจจัย ($T \times P \times Y$)

$$F_1 = \frac{M_2}{M_1} = \frac{\sigma^2 + n\sigma^2_{tpy}}{\sigma^2} = \frac{3.41}{2.25} = 1.52^{ns}$$

เปิดตาราง F ที่ df 8 และ 48 ตามลำดับ

2. ทดสอบปฏิกิริยาระหว่าง 2 ปัจจัย ($T \times Y$) ถ้า F_1 ไม่แตกต่างก็ใช้ MS pooled error (M_1) เป็นตัวหารดังนี้

$$F_2 = M_3/M_1 = 13.81/2.35 = 6.13^{**} \quad (df\ 4, 48)$$

ทดสอบปฏิกิริยา 2 ปัจจัย ($T \times P$)

$$F_3 = M_4/M_1 = 8.11/2.25 = 3.60^{**} \quad (df\ 4, 48)$$

อย่างไรก็ดีถ้า F_1 แตกต่างให้ใช้ MS(TPY) เป็นตัวหารทั้ง 2 รายการ

3. ทดสอบทรีตเมนต์หรือพันธุ์พืช ในการทดสอบทรีตเมนต์นั้น เมื่อพิจารณาจากตาราง 17.5.4 แล้วไม่พบว่ามี EMS ตัวใดที่เหมาะสมจะเป็นตัวหาร แต่สามารถจะดัดแปลงได้ การดัดแปลงนี้ทำให้เราได้เห็นขนาดของ σ^2_t วิธีนี้เรียกว่า approximate (F′) test ตามการแนะนำของ Cochran และ Cox (1955) ดังนี้

$$F = \frac{M_5 + M_2}{M_4 + M_3}$$

ซึ่งเมื่อนำมาหารกันแล้วได้ σ_t^2 เป็นส่วนต่าง

$$F = \frac{35.71+3.41}{8.11+13.81} = \frac{39.12}{21.92} = 1.78$$

การตัดสินว่า F′ จะมากถึงระดับแตกต่างทางสถิติได้ต้องมีค่า F มาเปรียบเทียบ ซึ่งแนะนำวิธีการคำนวณ df โดย Satterwaite (1936) ดังนี้

$$\text{df(ตัวตั้ง)} = \frac{(M_4+M_3)^2}{M_5^2/\text{df}(t)+M_2^2/\text{df}(TPY)}$$

$$= \frac{(35.71+3.41)^2}{(35.71)^2/4+(3.41)^2/8} = \frac{1{,}530.37}{320.25} = 4.78$$

หรือประมาณ 5

$$\text{df(ตัวหาร)} = \frac{(M_4+M_3)^2}{M_4^2/\text{df}(TP)+M_3^2/\text{df}(TY)}$$

$$= \frac{(8.11+13.81)^2}{(8.11)^2/8+(13.81)^2/4} = \frac{480.49}{55.90} = 8.60$$

หรือประมาณ 9

เมื่อเปิดตาราง F ที่ df 5 และ 9 พบว่า F มีค่าดังนี้ : $F_{0.05} = 3.48$ และ $F_{0.01} = 6.06$ ซึ่งทรีตเมนต์ไม่แตกต่างทางสถิติ

ตาราง 17.4.1 ผลผลิตของพืช 5 พันธุ์ที่ทดสอบใน 3 ท้องที่เป็นเวลา 2 ปี

ก. ผลผลิต

พันธุ์	ปีที่ 1 ท้องที่ 1 I	II	III	รวม	ท้องที่ 2 I	II	III	รวม	ท้องที่ 3 I	II	III	รวม	รวมยอด
ก	6	4	9	19	15	13	13	41	11	13	9	33	93
ข	14	11	12	37	13	16	18	47	10	11	14	35	119
ค	10	8	9	27	15	13	14	42	8	10	8	26	95
ง	11	8	10	29	12	13	12	37	8	9	9	26	92
จ	6	8	9	23	12	14	15	41	12	14	9	35	99
	47	39	49	135	67	69	72	208	49	57	49	155	498

พันธุ์	ปีที่ 2 ท้องที่ 1 I	II	III	รวม	ท้องที่ 2 I	II	III	รวม	ท้องที่ 3 I	II	III	รวม	รวมยอด
ก	9	9	8	26	9	7	8	24	9	7	6	22	72
ข	11	11	14	36	12	14	16	42	10	12	11	33	111
ค	8	8	11	27	11	15	12	38	7	8	9	24	89
ง	12	14	15	41	11	14	12	37	9	12	10	31	109
จ	8	7	9	24	8	7	11	26	8	10	9	27	77
	48	49	57	154	51	57	59	167	43	49	45	137	458

ข. ผลการวิเคราะห์วาเรียนซ์แยกย่อย ๆ

ผลปีที่ 1

Sources	df	ท้องที่ 1 SS	MS	ท้องที่ 2 SS	MS	ท้องที่ 3 SS	MS
Blocks	2	11.20	5.60	2.53	1.27	8.53	4.26
Varieties	4	61.33	15.33	17.06	4.26	28.67	7.17
Error	8	17.47	2.18	20.14	2.52	24.13	3.10
Total	14	90.00		39.73		61.33	

ผลปีที่ 2

Sources	df	ท้องที่ 1 SS	MS	ท้องที่ 2 SS	MS	ท้องที่ 3 SS	MS
Blocks		9.73	4.87	6.93	3.47	3.73	1.87
Varieties		71.60	17.90	83.73	20.93	28.40	7.10
Error		9.60	1.20	25.07	3.13	11.60	1.45
Total		90.93		115.73		43.73	

ตาราง 17.4.2 ปฏิกิริยาระหว่างอิทธิพลที่เกี่ยวข้อง

ก. ตารางท้องที่ × ปี

	P_1	P_2	P_3	
Y_1	135	208	155	498
Y_2	152	167	137	456
รวม	287	375	298	498

ข. ตารางพันธุ์ × ท้องที่

พันธุ์	P_1	P_2	P_3	รวม
ก	45	65	55	165
ข	73	89	68	230
ค	54	80	50	184
ง	70	74	57	201
จ	47	67	62	176
	289	375	292	956

ค. ตารางพันธุ์ × ปี

	ปีที่ 1	ปีที่ 2	รวม
ก	93	72	165
ข	119	111	230
ค	95	89	184
ง	92	109	201
จ	99	77	176
	498	458	956

ตาราง 17.4.3 ผลการวิเคราะห์วาเรียนซ์ข้อมูลใน ตาราง 17.6.2

Sources	df	SS	MS	F
Places (P)	2	158.82	79.41	
Years (Y)	1	17.78	17.78	
P × Y	2	61.09	30.55	
Error (a)	12	42.69	3.56	
Treatments (T)	4	142.83	35.71	
T × P	8	65.51	8.11	
T × Y	4	55.22	13.81	
T× P × Y	8	27.24	3.41	1.52
Pooled error	48	108.00	2.25	
Total	**89**	**679.16**		

ตาราง 17.4.4 ค่า expecxed mean square ของการทดลองหลายครั้งที่ทำหลายปีและหลายท้องที่

Sources	df	MS	
Places (P)	$p-1$	M_9	$\sigma^2 + n\sigma^2_{tpy} + nt\sigma^2_{py} - nty\sigma^2_p$
Years (Y)	$y-1$	M_8	$\sigma^2 + n\sigma^2_{tpy} + nt\sigma^2_{py} + ntp\sigma^2_y$
$P \times Y$	$(p-1)(y-1)$	M_7	$\sigma^2 + n\sigma^2_{tpy} + nt\sigma^2_{py}$
Error (a)	$py(n-1)$	M_6	$\sigma^2 + n\sigma^2_{tpy}$
Treatment(s)(T)	$(t-1)$	M_5	$\sigma^2 + n\sigma^2_{tpy} + np\sigma^2_{ty} + ny\sigma^2_{tp} + npy\sigma^2_t$
$T \times P$	$(p-1)(t-1)$	M_4	$\sigma^2 + n\sigma^2_{tpy} + ny\sigma^2_{tp}$
$T \times Y$	$(p-1)(y-1)$	M_3	$\sigma^2 + n\sigma^2_{tpy} + npk\sigma^2_{ty}$
$T \times P \times Y$	$(t-1)(p-1)(y-1)$	M_2	$\sigma^2 + n\sigma^2_{tpy}$
Pooled error	$py(t-1)(n-1)$	M_1	σ^2

4. ทดสอบปฏิกิริยาระหว่างท้องที่และปี ($P \times Y$)

$$F_5 = M_7/M_6 = 30.55/3.56 = 8.58^{**}$$

เปิดตาราง $F_{0.01}$ ที่ df 2, 12 = 6.93

5. ทดสอบปี (Y) เนื่องจากกำหนดไว้ว่าให้ปีเป็นปัจจัยสาม ดังนั้น

$$F_6 = M_8/M_7 = 17.78/30.35 = 0.58$$

ค่า $F < 1$ และ df ก็น้อยมาก ปกติไม่ทำการทดสอบ

6. ทดสอบท้องที่ (P) ในทำนองเดียวกันกับข้อ 5 ดังนี้

$$F_7 = M_9/M_7 = 79.41/30.55 = 2.60$$

ซึ่งค่า F ที่คำนวณได้ต่ำมาก และ df ก็น้อยมาก จึงให้ข้อสรุปเช่นเดียวกับข้อ 5

17.5 การแปลผลจากการทดลอง

ในการทดลองหลายครั้ง เช่น ในการเปรียบเทียบพันธุ์พืช เรียกว่า เป็นการทดลองในหลายสภาพแวดล้อม อันประกอบด้วยฤดู ปี และสถานที่ จะได้แยกแหล่งของความแปรปรวนแบ่งออกมาได้ ดังตาราง 17.3.1 ซึ่งจะเห็นได้ว่ามีการทดสอบสภาพแวดล้อม และทดสอบทรีตเมนต์เป็นขั้น ๆ ไป นอกนั้นได้เพิ่มการทดสอบปฏิกิริยาระหว่างทรีตเมนต์กับสภาพแวดล้อม ผู้ทดลองต้องเข้าใจวัตถุประสงค์ของผลการทดสอบเหล่านี้ จึงแปลความหมายได้อย่างมีประโยชน์

เมื่อปฏิกิริยามีความสำคัญ แปลความหมายดังนี้

1. ปฏิกิริยาระหว่างปีและท้องที่ เมื่อมีความสำคัญ หมายความว่า โดยเฉลี่ยแล้วในท้องที่หนึ่ง ปีหนึ่งจะให้ผลผลิตโดยเฉลี่ยดีกว่าอีกปีหนึ่ง แต่อีกท้องที่ให้ผลตรงกันข้าม ปฏิกิริยานี้ไม่มีผลต่อการแนะนำว่าใช้พืชพันธุ์ใด เพียงแต่บอกว่าปีหรือสภาพแวดล้อมการปลูกพืชแต่ละครั้งไม่คงเส้นคงวา

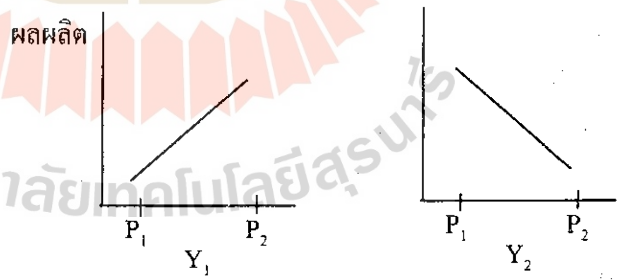
2. ปฏิกิริยาระหว่างพันธุ์และปี เมื่อค่านี้มีความสำคัญ หมายความว่า โดยเฉลี่ยทุกท้องที่แล้วพบว่าพืชบางพันธุ์ให้ผลดีกว่าพันธุ์อื่นในบางปี แต่บางปีผลผลิตต่ำกว่า ถ้าค่ามีนสแควร์ของพันธุ์สูงกว่าค่ามีนสแควร์ของพันธุ์ x ปี แสดงว่าบางพันธุ์ให้ผลผลิตสูงกว่าพันธุ์อื่นทุกปี

3. ปฏิกิริยาระหว่างพันธุ์และท้องที่ ถ้าค่านั้นแตกต่างในทางสถิติก็แสดงว่า โดยเฉลี่ยจากทุกปีพืชบางพันธุ์ให้ผลผลิตสูงในบางท้องที่แต่ให้ผลผลิตต่ำในท้องที่อื่น ถ้าหากว่าค่ามีนสแควร์ของพันธุ์สูงกว่ามีนสแควร์ของปฏิกิริยาระหว่างพันธุ์กับท้องที่ ก็แสดงว่ามีบางพันธุ์ให้ผลผลิตสูงในทุกท้องที่แต่ถ้าค่ามีนสแควร์ของพันธุ์ไม่แตกต่างจากของปฏิกิริยาระหว่างพันธุ์และท้องที่ ก็แสดงว่าควรแนะนำพันธุ์แต่ละท้องที่แยกกัน

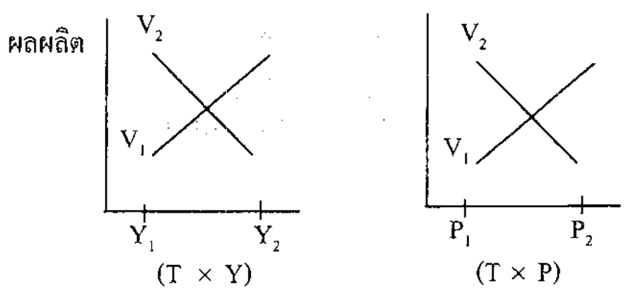
4. ปฏิกิริยาระหว่างพันธุ์-ปี-ท้องที่ แสดงว่าปฏิกิริยาระหว่างพันธุ์ และท้องที่ในแต่ละปีเป็นไปในทางตรงกันข้าม ถ้าค่านี้มีความสำคัญ หมายความว่า ในปีหนึ่งในสถานที่หนึ่งพืชพันธุ์หนึ่งให้ผลผลิตต่ำ แต่อีกพันธุ์ให้ผลผลิตสูง ในอีกสถานที่หนึ่งผลผลิตเป็นแบบตรงกันข้าม และในอีกปีผลจะเป็นตรงกันข้าม

ปฏิกิริยาดังที่กล่าวแล้วข้างบนอาจจะแสดงง่าย ๆ โดยใช้รูปดังแสดงในรูป 17.5.1

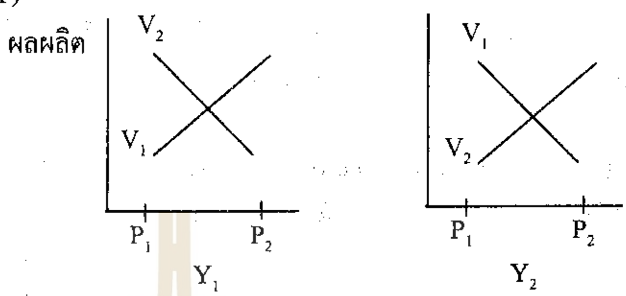
ก. เมื่อมีปฏิกิริยาระหว่างปีและท้องที่ (P x Y)



ข. เมื่อมีปฏิกิริยาระหว่าง
พันธุ์และปี (T x Y)
พันธุ์และท้องที่ (T x P)



ค. เมื่อมีปฏิกิริยา ระหว่างพันธุ์-ปี
และท้องที่ (T × P × Y)



รูป 17.5.1 อธิบายผลของปฏิกิริยา ระหว่างปัจจัยการทดลองในรูปแบบต่าง ๆ กัน

17.6 แบบฝึกหัด

ในการทดลองเปรียบเทียบพันธุ์ถั่วลิสง 5 พันธุ์ในไร่นาของกสิกร 2 จังหวัด จำนวน 2 ปี ได้ผลผลิต (กก./ไร่) ดังนี้

พันธุ์	บลอก				บลอก			
	I	II	III	รวม	I	II	III	รวม
	--------------สงขลา-2529--------------				--------------สงขลา-2530--------------			
ไทนาน 9	170	200	197	567	336	377	312	1,025
สข. 38	161	189	203	553	297	300	359	956
ลำปาง	211	231	218	660	241	269	298	808
ขอนแก่น 60	154	209	184	547	263	318	298	879
RCM 387	203	210	142	555	281	254	315	850
รวม	899	1,039	944	2,882	1,418	1,518	1,582	4,518
	--------------พัทลุง-2529--------------				--------------พัทลุง-2530--------------			
ไทนาน 9	268	336	264	868	228	207	235	670
สข. 38	292	324	260	876	243	262	267	772
ลำปาง	233	228	185	646	272	249	246	767
ขอนแก่น 60	262	288	233	783	278	263	240	781
RCM 387	226	210	242	678	240	237	233	710
รวม	1,281	1,382	1,184	3,851	1,261	1,218	1,221	3,700

จากข้อมูลดังกล่าวนี้

ก. จงวิเคราะห์วาเรียนซ์ข้อมูลแต่ละการทดลองแยกกัน แล้วทดสอบว่า MS error เท่ากันโดยใช้วิธีที่เหมาะสม

ข. จงวิเคราะห์วาเรียนซ์ในแต่ละท้องที่และทดสอบสมมุติฐาน ถ้าหากปีเป็นปัจจัยสุ่ม

ค. จงวิเคราะห์วาเรียนซ์ในแต่ละปี และทดสอบสมมุติฐาน ถ้าหากท้องที่เป็นปัจจัยคงที่

ง. จงวิเคราะห์วาเรียนซ์ร่วมปีและท้องที่ และคำนวณค่า F จนครบ ถ้าหากว่าพันธุ์เป็นปัจจัยคงที่ ส่วนปีและท้องที่เป็นปัจจัยสุ่ม

2. ในการทดสอบปุ๋ยไนโตรเจน 5 ระดับ ต่อผลผลิตของข้าวหอมมะลิ 105 ใน 2 ปี พบว่าได้ผลผลิตดังนี้

อัตราปุ๋ยไนโตรเจน	ผลผลิต (กก./ไร่)		
(กก./ไร่)	บลอก 1	บลอก 2	บลอก 3
		ปีที่ 1	
5	489	257	454
10	601	662	567
15	671	669	680
20	646	667	663
25	568	687	569
		ปีที่ 2	
5	500	350	536
10	635	632	658
15	607	597	589
20	482	402	581
25	344	405	374

ก. จงวิเคราะห์วาเรียนซ์ร่วมทั้ง 2 ปี

ข. ในการวิเคราะห์ ถ้าพบว่าผลของไนโตรเจนสำคัญ จงวิเคราะห์เพื่อแสดงให้เห็นว่าปุ๋ยนี้แสดงผลในแบบใดมากที่สุด คือแบบ linear, quadratic หรือ cubic

ค. จงแยกผลของปฏิกิริยาระหว่างปุ๋ยและปีออกเป็นส่วนย่อย ๆ เพื่อแสดงให้เห็นว่าปฏิกิริยาเป็นแบบใดสำคัญที่สุด คือแบบ NY linear, NY quadratic หรือ NY cubic

คำในบท

(1) combined analysis, (2) homogeneity of variance

บทที่ 18
แผนการทดลองแบบแลตติส

18.1 คำนำ

ในการทดลองแต่ละครั้ง อาจจำเป็นต้องเปรียบเทียบทรีตเมนต์ครั้งละมาก ๆ เช่นในการเปรียบเทียบสายพันธุ์พืชอาจกระทำครั้งละมากกว่า 100 สายพันธุ์ ดังนั้นจำเป็นต้องใช้หน่วยทดลองเป็นจำนวนมากเป็นเงาตามตัว ซึ่งเป็นการยากที่จะจัดทำบลอกได้อย่างมีประสิทธิภาพ เพราะเป็นแปลงขนาดใหญ่ ไม่อาจหาพื้นที่สม่ำเสมอได้ดีเพียงพอ ถ้าทำการทดลองการใช้แผนการทดลองที่ศึกษามาแล้ว เช่นแผนการทดลองแบบ RCB ก็จะเพิ่มความคลาดเคลื่อนอันเนื่องจากความไม่สม่ำเสมอของหน่วยทดลองขึ้นได้ ในกรณีของการทดลองแบบแฟกตอเรียลก็อาจแก้ปัญหาโดยใช้การทดลองแบบคอนฟาวด์ แต่ในการทดสอบพันธุ์พืช เราต้องการเปรียบเทียบระหว่างทุกคู่ของทรีตเมนต์อย่างเท่าเทียมกัน จึงไม่อำนวยให้ทำเช่นนั้น

เพื่อเป็นการแก้ปัญหาดังกล่าวมาแล้วข้างต้น ได้มีการพัฒนาแผนการทดลองขึ้นมาใหม่ (Yates, 1936) ในแผนการทดลองนี้ได้แบ่งการทดลอง 1 ชุด หรือ 1 ซ้ำ ออกเป็นบลอกย่อย ๆ หลายบลอก แต่ละบลอกรับทรีตเมนต์ไปชุดหนึ่ง ซึ่งไม่ครบทุกทรีตเมนต์ แต่เมื่อรวมทุกบลอกแล้วจะมีอยู่ครบ เราเรียกแผนการทดลองนี้ว่าแผนการทดลองแบบบลอกไม่สมบูรณ์ (incomplete block design) ซึ่งต่อมาได้มีการพัฒนาเพิ่มเติม เป็นชนิดย่อย ๆ หลายชนิด ซึ่งมีชื่อเรียกต่าง ๆ กันไป เช่น การทดลองที่มี 9 ทรีตเมนต์ อาจจัดแปลงทดลอง 1 ซ้ำได้ดังรูป 18.1.1

บลอกที่ 1	1	2	3
บลอกที่ 2	4	5	6
บลอกที่ 3	7	8	9

รูป 18.1.1 การทดลองแบบบลอกไม่สมบูรณ์ 1 ซ้ำ

จากรูป 18.1.1 เห็นได้ว่า ในการเปรียบเทียบ 9 ทรีตเมนต์ กระจายอยู่ใน 3 บลอกเล็ก ๆ บลอกเล็กขนาดนี้เป็นการง่ายที่จะควบคุมความคลาดเคลื่อนภายในบลอก จึงได้หน่วยทดลองที่สม่ำเสมอ ทำให้วาเรียนซ์ระหว่างแปลงในบลอกเดียวกันมีน้อย และน้อยกว่าวาเรียนซ์ระหว่างบลอกย่อยในแต่ละซ้ำ จึงเป็นการเพิ่มประสิทธิภาพของการทดลอง อย่างไรก็ดี ในการทดลองเช่นนี้มีข้อเสียเช่นกัน เช่น ข้อกำหนดที่เคร่งครัดในการวางแผน และขั้นตอนในการวิเคราะห์ แต่จำเป็นต้องใช้เมื่อเป็น

การทดลองขนาดใหญ่ ไม่ว่าเป็นการเปรียบเทียบพันธุ์พืชหรือการทดลองเกี่ยวกับสัตว์ หากใช้ครอกสัตว์เป็นบลอก แต่จำนวนสัตว์ไม่พอสำหรับ 1 ซ้ำ

แผนการทดลองแบบบลอกไม่สมบูรณ์มีอยู่หลายชนิด ชนิดที่นิยมใช้กันมากได้แก่พวกที่จำนวนทรีตเมนต์ในแต่ละบลอก (k) เท่ากับรากที่สองของจำนวนทรีตเมนต์ทั้งหมด (t) ซึ่งกล่าวได้ว่า $t = k^2$ หรือ $k = \sqrt{t}$ แผนการทดลองที่มีคุณสมบัติเช่นนี้เรียกว่าแผนการทดลองแบบแลตติส ส่วนแผนการทดลองที่จัดทรีตเมนต์ในลักษณะอื่น ๆ จะไม่รวมไว้ในระดับนี้ เพราะไม่ค่อยมีการใช้ประโยชน์ ผู้สนใจอาจศึกษาเพิ่มเติมจากตำราที่ใช้เป็นเอกสารอ้างอิง (Federer, 1955; Cochran และ Cox, 1957; จรัญ จันทลักขณา, 2523 , สุรพล อุปดิสสกุล, 2537)

การจำแนกชนิดของแผนการทดลองแบบแลตติส

การทดลองแบบแลตติส แบ่งออกได้เป็น 2 ชนิด ใหญ่ ๆ คือ

1. ชนิดที่สมดุล (balanced) ซึ่งแบ่งออกเป็นชนิดย่อย ๆ อีก เช่น
   (1) แลตติสสมดุล (balanced lattice)
   (2) แลตติสสแควร์ (lattice square)
   (3) ยูเดนสแควร์ (ลาตินสแควร์ที่ไม่สมดุล) (Youden square)
2. ชนิดที่สมดุลบางส่วน (partially balanced lattice)
   (1) แลตติสสองซ้ำ (double lattice) หรือแลตติสอย่างง่าย (simple lattice)
   (2) แลตติสสามซ้ำ (triple lattice)
   (3) แลตติสแบบมุมฉาก (rectangular lattice)

เนื่องจากแต่ละชนิดมีความแตกต่างกัน จึงอธิบายถึงลักษณะของแต่ละแผนการทดลองเมื่ออธิบายถึงแผนการทดลองชนิดนั้น ๆ

18.2 แผนการทดลองแลตติสแบบสมดุล

แผนการทดลองแลตติสแบบสมดุลเสนอโดย Yates ในปี 1937 แผนการทดลองมีลักษณะดังนี้

(1) จำนวนทรีตเมนต์ (t) เท่ากับกำลังสองของขนาดของบลอก (k) คือ $t = k^2$ หรือ $\sqrt{t} = k$ ดังนั้นในการทดลองแต่ละครั้งจำนวนทรีตเมนต์เท่ากับ 9, 16, 25, 36, 49, 64 ฯลฯ

(2) ในแต่ละการทดลองขนาดของบลอกเท่ากับค่ารากสองของจำนวนทรีตเมนต์ คือ $k = \sqrt{t}$ เช่นขนาดของบลอกเท่ากับ 3, 4, 5, 6, 7, 8,…แปลงย่อย

(3) ในแต่ละการทดลองจำนวนซ้ำเท่ากับขนาดของบลอกบวก 1 คือ $r = k + 1$ ดังนั้นจำนวนซ้ำเท่ากับ 4, 5, 6, 7, 8… สำหรับการทดลองที่มี 9, 16, 25, 36 และ 49 ทรีตเมนต์ตามอันดับ

(4) แต่ละคู่ของทรีตเมนต์ อยู่ในบลอกเดียวกัน 1 ครั้ง $(\lambda = 1)$

การสุ่ม

ในการใช้แผนการทดลองดังกล่าว การสุ่มกระทำโดยใช้แผนการทดลองมาตรฐานที่แสดงโดย Cochran และ Cox (1957) ยกตัวอย่างการทดลองขนาด 9 ทรีตเมนต์ ดังนั้นขนาดของบลอก (k) = 3, จำนวนซ้ำ (n) = k + 1 = 4 ซึ่งมีแผนการทดลองพื้นฐานดังตาราง 18.2.1

ตาราง 18.2.1 แผนการทดลองมาตรฐานของการทดลองที่มี 9 ทรีตเมนต์ให้ 1, 2,..., 9 เป็นหน่วยทดลองในแต่ละซ้ำ

บลอก	Rep 1			บลอก	Rep 2			บลอก	Rep 3			บลอก	Rep 4		
(1)	1	2	3	(4)	1	4	7	(7)	1	5	9	(10)	1	6	8
(2)	4	5	6	(5)	2	5	8	(8)	2	6	7	(11)	2	4	9
(3)	7	8	9	(6)	3	6	9	(9)	3	4	8	(12)	3	5	7

ทั้งนี้ในการจัดแผนการทดลองซ้ำที่ 3 และ 4 ได้จากการจัดออโทโกนัลลาตินสแควร์จากซ้ำที่ 1 และ 3 ตามลำดับนั่นเอง

ขั้นที่ 1 สุ่มลำดับของซ้ำ จากลำดับเดิมตามที่แสดงไว้ในตาราง 18.2.1 มีลำดับของซ้ำเป็น 1, 2, 3, 4 หลังจากสุ่มแล้วอาจได้ลำดับเป็น 4, 1, 2 และ 3 และให้เป็นลำดับใหม่ คือ I, II, III และ IV ดังนี้

บลอก	Rep I (4)			Rep II (1)			Rep III (2)			Rep IV (3)		
(1)	1	6	8	1	2	3	1	4	7	1	5	9
(2)	2	4	9	4	5	6	2	5	8	2	6	7
(3)	3	5	7	7	8	9	3	6	9	3	4	8

ขั้นที่ 2 ทำการสุ่มบลอกในแต่ละซ้ำเหมือนแต่ละบลอกเป็นหน่วยเดียวกัน และสุ่มแปลงภายในบลอก เป็นขั้นสุดท้าย

บลอก	Rep I			บลอก	Rep II			บลอก	Rep III			บลอก	Rep IV		
(1)	8	6	1	(4)	8	7	9	(7)	8	5	2	(10)	3	4	8
(2)	3	5	7	(5)	3	2	1	(8)	4	7	1	(11)	6	2	7
(3)	4	2	9	(6)	6	5	4	(9)	3	6	9	(12)	9	5	1

ขั้นที่ 3 ทำการทดลองโดยใช้ผลการสุ่มในขั้นที่ 2 บันทึกข้อมูลแล้วนำมาวิเคราะห์-วาเรียนซ์ต่อไป สมมุติว่าได้ผลดังแสดงในตาราง 18.2.2

ตารางที่ 18.2.2 ผลการทดลองเปรียบเทียบ 9 ทรีตเมนต์ โดยใช้แลตติสสมดุล

บลอก	Rep I			บลอก	Rep II			บลอก	Rep III			บลอก	Rep IV		
(1)	4	5	7	(4)	3	3	8	(7)	4	7	4	(10)	7	6	5
(2)	5	7	4	(5)	6	5	6	(8)	5	4	7	(11)	6	6	5
(3)	3	4	6	(6)	6	8	4	(9)	5	5	9	(12)	9	8	7

การวิเคราะห์วาเรียนซ์

การวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 18.2.2 กระทำเป็นขั้น ๆ ดังนี้

ขั้นที่ 1 คำนวณหาผลรวมของบลอก (B) ของซ้ำ (R) ของทรีตเมนต์ (T) ผลรวมทั้งหมด (G) ผลรวมของบลอกที่มีทรีตเมนต์ที่ i (B_T) และคำนวณค่า W ดังนี้

ผลรวมของบลอก (B) เช่น บลอก I = 4 + 5 + 7 = 16

ผลรวมของซ้ำ (R) เช่นซ้ำ I = 16 + 16 + 13 = 45

ผลรวมของทรีตเมนต์ เช่น T_1 = 7 + 6 + 7 + 7 = 27

ผลรวมของบลอกที่มีทรีตเมนต์ที่ i เช่น B_T = 16 + 17 + 16 + 24 = 73

W หาจากสมการดังนี้

$$W = kT - (k+1)B_T + G \qquad \ldots(18\text{-}1)$$

เช่น $$W_1 = (3)(27) - (3+1)(73) + 203 = -8$$

ผลการคำนวณแสดงไว้ในตาราง 18.2.4

แหล่งของความปรวนแปร

ตารางวิเคราะห์วาเรียนซ์ซึ่งมีแหล่งของความปรวนแปรดังตาราง 18.2.3

ตาราง 18.2.3 แหล่งของความปรวนแปรและ df ในการทดลองแบบแลตติสสมดุล 9 ทรีตเมนต์

Sources	df	
Replications	$n-1$	= 3
Treatments (unadj.)	k^2-1	= 8
Blocks (adj.)	$n(k-1)$	= 8
Intrablock error	$(k-1)(k^2-1)$	= 16
Total	nk^2-1	= 35

ขั้นที่ 2 การคำนวณค่า Sum of Squares ต่าง ๆ

$$CF = \frac{(\text{ผลรวม})^2}{(k+1)(k^2)} = \frac{(203)^2}{36} = 1{,}144.69$$

$$\text{TSS} = \sum X^2 - \text{CF}$$

$$= 4^2 + 5^2 + \ldots + 7^2 - 1{,}144.69 = 94.31$$

$$\text{Rep SS} = \frac{\sum R^2}{k^2} - \text{CF}$$

$$= \frac{45^2 + \ldots + 59^2}{9} - 1{,}144.69 = 11.64$$

ตาราง 18.2.4 **ผลการทดลองและการเตรียมการวิเคราะห์ในการทดลองแบบแลตติสสมดุลย์**

ซ้ำ	บลอก				รวม บลอก (B)	ทรีตเมนต์	รวม ทรีตเมนต์ (T)	รวม บลอก(B_T)	W
I	(1)	4(8)	5(6)	7(1)	16	1	27	73	– 8
	(2)	5(3)	7(5)	4(7)	16	2	19	62	12
	(3)	3(4)	4(2)	6(9)	13	3	23	70	– 8
		รวมซ้ำ I			45	4	18	65	– 3
II	(4)	3(8)	3(7)	8(9)	14	5	30	73	1
	(5)	6(3)	5(2)	6(1)	17	6	22	70	– 11
	(6)	6(6)	8(5)	4(4)	18	7	16	63	– 1
		รวมซ้ำ II			49	8	16	63	– 1
III	(7)	4(8)	7(5)	4(2)	15	9	32	70	19
	(8)	5(4)	4(7)	7(1)	16		203	609	0
	(9)	5(3)	5(6)	9(9)	19				
		รวมซ้ำ III			50				
IV	(10)	7(3)	6(4)	5(8)	18				
	(11)	6(6)	6(2)	5(7)	17				
	(12)	9(9)	8(5)	7(1)	24				
		รวมซ้ำ IV			59				

$$\text{SSTr (unadj.) SS} = \frac{\sum T^2}{(k+1)} - CF$$

$$= \frac{27^2 + 19^2 + ... + 32^2}{4} - 1{,}144.69 = 71.06$$

$$\text{Block (adj.) SS} = \frac{\sum W^2}{(k+1)(k^3)} \quad \text{...(18-2)}$$

$$= \frac{(-8)^2 + 12^2 + ... + 19^2}{(4)(27)} = 7.09$$

Intrablock error SS = Total SS – Rep SS – SSTr (unadj.) – Block (adj.) SS

= 94.31 – 11.64 – 71.06 – 7.09 = 4.52

ขั้นที่ 3 คำนวณค่า mean squares ต่าง ๆ นำค่า sum of squares ที่คำนวณได้ลงตาราง วิเคราะห์วาเรียนซ์ ดังตาราง 18.2.5 แล้วคำนวณค่า mean squares ต่อไป

ตาราง 18.2.5 ผลการวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 18.2.2

Sources	df	SS	MS
Replications	3	11.64	
Treatments (unadj.)	8	71.06	8.88
Blocks (adj.)	8	7.09	(E_b) 0.89
Intrablock error	16	4.52	(E_e) 0.28
Total	35	94.31	

ขั้นที่ 4 ปรับค่าผลรวมของทรีตเมนต์และคำนวณ SSTr (adjusted)

ในกรณีที่ E_b มากกว่า E_e ต้องมีการปรับผลรวมของทรีตเมนต์ โดยให้สมการ

$$T' = T + \mu W$$

เมื่อให้ T' = ผลรวมของทรีตเมนต์เมื่อปรับค่าแล้ว

T = ผลรวมของทรีตเมนต์ที่ต้องปรับ

$$\mu = \frac{E_b - E_e}{k^2 E_b} \quad \text{...(18-3)}$$

W = ค่าที่คำนวณจากสมการ (18-1)

ซึ่งหาได้ว่า

$$\mu = \frac{0.89 - 0.30}{(9)(0.89)} = \frac{0.61}{8.01} = 0.076$$

ดังนั้น

$$
\begin{aligned}
T_1' &= 27+(0.076)(-8)=26.39\\
T_2' &= 19+(0.074)(12)=19.91\\
T_3' &= 23+(0.074)(-8)=22.39\\
T_4' &= 18+(0.074)(-3)=17.77\\
T_5' &= 30+(0.074)(1)=30.08\\
T_6' &= 22+(0.074)(-11)=21.16\\
T_7' &= 16+(0.074)(-1)=15.92\\
T_8' &= 16+(0.074)(-1)=15.92\\
T_9' &= 32+(0.074)(19)=33.44
\end{aligned}
$$

$$
\begin{aligned}
\text{SSTr (adj)} &= \frac{\sum T_i'^{2}}{k+1} - CF\\
&= \frac{(26.4)^{2}+\ldots+(33.4)^{4}}{4} - 1{,}144.69 = 77.21
\end{aligned}
$$

ดังนั้น $\text{MSTr (adj.)} = \dfrac{\text{SSTr(adj)}}{k^2-1} = \dfrac{77.21}{8} = 9.65$

ขั้นที่ 5 ทดสอบผลของทรีตเมนต์หลังจากปรับค่า

การทดสอบผลของทรีตเมนต์ภายหลังการปรับค่าแล้ว ใช้สมการดังนี้

$$F = \frac{\text{MSTr(adj.)}}{\text{Effective error MS }(E_e')}$$

ทั้งนี้

$$E_e' = E_e(1+k\mu) \qquad \ldots(18\text{-}4)$$

$$= 0.28\,[1+(3)(0.076)] = 0.34$$

ดังนั้นสามารถทดสอบ MSTr (adj.) ดังนี้

$$F = \frac{9.65}{0.34} = 28.38^{**} \qquad (\text{df } 8, 16)$$

ขั้นที่ 6 คำนวณความคลาดเคลื่อนเพื่อเปรียบเทียบค่าเฉลี่ย

จากผลการวิเคราะห์วาเรียนซ์แสดงไว้ในตาราง 18.2.5 พบว่าทรีตเมนต์มีความแตกต่างกันในทางสถิติ ในขั้นต่อไปอาจคำนวณ s_d^2 และ CV ดังนี้

$$s_{\bar{d}}^2 = \frac{2E'_e}{k+1} = \frac{(2)(0.34)}{4} = 0.17 \qquad \ldots(18\text{-}5)$$

ดังนั้น $s_{\bar{d}} = \sqrt{0.185} = 0.41$

$$CV(\%) = \frac{\sqrt{E'_e}}{\text{Grand mean}} \times 100$$

$$= \frac{\sqrt{0.34}}{203/36} \times 100 = 10.34\%$$

ขั้นที่ 7 คำนวณค่าบ่งชี้ความเที่ยงตรงเปรียบเทียบกับแผนการทดลองแบบ RCB โดยใช้สมการดังนี้

$$\text{Relative precision} = \frac{\text{Error MS(RCB)}}{\text{Effective error}}$$

ทั้งนี้ $$\text{Error MS(RCB)} = \frac{\text{Block/rep SS} + \text{Intrablock error SS}}{\text{df block} + \text{df intrablock}}$$

$$= \frac{7.09 + 4.52}{8 + 16} = 0.48$$

ดังนั้น $\text{Relative precision} = \frac{0.48}{0.34} = 1.41$ หรือ 141%

คือสรุปว่าการใช้แผนการทดลองแบบแลตติสสมดุลให้ผลดีกว่าการใช้ RCB ปกติประมาณ 41 เปอร์เซ็นต์ ซึ่งหมายถึงว่าการใช้แผนการทดลองแลตติสแบบสมดุล 4 ซ้ำ จะให้ความถูกต้องเที่ยงตรงเทียบเท่าการทดลองแบบ RCB มากกว่า 5 ซ้ำ

ค่าสูญหายในแผนการทดลองแลตติสสมดุล

เนื่องจากเป็นการทดลองที่มีขนาดใหญ่ ดังนั้นย่อมเป็นธรรมดาที่อาจมีค่าสูญหายเกิดขึ้นได้ เช่น พืชล้มตายหรือถูกทำลาย บางครั้งค่าสูญหายเกิดจากความผิดพลาดในการทดลอง การเก็บข้อมูล ฯลฯ ดังนั้นการเกิดค่าสูญหายในแผนการทดลองแบบแลตติสกลายเป็นเรื่องธรรมดา การคำนวณค่าสูญหายใช้สมการ ดังนี้

$$\hat{X} = \frac{k^2T + k(k+1)B - R + G - kT_b - kB_t}{k(k-1)^2} \qquad \ldots(18\text{-}6)$$

ค่าต่าง ๆ ในสมการมีคำอธิบาย ดังนี้ :

$\hat{X}$ = ข้อมูลสูญหาย

T = ผลรวมของทรีตเมนต์ที่มีข้อมูลสูญหายจากซ้ำอื่น ๆ

B = ผลรวมของบลอกที่มีข้อมูลสูญหาย

R = ผลรวมของซ้ำที่มีข้อมูลสูญหาย

G = ผลรวมทั้งหมด

T_b = ผลรวม (จากทุกซ้ำ) ของทุกทรีตเมนต์ที่อยู่ในบลอกที่มีค่าสูญหาย

B_t = ผลรวมของทุกบลอกที่ทรีตเมนต์มีค่าสูญหายปรากฏขึ้น

สมมุติว่า T_3 ในซ้ำที่ 1 (คือ 5) เป็นค่าสูญหาย ถ้าจะคำนวณหาค่าสูญหายก็มีค่าต่าง ๆ ดังนี้

$T = 18,\ B = 11, R = 40, G = 198,\ T_b = 18+30+16 = 64\ ;\ B_t = 11+17+19+18 = 65$

แล้วนำค่าเหล่านี้ลงในสมการ (18-6) ได้ผลดังนี้

$$\hat{X} = \frac{(3^2)(18) + (3)(3+1)(11) - 40 + 198 - (3)(64) - (3)(65)}{(3)(3-1)^2} = 5.42$$

เมื่อได้ข้อมูลสูญหายแล้วก็ทำการวิเคราะห์ตามวิธีปกติ แต่ df ของ total และ intrablock จะลดลงไปเท่ากับจำนวนค่าสูญหาย (คือ 1) ในการทำ t-test ต้องประมาณจำนวนซ้ำสัมฤทธิ์ (effective number) เพื่อคำนวณ $s_{\bar{d}}$ ถ้าเปรียบเทียบระหว่าง A และ B ถ้า A เป็นทรีตเมนต์ที่มีค่าสูญหายให้กำหนดค่า ดังนี้

(1) สำหรับ A ในซ้ำที่ไม่มี A ให้ค่า 0 ซ้ำที่มี A ให้ค่า 1

(2) สำหรับ B ถ้าอยู่ในบลอกที่ไม่มี A ให้ค่า 0

(3) ถ้าผิดจากข้อ 1 – 2 ให้ค่า 1

เช่นถ้าต้องการเปรียบเทียบระหว่างทรีตเมนต์ 3 และ 5 ซึ่งข้อมูลทรีตเมนต์ที่ 3 ของซ้ำที่ 1 มีค่าสูญหาย

ทรีตเมนต์ 3 $0 + 1 + 1 + 1 = 3$ ซ้ำ

ทรีตเมนต์ 5 $0 + 1 + 1 + 1 = 3$ ซ้ำ

ถ้าเปรียบเทียบระหว่างทรีตเมนต์ 3 และ 4

ทรีตเมนต์ 3 $0 + 1 + 1 + 1 = 3$ ซ้ำ

ทรีตเมนต์ 4 $1 + 1 + 1 + 1 = 4$ ซ้ำ

ในกรณีที่มีค่าสูญหายหลายค่าแนะนำให้ใช้วิธีคำนวณซ้ำ ๆ หลายครั้ง ดังวิธีการที่ใช้ในแผนการทดลองแบบ RCB และนำผลที่ได้มาวิเคราะห์แบบ RCB ต่อไป

18.3 แผนการทดลองแบบแลตติสสมดุลบางส่วน

จากการทดลองแบบแลตติสสมดุลที่กล่าวมาแล้วในตอน 18.2 เห็นได้ว่ามีข้อเสีย คือ จำนวนซ้ำถูกกำหนดโดย $k+1$ ดังนั้นถ้ามีทรีตเมนต์มาก ๆ ก็ไม่สามารถทดลองได้ จึงได้มีผู้คิดค้นแผนการทดลองที่ใกล้เคียงมาใช้แทนโดยลดจำนวนซ้ำให้น้อยลง ถ้าใช้เพียง 2 ซ้ำแรก (ดูตาราง 18.2.1)

เรียกว่าแลตติสสองซ้ำ (double lattice หรือ simple lattice) ถ้าใช้ 3 ซ้ำ ก็เรียก แลตติสสามซ้ำ (triple lattice)

การทดลองแลตติสสองซ้ำ 1 ชุด

การทดลองที่มี 1 ชุด คือ การนำซ้ำที่ 1 ให้ชื่อว่า X และซ้ำที่สองให้ชื่อว่า Y ในตาราง 18.2.1 มาทำการทดลองเพียง 1 ครั้ง ก่อนทดลองทำการสุ่มตามปกติ คือ ทำการสุ่มเพื่อวางบลอกในแต่ละซ้ำ แล้วทำการสุ่มทรีตเมนต์ในแต่ละบลอก นำผลการสุ่มนี้ไปทำการทดลองต่อไป สมมุติว่าจากการทดลองได้ข้อมูลดังตาราง 18.3.2 ก็นำผลการทดลองมาวิเคราะห์วาเรียนซ์ ซึ่งแสดงแหล่งของความปรวนแปร ดังตาราง 18.3.1

ตาราง 18.3.1 แหล่งของความปรวนแปรของการทดลองแบบแลตติสสมดุลบางส่วนที่มี 9 ทรีตเมนต์

Sources	Df
Replications	$n-1 = 1$
Treatments (unadj.)	$k^2-1 = 8$
Blocks within rep (adj.)	$n(k-1) = 4$
Intrablocks	$(k-1)(nk-k-1) = 4$
	$nk^2-1 = 17$

ตัวอย่าง สมมุติว่าทำการทดลองเปรียบเทียบทรีตเมนต์ 9 ทรีตเมนต์ ดังตาราง 18.3.2

ตาราง 18.3.2 ผลการทดลองเปรียบเทียบ 9 ทรีตเมนต์โดยใช้แลตติสสมดุลบางส่วน 2 ซ้ำ 1 ครั้ง

บลอก	บลอก X			B_X	บลอก	บลอก Y			B_Y
1	(1)	(2)	(3)		1	(1)	(4)	(7)	
	9	3	5	17		7	6	8	21
2	(4)	(5)	(6)		2	(2)	(5)	(8)	
	4	5	2	11		4	7	3	14
3	(7)	(8)	(9)		3	(3)	(6)	(9)	
	8	4	3	15		4	5	4	13
				43					48

การวิเคราะห์วาเรียนซ์

การวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 18.3.2 กระทำเป็นขั้น ๆ ดังนี้

ขั้นที่ 1 คำนวณผลรวมและผลต่างของค่าต่าง ๆ ดังนี้

ผลรวมของบลอก X และ Y จากตาราง 18.3.2 เช่นได้ $B_{X_1} = 17, B_{Y_1} = 21$

ผลรวมของทรีตเมนต์ เช่น $T_1 = 9 + 7 = 16$

ผลรวมของทรีตเมนต์ที่จัดบลอกแบบ X เช่น $T_{X_1} = 16 + 7 + 9 = 32$

ผลรวมของทรีตเมนต์ที่จัดบลอกแบบ Y เช่น $T_{Y_1} = 16 + 10 + 16 = 42$

ความคลาดเคลื่อนเนื่องจากการวางบลอกแบบ X และ Y หรือค่าปรับ C คำนวณดังนี้ :

$C_{X_1} = T_{X_1} - 2B_{X_1} = 32 - 2(17) = -2$ ทั้งนี้ผลรวมของ C_X และ C_Y เท่ากับศูนย์ แล้วนำผลการคำนวณลงตาราง 18.3.2 และ 18.3.3 แล้วทำการวิเคราะห์วาเรียนซ์ ต่อไป

ขั้นที่ 2 คำนวณหาค่า sum of squares ต่าง ๆ

$$CF = \frac{(\text{ผลรวม})^2}{(n)(k^2)} = \frac{(91)^2}{18} = 460.06$$

$$TSS = 9^2 + 3^2 + \ldots + 4^2 - 460.16 = 529 - 460.06 = 68.94$$

$$\text{Rep SS} = \frac{\sum R^2}{k^2} - CF = \frac{43^2 + 48^2}{9} - 460.06 = 461.44 - 460.06 = 1.38$$

$$\text{SSTr (unadj.)} = \frac{\sum T_i^2}{n} - CF = \frac{16^2 + \ldots + 7^2}{2} - 460.06 = 516.50 - 460.06 = 56.44$$

$$\text{Block within rep (adj.) SS} = \frac{\sum C_x^2 + \sum C_y^2}{kn(n-1)} - \frac{(\sum C_x)^2 + (\sum C_y)^2}{k^2 n(n-1)} \qquad \ldots(18\text{-}7)$$

$$= \frac{(-2)^2 + \ldots + (-3)^2}{6} - \frac{5^2 + (-5)^2}{18} = 11.00 - 2.78 = 8.22$$

$$\text{Intrablock SS} = \text{TSS} - \text{RepSS} - \text{SSTr (unadj)} - \text{Blocks within rep SS}$$

$$= 68.94 - 1.38 - 56.44 - 8.22 = 2.90$$

ตาราง 18.3.3 ผลการทดลองและการเตรียมวิเคราะห์ในการทดลองแบบแลทติสสมดุลของส่วนที่มี 9 ทรีตเมนต์ 2 ซ้ำ 1 ชุด

บลอก	รวม X			T_X	$C_X = T_X - 2B_X$	μC_X
1	(1)	(2)	(3)			
	16	7	9	32	$-2 = 32 - 2(17)$	−0.44
2	(4)	(5)	(6)			
	10	12	7	29	$7 = 29 - 2(11)$	1.54
3	(7)	(8)	(9)			
	16	7	7	30	$0 = 30 - 2(15)$	0
				91	5	1.10

บลอก	รวม Y			T_Y	$C_Y = T_Y - 2B_Y$	
1	(1)	(4)	(7)			
	16	10	16	42	0 = 42 – 2(21)	0
2	(2)	(5)	(8)			
	7	12	7	26	–2 = 26 – 2 (14)	–0.44
3	(3)	(6)	(9)			
	9	7	7	23	–3 = 23 – 2(13)	–0.66
				91	–5	–1.10

ขั้นที่ 3 นำค่า sum squares ลงตารางวิเคราะห์วาเรียนซ์แล้วหาค่า mean squares ดังตาราง 18.3.4

ตาราง 18.3.4 ตารางวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 18.3.2

Sources	df	SS	MS
Replications	1	1.38	
Treatments (unadj.)	8	56.44	
Blocks within rep (adj.)	4	8.22	2.06 (E_b)
Intrablock error	4	2.90	0.73 (E_e)
Total	17	68.94	

ขั้นที่ 4 ปรับค่าเฉลี่ย

จากผลการวิเคราะห์วาเรียนซ์พบว่า $E_b > E_e$ ดังนั้นคำนวณค่า μ เพื่อจะนำไปปรับค่าเฉลี่ย

$$\mu = \frac{(E_b - E_e)}{k(n-1)E_b} \qquad \ldots(18\text{-}8)$$

$$= \frac{2.06 - 0.73}{(3)(2.06)} = \frac{1.33}{6.18} = 0.22$$

ต่อจากนั้นก็นำ μ มาคูณค่า C ของแต่ละบลอกแล้วปรับผลรวมของทรีตเมนต์ ดังนี้

$$T_i' = T_i + \mu C_X + \mu C_Y \qquad \ldots(18\text{-}9)$$

ทั้งนี้ T_i = ผลรวมของทรีตเมนต์ที่ไม่ปรับค่า

ดังนั้นหาได้ว่า

$$T_1' = 16 + (-2)(0.22) + (0)(0.22) = 15.56$$

$$T_2' = 7 - 0.44 - 0.44 = 6.12$$

$$T_3' = 9 - 0.44 - 0.66 = 7.90$$

$$T_4' = 10 + 1.54 - 0 = 11.54$$
$$T_5' = 12 + 1.54 - 0.44 = 13.10$$
$$T_6' = 7 + 1.54 - 0.66 = 7.88$$
$$T_7' = 16 + 0 + 0 = 16.00$$
$$T_8' = 7 + 0 - 0.44 = 6.56$$
$$T_9' = 7 + 0 - 0.66 = 6.34$$

ถ้าหากว่า E_b น้อยกว่า E_e ก็ถือว่าเท่ากับศูนย์ ไม่ต้องปรับผลรวมของทรีตเมนต์ และวิเคราะห์วาเรียนซ์แบบ RCB ปกติต่อไป

ตารางวิเคราะห์วาเรียนซ์ข้างบนไม่อาจใช้ทดสอบทรีตเมนต์หลังปรับค่าแล้ว แต่ถ้าจะทดสอบทรีตเมนต์ก่อนปรับค่าก็อาจใช้วิธีการวิเคราะห์ RCB ปกติ ซึ่งหาได้ว่า

	df	SS	MS	F
Treatments	8	56.44	7.06	5.08*
Error	8	11.12	1.39	

ขั้นที่ 5 ทดสอบผลของทรีตเมนต์ที่ปรับค่าแล้ว

คำนวณหา Treatment SS (adjusted)

ถ้าหากต้องการที่จะทดสอบทรีตเมนต์ภายหลังจากปรับค่าแล้ว อาจคำนวณหา SSTr(adj.) ดังนี้

$$\text{SSTr(adj.)} = \text{SSTr (unadj.)} - k(n-1)\,\mu\left[\frac{n}{1+k\mu}\right](B_\mu - B_a) \qquad \ldots(18\text{-}10)$$

ทั้งนี้
$$B_\mu = \frac{\sum B_X^2 + \sum B_Y^2}{k} - \frac{(\sum B_X)^2 + (\sum B_Y)^2}{k^2} \qquad \ldots(18\text{-}11)$$
$$= \frac{(17^2 + 11^2 + 15^2) + (21^2 + 14^2 + 13^2)}{(3)} - \frac{43^2 + 48^2}{9}$$
$$= 480.33 - 461.44 = 18.89$$

และ B_a = Blocks within rep SS = 8.22

ดังนั้นหาได้ว่า

$$\text{SSTr (adj)} = 56.44 - (3)(1)(0.22)\left[\frac{2}{1+(3)(0.22)}\right](18.90 - 8.22)$$
$$= 47.96$$

เมื่อดำเนินการทดสอบทรีตเมนต์ได้ผลดังนี้

	df	SS	MS	F
Treatments	8	47.96	5.99	8.20*
Intrablock error	4	2.90	0.73	

ซึ่งผลการทดสอบอยู่ในระดับเดียวกับการทดสอบทรีตเมนต์ที่มิได้ปรับค่า

ขั้นที่ 6 คำนวณความคลาดเคลื่อนเพื่อเปรียบเทียบค่าเฉลี่ย

ในการเปรียบเทียบค่าเฉลี่ย สามารถคำนวณวาเรียนซ์ของความแตกต่าง ($s_{\bar{d}}^2$) ดังนี้ :

ก. ทรีตเมนต์ในบลอกเดียวกัน

$$s_{\bar{d}}^2 = \frac{2E_e}{n}(1+\mu) = \frac{(2)(0.73)}{2}(1+0.22) = 0.89$$

ข. ทรีตเมนต์คนละบลอก

$$s_{\bar{d}}^2 = \frac{2E_e}{n}(1+n\mu) = \frac{(2)(0.73)}{2}(1+(2)(0.22)) = 1.05$$

ค. วาเรียนซ์เฉลี่ยทั่วไป

$$s_{\bar{d}}^2 = \frac{2E_e}{n}\left(1+\frac{nk\mu}{(k+1)}\right) = \frac{(2)(0.73)}{2}\left(1+\frac{(2)(3)(0.22)}{4}\right) = 0.97$$

ซึ่งให้ $s_{\bar{d}} = 0.94$, 1.02 และ 0.98 ตามลำดับ โดยปกติแล้วใช้ $s_{\bar{d}}$ เฉลี่ยสำหรับการเปรียบเทียบระหว่างคู่ของทรีตเมนต์

การประมาณการได้เปรียบว่าวิธีนี้เหนือแผนการทดสอบแบบ RCB ต้องคำนวณค่า effective error (E_e') ดังนี้

$$E_e' = E_e\left[1+\frac{nk\mu}{(k+1)}\right] = 0.73\left[1+\frac{(2)(3)(0.22)}{4}\right] = 0.97$$

เมื่อเปรียบเทียบค่า RCB ก็พบว่ามีข้อได้เปรียบ = (1.39/0.97) × 100 = 143%

ขั้นที่ 7 ค่าสูญหาย

เมื่อมีค่าสูญหายอาจคำนวณโดยใช้สมการ ดังนี้

$$\hat{X} = \frac{(n-1)k^2T - nR + G - nkC + kC'}{(n-1)(k-1)(nk-k-1)} \qquad \ldots(18.12)$$

กำหนดให้

$\hat{X}$ = ค่าสูญหาย

T = ผลรวมของทรีตเมนต์ที่สูญหายจากซ้ำอื่น ๆ

R = ผลรวมของซ้ำที่มีค่าสูญหาย

G = ผลรวมทั้งหมด

C = $T - 2B$ จากบลอกที่มีค่าสูญหาย

C' = ผลรวมของ C จากทุกบลอกที่มีทรีตเมนต์สูญหาย $(C_x + C_y)$

ตัวอย่าง จากตัวอย่างในตาราง 18.3.2 สมมุติว่าทรีตเมนต์ที่ 1 (T_1) ในกลุ่ม X เป็นค่าสูญหาย ดังนั้น

$T = 7 + X$

$R = 34$

$G = 82$

$C = C_X = T_X - 2B_X = 7 + 7 + 9 - 2(8) = 7$

$C' = C_X + C_Y = 7 + [(7 + 10 + 16) - 2(21)] = -2$

$$\hat{X} = \frac{(9)(7) - (2)(34) + 82 - (2)(3)(7) + (3)(-2)}{(2)(2)} = 7.25$$

18.4 การทดลองแบบแลตติสสองซ้ำสองชุด

การทดลองที่มี 2 ชุด คือทั้งซ้ำ X และ Y ต่างก็มี 2 ชุด จำนวนซ้ำก็เพิ่มเป็น 2 เท่า อาจแยกความปรวนแปรดังตาราง 18.4.1 ในตารางนี้จะเห็นได้ว่ามีการแยก block within rep (adj.) ออกเป็นชุด คือ Component (a) และ component (b) ทั้งนี้ component (a) คือความแตกต่างระหว่างบลอกเดียวกันแต่อยู่ในคนละชุด ส่วน component (b) คือความแตกต่างระหว่างบลอกในซ้ำเดียวกัน

ตาราง 18.4.1 แหล่งของความปรวนแปรและ df ในการทดลองแบบแลตติสที่มี 2 ซ้ำ จำนวน 2 ชุด

Sources*	Df
Replications	$2q - 1 = n - 1$
Treatments (unadj.)	$k^2 - 1$
Blocks within rep (adj.)	$2q(k - 1)$
Component (a)	$2(q - 1)(k - 1)$
Component (b)	$2(k - 1)$
Intrablock (Error)	$(k - 1)(2qk - k - 1)$
Total	$2qk^2 - 1$

* q = จำนวนชุด , n = จำนวนซ้ำ

ตัวอย่าง สมมติว่าได้ทำการทดลองเปรียบเทียบ 9 ทรีตเมนต์ โดยใช้แลตติสสองซ้ำจำนวน 2 ชุด ให้ผลดังตาราง 18.4.2

ตาราง 18.4.2 ผลการทดลองเปรียบเทียบ 9 ทรีตเมนต์โดยใช้แลททิสสมดุลบางส่วน จำนวน 2 ซ้ำ 2 ชุด

บลอก	X1			รวม	บลอก	X2			รวม	บลอก	Y1			รวม	บลอก	Y2			รวม
1	(1)	(2)	(3)		1	(1)	(2)	(3)		1	(1)	(4)	(7)		1	(1)	(4)	(7)	
	9	3	5	17		6	2	2	10		7	6	8	21		5	4	7	16
2	(4)	(5)	(6)		2	(4)	(5)	(6)		2	(2)	(5)	(8)		2	(2)	(5)	(8)	
	4	5	2	11		3	2	2	7		4	7	3	14		3	3	2	8
3	(7)	(8)	(9)		3	(7)	(8)	(9)		3	(3)	(6)	(9)		3	(3)	(6)	(9)	
	8	4	3	15		6	3	3	12		4	5	4	13		4	4	3	11
				43					29					48					35

ขั้นตอนการวิเคราะห์วาเรียนซ์

ขั้นที่ 1 จัดทำตารางผลรวมของบลอกและทรีตเมนต์

ผลรวมของบลอกแสดงไว้ในตาราง 18.4.2 ผลรวมของทรีตเมนต์แสดงในตารางดังต่อไปนี้ เช่น $T_1 = 9+6+7+5 = 27$

	(1)	(2)	(3)	รวม X
	27	12	15	54
	(4)	(5)	(6)	
	17	17	13	47
	(7)	(8)	(9)	
	29	12	13	54
รวม Y	73	41	41	155

แล้วคำนวณค่า sum of squares บางค่า ดังนี้

$$\text{CF} = \frac{(\text{ผลรวม})^2}{nk^2} = \frac{(155)^2}{36} = 667.36$$

$$\text{TSS} = 9^2+3^2+\ldots+3^2-\text{CF} = 797.00-667.36 = 129.64$$

$$\text{Rep SS} = \frac{43^2+\ldots+35^2}{k^2=9}-\text{CF} = 691.00-667.36 = 23.64$$

$$\text{SSTr (unadj.)} = \frac{27^2+\ldots+13^2}{(k+1)=4}-\text{CF} = 749.75-667.36 = 82.39$$

ขั้นที่ 2 คำนวณ SS ของ block เนื่องจากเป็นการทดลอง 2 ชุด แต่ละชุดมี 2 ซ้ำ จึงแยกผลของบลอกออกได้เป็น 2 ส่วน เราเรียก ดังนี้

Component (a) ความแตกต่างระหว่างบลอกในการทดลองชุดเดียวกัน เช่น ชุด X1 และ X2 ดังนั้น มี sum of square นี้ได้เมื่อมีการทดลองซ้ำในแต่ละชุดเท่านั้น

Component (b) ความแตกต่างระหว่างบลอกในการทดลองคนละชุดเหมือนการทดลองทั่วไป

(ก) Component (a) SS ได้จากความแตกต่างในกลุ่มบลอก X และ Y

X1 (ซ้ำที่ 1)	X2 (ซ้ำที่ 2)	d_X (ผลต่าง)	Y1	Y2	d_Y
17	10	7	21	16	5
11	7	4	14	8	6
15	12	3	13	11	2
43	29	14	48	35	14

ซึ่งหาต่อไปได้ว่า

$$\text{Component (a) SS} = \frac{\sum d_x^2 + \sum d_y^2}{2k} - \frac{(\sum d_x)^2 + (\sum d_y)^2}{2k^2} \qquad \ldots(18.13)$$

$$= \frac{7^2 + 4^2 + \ldots + 3^2}{(2)(3)} - \frac{14^2 + 14^2}{(2)(9)} = \frac{139}{6} - \frac{365}{18} = 2.89$$

(ข) Component (b) SS ได้จากความแตกต่างระหว่างบลอก X และ Y ในการหาค่านี้ ต้องมีการรวมทรีตเมนต์ที่เป็นบลอก X กลุ่มหนึ่ง และ Y อีกกลุ่มหนึ่ง แล้วหาความแตกต่าง ถ้าแต่ละกลุ่มนั้นได้จากกลุ่ม X อย่างเดียว และ Y อย่างเดียว ดังนั้น

ผลรวมกลุ่ม X			B_X (ผลรวม)	$C_X = T_X - 2B_X$
(1)	(2)	(3)		
15	5	7	27	0 = 54 − 2(27)
(4)	(5)	(6)		
7	7	4	18	11 = 47 − 2(18)
(7)	(8)	(9)		
14	7	6	27	0 = 54 − 2(27)
			72	11

ผลรวมกลุ่ม Y			B_Y (ผลรวม)	$C_Y = T_Y - 2B_Y$
(1)	(4)	(7)		
12	10	15	37	$-1 = 73 - 2(37)$
(2)	(5)	(8)		
7	10	5	22	$-3 = 41 - 2(22)$
(3)	(6)	(9)		
8	9	7	24	$-7 = 41 - 2(24)$
			83	-11

ซึ่งหาต่อไปได้ว่า

$$\text{Component (b) SS} = \frac{\sum C_x^2 + C_y^2}{nk} - \frac{(\sum C_x)^2 + (\sum C_y)^2}{nk^2} \qquad \ldots(18\text{-}14)$$

$$= \frac{0^2 + 11^2 + \ldots + (-7)^2}{(3)(4)} - \frac{11^2 + (-11)^2}{(9)(4)} = \frac{180}{12} - \frac{242}{36} = 8.28$$

ดังนั้น

Block within rep (adj.) SS = 2.89 + 8.28 = 11.17

คำนวณ Intrablock (error) SS ดังนี้ :

= TSS – Rep SS – SSTr (unadj.) – Block/rep SS

= 129.64 – 23.64 – 82.39 – 11.17 = 12.44

ขั้นที่ 3 นำผลลงตารางวิเคราะห์วาเรียนซ์ดังตาราง 18.4.3 แล้วคำนวณค่า mean square ต่าง ๆ ต่อไป

ตาราง 18.4.3 ตารางวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 18.4.2

Sources	Df	SS	MS
Replications	3	23.64	7.88
Treatments (unadj.)	8	82.39	
Block within rep (adj.)	8	11.17	1.39 (E_b)
Component (a)	4	2.89	
Component (b)	4	8.28	
Intrablock (Error)	16	12.44	0.78 (E_e)
Total	35	129.64	

อย่างไรก็ดี การทดสอบทางสถิติ เพื่อแสดงความแตกต่างระหว่างทรีตเมนต์ อาจใช้วิธีการวิเคราะห์แบบ Randomized Complete Block ตามปกติ โดยรวม SS ของ block/rep และ error เข้าด้วยกัน ดังนี้

Sources	df	SS	MS	F
Replication	3	23.64	7.88	
Treatments	8	82.39	10.30	10.51**
Error	24	23.61	0.98	
Total	35	129.64		

ซึ่งพบว่าทรีตเมนต์แตกต่างกันอย่างมีนัยสำคัญยิ่ง จึงไม่จำเป็นต้องทดสอบอื่นใดอีก

ขั้นที่ 4 การปรับค่าเฉลี่ย

จากผลการวิเคราะห์วาเรียนซ์ในตาราง 18.4.3 ถ้าหากว่า E_b น้อยกว่า E_e ก็ไม่จำเป็นต้องปรับค่าเฉลี่ยของทรีตเมนต์โดยสามารถนำค่าเฉลี่ยจากข้อมูลเดิมไปใช้ประโยชน์โดยตรง แต่ถ้า E_b มากกว่า E_e ต้องมีการปรับค่าโดยใช้สมการ

$$T' = T + \mu C_x + \mu C_y \qquad \ldots(18\text{-}15)$$

$$\mu = \frac{q(E_b - E_e)}{k(n-q)E_b + (q-1)E_e} \qquad \ldots(18\text{-}16)$$

$$= \frac{2(1.39-0.78)}{3[(4-2)(1.39)+(2-1)(0.78)]} = 0.11$$

ต่อไปก็หาได้ว่า

$$
\begin{aligned}
T_1 &= 27 + (0.11)(0) + (0.11)(-1) = 26.89\\
T_2 &= 12 + (0.11)(0) + (0.11)(-3) = 11.67\\
T_3 &= 15 + (0.11)(0) + (0.11)(-7) = 14.23\\
T_4 &= 17 + (0.11)(11) + (0.11)(-1) = 18.10\\
T_5 &= 17 + (0.11)(11) + (0.11)(-3) = 17.88\\
T_6 &= 13 + (0.11)(11) + (0.11)(-7) = 13.44\\
T_7 &= 29 + (0.11)(0) + (0.11)(-1) = 28.89\\
T_8 &= 12 + (0.11)(0) + (0.11)(-3) = 11.67\\
T_9 &= 13 + (0.11)(0) + (0.11)(-7) = 12.23
\end{aligned}
$$

ขั้นที่ 5 คำนวณ SSTr (adj.)

คำนวณค่า Treatment SS (adjusted) จากสมการ ดังนี้

$$\text{SSTr (adj.)} = \text{SSTr (unadj.)} - k\mu\left[\frac{n}{1+k\mu}(B_\mu - B_a)\right] \qquad \ldots(18\text{-}17)$$

โดยที่คำนวณ B_μ ได้จาก Component (b) SS (unadjusted)

$$B_\mu = \frac{\sum B_x^2 + \sum B_y^2}{nk} - \frac{(\sum B_x)^2 + (\sum B_y)^2}{nk^2}$$

$$= \frac{27^2 + 18^2 + \ldots + 24^2}{(4)(3)} - \frac{72^2 + 83^2}{(4)(9)}$$

$$= 350.92 - 335.36 = 15.56$$

$$B_a = \text{Component (a) SS} = 2.89$$

ดังนั้น

$$\text{SSTr (adj.)} = 82.39 - (3)(0.11)\left[\frac{4}{1+(3)(0.11)}(15.56) - 2.89\right]$$

$$= 82.39 - 12.57$$

$$= 69.82$$

$$\text{MSTr (adj.)} = \frac{69.82}{8} = 8.73$$

Effective error MS, E'_e

$$E'_e = E_e\left[1 + \frac{2k\mu}{k+1}\right] = 0.78\left[1 + \frac{(2)(3)(0.11)}{3+1}\right] = 0.91$$

$$F\text{-test} = \frac{\text{MSTr(adj)}}{E'_e} = \frac{8.73}{0.91} = 9.59^{**}(\text{df } 8, 16)$$

$$CV(\%) = \frac{\sqrt{\text{(Intrablock error Ms)}}}{\text{Grand mean}} \times 100 = \frac{\sqrt{0.78}}{4.31} \times 100 = 20.49$$

ขั้นที่ 6 คำนวณความคลาดเคลื่อนมาตรฐาน

ก่อนคำนวณความคลาดเคลื่อนมาตรฐานอาจหา MS error ต่าง ๆ ดังนี้

1. สำหรับทรีตเมนต์ในบลอกเดียวกัน

   $$\text{Error MS} = E_e(1+\mu) = 0.78(1+0.11) = 0.87$$

2. สำหรับทรีตเมนต์ต่างบลอก

   $$\text{Error MS} = E_b(1+n\mu\mu = 1.39[1+(4)(0.11)] = 2.00$$

3. สำหรับความแตกต่างระหว่างค่าเฉลี่ยของทรีตเมนต์ทั่วไป

$$\text{Error MS} = E_e\left[1+\frac{2k\mu}{k+1}\right] = 0.82\left[1+\frac{(2)(3)(0.11)}{4}\right]$$

$$= 0.96$$

การคำนวณ $s_{\bar{d}}$ สำหรับเปรียบเทียบค่าเฉลี่ย

$$s_{\bar{d}} = \sqrt{\frac{(2)\,\text{Error MS}}{n}}$$

เช่นในการเปรียบเทียบค่าเฉลี่ยของทรีตเมนต์ทั่วไป

$$s_{\bar{d}} = \frac{\sqrt{(2)(0.96)}}{\sqrt{4}} = 0.69$$

ขั้นที่ 7 ค่าสูญหาย

เมื่อมีค่าสูญหายก็ใช้วิธีการคำนวณเช่นเดียวกับการทดลองแบบ 1 ชุด โดยมีสมการ ดังนี้

$$\hat{X} = \frac{k^2T + nR + G - 2kC + kC' - 4R'}{(k-1)(4-k-1)} \qquad \ldots(18\text{-}18)$$

ทั้งนี้แต่ละค่ามีคำอธิบายที่สมการ (18-11) ยกเว้น R' ซึ่งเท่ากับผลรวมของซ้ำที่มีค่าสูญหาย วิธีการคำนวณคล้ายกับการทดลอง 1 ชุด ดังที่กล่าวมาแล้วในตอน 18.3

18.5 แผนการทดลองแบบแลตติสสามซ้ำ

แผนการทดลองแบบแลตติสสามซ้ำ คือการทดลองที่มี 3 ซ้ำ คือ X , Y และ Z ซึ่งอาจทำครั้งเดียวหรือ 2-3 ครั้งก็ได้ ทั้งนี้ซ้ำ X และ Y เหมือนกับแผนการทดลองแบบสองชุด แต่ซ้ำที่สามจัดสร้างโดยวิธีจัดทรีตเมนต์แบบละตินสแควร์ เช่นการทดลอง 16 ทรีตเมนต์ มี 3 ซ้ำ ดังนี้

บลอก	ซ้ำ x				บลอก	ซ้ำ y				บลอก	ซ้ำ z			
1	1	2	3	4	5	1	5	9	13	9	1	6	11	16
2	5	6	7	8	6	2	6	10	14	10	2	7	12	13
3	9	10	11	12	7	3	7	11	15	11	3	8	9	14
4	13	14	15	16	8	4	8	12	16	12	4	5	10	15

ทั้งนี้ซ้ำที่ 3 ได้จากการใส่ตัวอักษร A, B, C, D แบบละตินสแควร์ลงในซ้ำ X แล้วจัดระเบียบใหม่ตามตัวอักษร ดังนี้

ซ้ำ X + (A, B, C, D)				บลอก	ซ้ำ z			
1A	2B	3C	4D	9(A)	1	6	11	16
5D	6A	7B	8C	10(B)	2	7	12	13
9C	10D	11A	12B	11(C)	3	8	9	14
13B	14C	15D	16A	12(D)	4	5	10	15

เมื่อนำ A, B, C และ D มาเรียงไว้ในบรรทัดเดียวกันก็จะให้บลอก Z ดังแสดงไว้ในด้านขวามือนั่นเอง ต่อจากนั้นก็นำบลอกเหล่านี้ไปจัดสุ่มตามปกติแล้วนำไปใช้ประโยชน์ต่อไป

ตัวอย่าง ในการทดลองเปรียบเทียบ 16 ทรีตเมนต์ โดยใช้แลททิส 3 ซ้ำ 1 ชุด ดังแสดงในตาราง 18.5.1

การวิเคราะห์วาเรียนซ์

ขั้นที่ 1 จัดทำตารางผลรวมของบลอก X, Y, Z ผลรวมของทรีตเมนต์ ค่า Tx, Ty, Tz และคำนวณค่าต่าง ๆ ดังแสดงในตาราง 18.5.1 และ 18.5.2 แล้วคำนวณค่า sum of squares ต่าง ๆ

ตาราง 18.5.1 ผลการทดลองเพื่อเปรียบเทียบ 16 ทรีตเมนต์ โดยใช้แลตติส 3 ซ้ำ 1 ชุด

บลอก	ซ้ำ X				รวม	บลอก	ซ้ำ Y				รวม	บลอก	ซ้ำ Z				รวม
2	(6)	(8)	(5)	(7)		8	(8)	(12)	(16)	(4)		9	(1)	(11)	(16)	(6)	
	3	4	3	5	15		1	2	4	2	9		1	2	4	2	9
4	(13)	(16)	(15)	(14)		5	(9)	(13)	(5)	(1)		11	(8)	(14)	(3)	(9)	
	1	3	2	2	8		2	3	1	2	8		1	1	2	1	5
1	(1)	(3)	(2)	(4)		6	(14)	(2)	(6)	(10)		12	(4)	(10)	(15)	(5)	
	1	2	1	3	7		4	3	2	4	13		3	2	5	2	12
3	(10)	(12)	(9)	(11)		7	(11)	(3)	(15)	(7)		10	(7)	(2)	(13)	(12)	
	4	3	2	1	10		4	5	3	4	16		4	2	2	5	13
รวม					40	รวม					46	รวม					39

วิเคราะห์

$$CF = \frac{(\text{ผลรวม})^2}{(n)(k^2)} = \frac{(125)^2}{(3)(16)} = 325.52$$

$$TSS = 3^2 + 4^2 + \ldots + 5^2 - 325.52 = 399 - 325.52 = 73.48$$

ตาราง 18.5.2 การเตรียมการวิเคราะห์ข้อมูลในตาราง 18.5.1

บลอก	รวม X				T_x	$C_x = T_x - 3B_x$	μC_x
1	(1)	(2)	(3)	(4)			
	4	6	9	8	27	6 = 27 – 3(7)	0.576
2	(5)	(6)	(7)	(8)			
	6	7	13	6	32	–13 = 32 – 3(15)	–1.248
3	(9)	(10)	(11)	(12)			
	5	10	7	10	32	2 = 32 – 3(10)	0.192
4	(13)	(14)	(15)	(16)			
	6	7	10	11	34	10 = 34 – 3(8)	0.960
					125	5	
	รวม Y				T_y	$C_y = T_y - 3B_y$	μC_y
5	(1)	(5)	(9)	(13)			
	4	6	5	6	21	–3 = 21 – 3(8)	–0.288
6	(2)	(6)	(10)	(14)			
	6	7	10	7	30	–9 = 30 – 3 (13)	–0.864
7	(3)	(7)	(11)	(15)			
	9	13	7	10	39	–9 = 39 – 3(16)	–0.864
8	(4)	(8)	(12)	(16)			
	8	6	10	11	35	8 = 35 – 3(9)	0.768
					125	–13	
	รวม Z				T_z	$C_z = T_z - 3B_z$	μC_z
9	(1)	(6)	(11)	(16)			
	4	7	7	11	29	2 = 29 – 3(9)	0.192
10	(2)	(7)	(12)	(13)			
	6	13	10	6	35	–4 = 35 – 3 (13)	–0.384
11	(3)	(8)	(9)	(14)			
	9	6	5	7	27	12 = 27 – 3(5)	1.152
12	(4)	(5)	(10)	(15)			
	8	6	10	10	34	–2 = 34 – 3(12)	–0.192
					125	8	

$$\text{Rep SS} = \frac{\sum R^2}{k^2} - CF = \frac{40^2 + 46^2 + 39^2}{16} - 325.52$$

$$= 327.31 - 325.52 = 1.79$$

$$\text{SSTr(unadj.)} = \frac{\sum T_i^2}{n} - CF = 355.67 - 325.52 = 30.15$$

Block within rep. SS (adj) = Component (a) SS + Component (b) SS

เนื่องจากเป็นการทดลอง 1 ชุด ดังนั้น Component (a) = 0 ส่วน

$$\text{Com(b) SS} = \frac{\sum C_x^2 + \sum C_y^2 + \sum C_z^2}{2kn} - \frac{(\sum C_x)^2 + (\sum C_y)^2 + (\sum C_z)^2}{2k^2 n} \quad \ldots(18\text{-}19)$$

$$= \frac{6^2 + (-13)^2 + \ldots + (-2)^2}{(2)(4)(3)} - \frac{5^2 + (-13)^2 + 8^2}{(2)(16)(3)}$$

$$= 29.67 - 2.69 = 26.98$$

$$\text{Intrablock SS} = \text{TSS} - \text{Rep SS} - \text{SSTr (unadj)} - \text{Com (b) SS}$$

$$= 73.48 - 1.79 - 30.15 - 26.98 = 14.56$$

นำค่า SS ที่สามารถหาได้ลงตาราง 18.5.3 เพื่อคำนวณค่า MS ต่าง ๆ ต่อไป

ตาราง 18.5.3 ผลการวิเคราะห์วาเรียนซ์ของข้อมูลในตาราง 18.5.1

Source	df*		SS	MS
Replications	$n - 1$	= 2	1.79	0.89
Treatments (unadj)	$k^2 - 1$	= 15	30.15	2.01
Block within rep (adj)	$3q(k - 1)$	= 9	26.98	3.00 (E_b)
Component (a)	$3(q - 1)(k - 1)$	= 0	0	
Componen (b)	$3(k - 1)$	= 9	26.98	
Intrablock (error)	$(k - 1)(3qk - k - 1)$	= 21	14.56	0.69 (E_e)
	$3qk^2 - 1$	= 47	73.48	

*n = จำนวนซ้ำ, q = จำนวนกลุ่ม, k = จำนวนทรีตเมนต์ต่อบลอก

ในการทดลองแบบแลตติสไม่อาจทดสอบผลของทรีตเมนต์แบบ F-test จากตาราง 18.5.3 ได้โดยตรง แต่อาจทดสอบโดยใช้การวิเคราะห์แบบ RCB ซึ่งสามารถหาได้ว่า MSTr = 30.15/15 = 2.01, MSE = 41.54/30 = 1.38 ดังนั้น F = 2.01/1.38 = 1.45

ในการทดลองนี้พบว่า $E_b > E_a$ ดังนั้น อาจคำนวณค่า μ ดังนี้

$$\mu = \frac{q(E_b - E_e)}{k[(n-q)E_b + (q-1)E_e]} = \frac{1(3.00-0.69)}{4(2)(3.00)+(0)(0.69)} = 0.096$$

ซึ่งสามารถปรับค่าของทรีตเมนต์ (หา T') ได้ดังนี้

$$T_i' = T + \mu C_x + \mu C_y + \mu C_z$$

$$T_1' = 4+(0.096)(6)+(0.096)(-3)+(0.096)(2) = 4.48$$

$$T_2' = 6+(0.096)(6)+(0.096)(-9)+(0.096)(-4) = 5.328$$

$$\vdots$$

$$T_{16}' = 11+(0.096)(10)+(0.096)(8)+(0.096)(2) = 12.92$$

คำนวณ SSTr (adj) จากสมการ ดังนี้

$$\text{SSTr (adj)} = \text{SSTr(unadj.)} - 2k\mu\left[\frac{3}{2[1+k\mu]}\right](B_\mu - B_a)$$

$$B_\mu = \text{Block within group SS}$$

$$= \frac{15^2+8^2+\ldots+13^2}{4} - \frac{40^2+46^2+39^2}{16}$$

$$= 356.75 - 327.31 = 29.44$$

$$B_a = \text{Component (b) SS} = 26.98$$

ดังนั้นอาจหาได้ว่า

$$\text{SSTr(adj)} = 30.15-(2)(4)(0.096)\left[\frac{3}{2[(1)+(4)(0.096)]}(29.44-26.98)\right]$$

$$= 28.10$$

$$\text{MSTr (adj.)} = \frac{28.10}{15} = 1.87$$

ทดสอบ MSTr (adj.) ด้วย E_e'

$$E_e' = E_e\left[1+\frac{3k\mu}{(k+1)}\right] = 0.69\left[1+\frac{(3)(4)(0.096)}{5}\right] = 0.85$$

ดังนั้น

$$F = \frac{1.87}{0.85} = 2.2^{ns} \quad (\text{df } 9, 21)$$

ถ้าหากว่าทรีตเมนต์แตกต่างกัน เราดำเนินการเปรียบเทียบค่าเฉลี่ย โดยคำนวณ $s_{\bar{d}}$ ดังนี้

(1) ทรีตเมนต์ในบลอกเดียวกัน

$$s_{\bar{d}} = \sqrt{[(2E_e / n)(1 + 2\mu)]} = \sqrt{[(2)(0.69)/3][(1 + (2)(0.19)]} = 0.80$$

(2) ทรีตเมนต์คนละบลอก

$$s_{\bar{d}} = \sqrt{[(2E_e / n)(1 + 3\mu)]} = \sqrt{[(2)(0.69)/3][1 + (3)(0.19]} = 0.85$$

(3) ทรีตเมนต์ทั่วไป

$$s_{\bar{d}} = \sqrt{[(2E_e / n)(1 + 3k\mu / k + 1)]} = \sqrt{[(2)(0.69)/3][1 + (3)(4)(0.19)/5]} = 0.82$$

การประเมินความเที่ยงตรงเมื่อเปรียบเทียบกับ RCB ปกติ

$$\text{อัตราเที่ยงตรง} = \frac{MSE(RCB)}{E'_e} = \frac{[SS(Block+Intrablock)/df(Block+Intrablock)]}{E'_e}$$

$$= \frac{[(26.98 + 14.56)/(9 + 21)]}{0.85} = \frac{1.38}{0.85} = 163\%$$

ซึ่งหมายความว่าแผนการทดลองแบบแลตติสให้ผลดีกว่าการทดลองแบบ RCB

บทที่ 19
แมตริกซ์

19.1 คำนำ

ความเข้าใจเกี่ยวกับแมตริกซ์ นับว่าจำเป็นสำหรับการศึกษารีเกรซชันในเบื้องสูงขึ้นไป โดยเฉพาะอย่างยิ่งเมื่อมีตัวแปรเกิน 1 ชุด ซึ่งมีวิธีการวิเคราะห์ที่สลับซับซ้อนเกินกว่าที่จะใช้วิธีปกติที่ใช้ทั่วไป ดังนั้นบทนี้เป็นการศึกษาถึงวิธีการทางแมตริกซ์โดยย่อ ๆ ศัพท์และหลักวิธีที่ไม่เกี่ยวข้องกับการศึกษาในเรื่องรีเกรซชันจะไม่รวมไว้ในที่นี้ ผู้สนใจอาจหาอ่านหรือศึกษาเพิ่มเติมได้จากหนังสือเกี่ยวกับเรื่องนี้โดยเฉพาะ

19.2 แมตริกซ์ [1]

ในวิชาคำนวณสาขาต่าง ๆ มีอยู่บ่อยครั้งที่เราพบว่า มีการจัดข้อมูลในแบบมีแถวและสดมภ์ ดังนี้

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & a_{mn} \end{bmatrix}$$

เลขที่จัดระเบียบข้อมูลแบบนี้เรียกว่าแมตริกซ์

ขนาดของแมตริกซ์ถือตามจำนวนแถวและสดมภ์ เลขที่ของแถวแทนด้วย i และสดมภ์แทนด้วย j ทั้งนี้ i คือแถวที่ 1 ถึง m และ j คือสดมภ์ที่ 1 ถึง n ดังนั้นแมตริกซ์ขนาด $m \times n$ คือแมตริกซ์ที่มี m แถวและ n สดมภ์ เช่น แมตริกซ์ที่มี 3 แถว 3 สดมภ์ เรียกว่าแมตริกซ์ขนาด 3×3 เราอาจเขียนแมตริกซ์โดยใช้สัญลักษณ์ว่า $A = (a_{ij})(m, n)$ คือแมตริกซ์ A มีสมาชิกย่อย ๆ a_{ij} และมีขนาด $m \times n$ สมาชิกในแมตริกซ์ หรือ a_{ij} นิยมเรียกว่าเอลิเมนต์[2] เพื่อแสดงตัวอย่างง่าย ๆ เกี่ยวกับแมตริกซ์จึงยกสมการดังนี้

$$2x + 6y = 1$$
$$4x - y = -3$$

เราอาจนำค่าจริงหรือสัมประสิทธิ์ของ x และ y มาแสดงในรูปแมตริกซ์ขนาด 2×2 ได้ดังนี้

$$\begin{bmatrix} 2 & 6 \\ 4 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$$

ค่า 2, 6, 4 และ −1 เรียกว่าสมาชิกของแมตริกซ์

แมตริกซ์ที่มีจำนวนแถวและจำนวนสดมภ์เท่ากัน เรียกว่าแมตริกซ์จัตุรัส[3] แมตริกซ์จัตุรัสยังมีหลายชนิดย่อย ๆ ที่ควรรู้จักคือ

(1) แมตริกซ์ทแยง[4] คือแมตริกซ์จัตุรัสที่สมาชิกนอกเส้นทแยงมุมเป็นศูนย์ เช่น

$$\begin{bmatrix} a_{11} & 0 & \ldots\ldots & 0 \\ 0 & a_{22} & \ldots\ldots & a \\ . & . & . & . \\ 0 & 0 & \ldots\ldots & a_{mn} \end{bmatrix} \quad \text{ตัวอย่าง} \quad \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

(2) แมตริกซ์สเกลาร์[5] คือแมตริกซ์ทแยงที่มีทุกค่าเท่ากันหมด

(3) แมตริกซ์เอกลักษณ์[6] คือแมตริกซ์ทแยงที่มีทุกค่าเท่ากับ 1 ซึ่งใช้สัญลักษณ์ว่า $A = (a_{ij})(m = n = I)$ ดังตัวอย่าง

แมตริกซ์สเกลาร์	แมตริกซ์เอกลักษณ์
$\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

19.3 การทำงานเกี่ยวกับแมตริกซ์

การย้ายรูปแมตริกซ์

ในการกระทำกับแมตริกซ์ อาจมีการย้ายแถวไปเป็นสดมภ์ ย้ายสดมภ์เป็นแถว การกระทำเช่นนี้เรียกว่าการย้ายรูปแมตริกซ์หรือการทรานสโพสแมตริกซ์[7] การใช้ย้ายรูปแมตริกซ์ใช้สัญลักษณ์ว่า A^T ดังนี้

แมตริกซ์		ย้ายรูป
$[A_{ij}]mn$	$\longrightarrow$	$[A_{ij}^{T}]nm$

ตัวอย่าง เช่นแมตริกซ์เหล่านี้เมื่อย้ายรูปจะได้

$$\overset{\mathbf{A}}{\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}} \longrightarrow \overset{\mathbf{A}^{T}}{\begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}}$$

เช่น
$$\begin{bmatrix} 2 & 6 & -1 \\ 4 & -1 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & 4 \\ 6 & -1 \\ -1 & 3 \end{bmatrix}$$

แมตริกซ์ที่ย้ายรูปแล้วยังมีค่าของสมาชิกเหมือนเดิมเรียกว่าแมตริกซ์สมมาตร[8] เช่น

$$\begin{bmatrix} 4 & 2 & -1 \\ 2 & 6 & 12 \\ -1 & 12 & 8 \end{bmatrix} \longrightarrow \begin{bmatrix} 4 & 2 & -1 \\ 2 & 6 & 12 \\ -1 & 12 & 8 \end{bmatrix}$$

ซึ่งจะเห็นว่าแม้มีการย้ายรูปแล้ว ก็ยังมีสมาชิกเหมือนเดิม

การคูณแมตริกซ์

ถ้ามีแมตริกซ์อยู่ 2 ชนิด คือ $A = (a_{ij})(3 \times 2)$ และ $B = (a_{ij})(2 \times 2)$ เช่น

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

เมื่อนำมาคูณกันในรูป $A \times B$ ก็จะได้แมตริกซ์ C ดังนี้

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} \end{bmatrix}$$

$$AB = C$$

วิธีการคูณโดยย่อ ๆ ก็คือนำค่าที่ 1 ในแถว 1 คูณค่าที่ 1 ในสดมภ์ 1 บวกกับค่าที่ 2 ในแถวที่ 1 คูณด้วยค่าที่ 2 ในสดมภ์ที่ 1 กระทำเช่นนี้กระทั่งทุกแถวและทุกสดมภ์ได้คูณกันจนครบ ซึ่งเราอาจจะ แสดงได้โดยวิธีใช้ตัวเลขได้ดังนี้

$$\begin{bmatrix} 3 & 0 & -1 \\ 0 & 1 & 1 \\ 2 & 4 & -3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 4 & -3 & 2 \\ 6 & 6 & 1 \end{bmatrix} = \begin{bmatrix} -3 & -6 & 2 \\ 10 & 3 & 3 \\ 0 & -30 & 7 \end{bmatrix}$$

ซึ่งแต่ละค่าได้จากการใช้แถวคูณสดมภ์โดยเริ่มจากแถวที่ 1, 2, 3 ของ A คูณสดมภ์ที่ 1 ของ B

$$\begin{aligned} (3)(1) + (0)(4) + (-1)(6) &= -3 \\ (0)(1) + (1)(4) + (1)(6) &= 10 \qquad \text{สดมภ์ที่ 1 ของ C} \\ (2)(1) + (4)(4) + (-3)(6) &= 0 \end{aligned}$$

แถว 1, 2, 3 ของ A คูณสดมภ์ที่ 2 ของ B

$$\begin{aligned} (3)(0) + (0)(-3) + (-1)(6) &= -6 \\ (0)(0) + (1)(-3) + (1)(6) &= 3 \qquad \text{สดมภ์ที่ 2 ของ C} \\ (2)(0) + (4)(-3) + (-3)(6) &= -30 \end{aligned}$$

แถว 1, 2, 3 ของ A คูณสดมภ์ที่ 3 ของ B

$$\begin{aligned} (3)(1) + (0)(2) + (-1)(1) &= 2 \\ (0)(1) + (1)(2) + (1)(1) &= 3 \qquad \text{สดมภ์ที่ 3 ของ C} \\ (2)(1) + (4)(2) + (-3)(1) &= 7 \end{aligned}$$

จากวิธีการเช่นเดียวกันอาจหาได้ว่า

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} 22 & 28 \\ 49 & 64 \\ 76 & 100 \end{bmatrix}$$

$$AB = C$$

เช่นทุกตัวในแถวที่ 3 ของแมตริกซ์ A (7, 8, 9) คูณกับทุกตัวในสดมภ์ที่ 2 ของแมตริกซ์ B แล้วบวกกันดังนี้

$$(7)(2) + (8)(4) + (9)(6) = 100$$

ซึ่งเป็นสมาชิกในแถวที่ 3 และสดมภ์ที่ 2 ของแมตริกซ์ C จึงอาจสรุปได้ว่าแมตริกซ์ที่จะคูณกันได้นั้นจำนวนสดมภ์ของแมตริกซ์ A ต้องเท่ากับจำนวนแถวของแมตริกซ์ B

สมการต่าง ๆ ในทางพีชคณิต อาจแทนด้วยผลคูณของแมตริกซ์

$$\begin{bmatrix} 5 & -2 & 1 \\ 2 & 1 & -5 \\ 4 & -2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ -6 \\ 1 \end{bmatrix}$$

เมื่อทำการคูณตามหลักที่อธิบายมาแล้วก็จะได้

$$\begin{aligned} 5x - 2y + z &= 3 \\ 2x + y - 5z &= -6 \\ 4x - 2y + z &= 1 \end{aligned}$$

จากการคูณที่ต้องนำเอาแถวของแมตริกซ์ A ไปคูณกับสดมภ์ของแมตริกซ์ B นี่เองสามารถแสดงได้ว่า A คูณ B ได้ผลลัพธ์ไม่เท่ากับ B คูณ A ($AB \neq BA$) ดังตัวอย่าง

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$$

$$AB = C$$

เมื่อนำแมตริกซ์ B มาไว้ข้างหน้าก็จะได้

$$\begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \neq \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$$

การอินเวอร์สแมตริกซ์

การอินเวอร์ส[9] แมตริกซ์ หมายถึงการกลับแมตริกซ์ เช่น แมตริกซ์ A เมื่ออินเวอร์สแล้วก็จะได้แมตริกซ์ A^{-1} แมตริกซ์ที่นำมาอินเวอร์สต้องเป็นแมตริกซ์จัตุรัสเท่านั้น

แมตริกซ์ที่ได้รับการอินเวอร์สแล้ว เมื่อนำไปคูณกับแมตริกซ์เดิม ก็จะได้แมตริกซ์เอกลักษณ์ (I) ดังนี้

$$AA^{-1} = A^{-1}A = I$$

เช่น แมตริกซ์

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \text{เมื่ออินเวอร์สก็ได้} \quad A^{-1} = \begin{bmatrix} -2.0 & 1.0 \\ 1.5 & -0.5 \end{bmatrix}$$

เมื่อคูณ $A \times A^{-1}$ ก็ได้แมตริกซ์ C ซึ่งเป็นแมตริกซ์เอกลักษณ์ ดังนี้

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} -2.0 & 1.0 \\ 1.5 & -0.5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$AA^{-1} = I$$

การอินเวอร์สแมตริกซ์ 2 x 2

การอินเวอร์สแมตริกซ์ 2×2 มีวิธีการดังนี้ เช่น มี

แมตริกซ์ A, $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ ก็อาจอินเวอร์สโดยใช้ขั้นตอนดังต่อไปนี้

(1) สับเปลี่ยนตำแหน่งสมาชิกทแยง คือเปลี่ยนตำแหน่ง a กับ d ให้ b และ c อยู่ที่เดิม แต่ให้ใส่เครื่องหมายลบข้างหน้า b และ c ดังนี้

เช่น $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ เมื่อเปลี่ยนตำแหน่งก็ได้ $\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

(2) หารแต่ละสมาชิกในข้อ 1 ด้วยความแตกต่างของผลคูณ ad - bc เมื่อ $ad - bc \neq 0$

$$A^{-1} = \begin{bmatrix} \dfrac{d}{ad-bc} & \dfrac{-b}{ad-bc} \\ \dfrac{-c}{ad-bc} & \dfrac{a}{ad-bc} \end{bmatrix}$$

เช่น $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $ad - bc = (1)(4) - (2)(3) = -2$

ดังนั้น $A^{-1} = \begin{bmatrix} \dfrac{4}{-2} & \dfrac{-2}{-2} \\ \dfrac{-3}{-2} & \dfrac{1}{-2} \end{bmatrix} = \begin{bmatrix} -2.0 & 1.0 \\ 1.5 & -0.5 \end{bmatrix}$

เช่นนี้เราจะได้ $A^{-1}A = AA^{-1} = I$ (คือเท่ากับแมตริกซ์เอกลักษณ์)

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} -2.0 & 1.0 \\ 1.5 & -0.5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

ตัวอย่างข้างล่างเป็นตัวเลขชุด X, Y ซึ่ง Y มีรีเกรซชันต่อ X จงคำนวณหาค่า a และ b และเขียนสมการรีเกรซชันเส้นตรง

X	Y
1	10
2	11
3	13
4	17
5	19

วิธีทำ เมื่อนำตัวเลขข้างบนมาจัดเข้าสมการ ก็จะได้ดังนี้

$$Y_i = a + bX_i + d_{y.x}$$

$$10 = a + b(1) + d_{y.x}$$

$$11 = a + b(2) + d_{y.x}$$
$$13 = a + b(3) + d_{y.x}$$
$$17 = a + b(4) + d_{y.x}$$
$$19 = a + b(5) + d_{y.x}$$

ซึ่งอาจถอดสมการข้างบนออกมาในรูปแมตริกซ์คูณเป็น

$$\begin{bmatrix} 10 \\ 11 \\ 13 \\ 17 \\ 19 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} + d_{y.x(i)}$$

ซึ่งอาจเขียนได้ว่า

$$Y = X\begin{bmatrix} a \\ b \end{bmatrix} + d_{y.x(i)}$$

คูณด้วย X^T ทั้งสมการจะได้ $X^TY = (X^TX)\begin{bmatrix} a \\ b \end{bmatrix} + X^Td_{y.x(i)}$

ทั้งนี้อาจพิสูจน์ได้ว่า $X^Td_{y.x(i)} = 0$ ดังนั้น

$$X^TY = X^TX\begin{bmatrix} a \\ b \end{bmatrix}$$

หรือ $$\begin{bmatrix} a \\ b \end{bmatrix} = (X^TX)^{-1}(X^TY)$$

ทั้งนี้ $$X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}$$

จึงหาได้ว่า

$$X^TY = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}\begin{bmatrix} 10 \\ 11 \\ 13 \\ 17 \\ 19 \end{bmatrix} = \begin{bmatrix} 70 \\ 234 \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum XY \end{bmatrix}$$

$$X^TX = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 5 & 15 \\ 15 & 55 \end{bmatrix} = \begin{bmatrix} n & \sum X \\ \sum X & \sum X^2 \end{bmatrix}$$

ต่อจากนี้ก็อินเวอร์ส X^TX จนได้ $(X^TX)^{-1}$ โดยวิธีการดังนี้

1. สับเปลี่ยนตำแหน่งในแถวทแยงและเปลี่ยนเครื่องหมายบวกเป็นลบ

$$\begin{bmatrix} 5 & 15 \\ 15 & 55 \end{bmatrix} \longrightarrow \begin{bmatrix} 55 & -15 \\ -15 & 5 \end{bmatrix}$$

2. หารแต่ละค่าด้วย $(55)(5)-(15)(15) = 50$ ดังนั้น

$$(X^TX)^{-1} = \begin{bmatrix} \frac{55}{50} & \frac{-15}{50} \\ \frac{-15}{50} & \frac{5}{50} \end{bmatrix}$$

จากสมการ $\begin{bmatrix} a \\ b \end{bmatrix} = (X^TX)^{-1}(X^TY)$

ดังนั้น $\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \frac{55}{50} & \frac{-15}{50} \\ \frac{-15}{50} & \frac{5}{50} \end{bmatrix} \begin{bmatrix} 70 \\ 234 \end{bmatrix} = \begin{bmatrix} \frac{55(70)}{50} & \frac{-15(234)}{50} \\ \frac{-15(70)}{50} & \frac{5(234)}{50(1)} \end{bmatrix}$

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 6.8 \\ 2.4 \end{bmatrix}$$

ดังนั้น $a = 6.8,\ b = 2.4$

ซึ่งหากนำค่าในแมตริกซ์ (1) มาแทนด้วยสัญลักษณ์ก็จะได้

$$a = \frac{(\sum X^2)(\sum Y) - (\sum X)(\sum XY)}{n\sum X^2 - (\sum X)^2} \qquad \ldots(19\text{-}1)$$

$$b = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2} \qquad \ldots(19\text{-}2)$$

ซึ่งจะเห็นได้ว่าสมการ (19-2) มีรูปเหมือนกับสมการ (5-12) ทุกประการ

ตัวอย่าง จากสมการ

$$5x - 2y = 4$$

$$x + y = 5$$

จงหาค่า x และ y

วิธีทำ

จากสมการข้างบนจะได้แมตริกซ์

$$\begin{bmatrix} 5 & -2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

$$[A]\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = A^{-1}\begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} 1/7 & 2/7 \\ -1/7 & 5/7 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1/7 & 2/7 \\ -1/7 & 5/7 \end{bmatrix}\begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

$$= \begin{bmatrix} 4/7 + 10/7 \\ -4/7 + 25/7 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

ดังนั้นจึงหาได้ว่า x = 2, y = 3

การอินเวอร์สแมตริกซ์ขนาด 3 × 3 มีวิธีการที่ซับซ้อนกว่าเดิม อย่างไรก็ดีเราอาจเขียนเป็นสูตรได้ดังนี้

จากแมตริกซ์

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix} \quad \text{เมื่ออินเวอร์สแล้วจะได้} \quad A^{-1} = \begin{bmatrix} A & B & C \\ D & E & F \\ G & H & K \end{bmatrix}$$

ต่อไปก็หาได้ว่า

$$\begin{array}{lll} A = (ek - fh)/Z & B = -(bk - ch)/Z & C = (bf - ce)/Z \\ D = -(dk - fg)/Z & E = (ak - cg)/Z & F = -(af - cd)/Z \\ G = (dh - eg)/Z & H = -(ah - bg)/Z & K = (ae - bd)/Z \end{array}$$

โดยที่ $Z = a(ek - fh) - b(dk - fg) + c(dh - eg)$

ตัวอย่าง เมื่อ $A = \begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}$ จงหา A^{-1}

วิธีทำ

$$Z = 2(4-3) - 3(2-9) + 1(1-6) = 18$$

$$\begin{array}{lll} A = (4-3)/18 = 1/18 & B = -(6-1)/18 = -5/18 & C = (9-2)/18 = 7/18 \\ D = -(2-9)/18 = 7/18 & E = (4-3)/18 = 1/18 & F = -(6-1)/18 = -5/18 \\ G = (1-6)/18 = -5/18 & H = -(2-9)/18 = 7/18 & K = (4-3)/18 = 1/18 \end{array}$$

ดังนั้น $$A^{-1} = \begin{bmatrix} 1/18 & -5/18 & 7/18 \\ 7/18 & 1/18 & -5/18 \\ -5/18 & 7/18 & 1/18 \end{bmatrix}$$

19.4 การอินเวอร์สแมตริกซ์โดยวิธี abbreviated Doolittle

การอินเวอร์สแมตริกซ์นั้นมีวิธีการแตกต่างกันหลายวิธี แมตริกซ์ที่มีขนาดเกิน 2 แถวและ 2 สดมภ์ มีวิธีการยุ่งยากยิ่งขึ้น การอินเวอร์สแมตริกซ์โดยวิธี abbreviated Doolittle นับเป็นวิธีการที่เหมาะสมสำหรับการศึกษาเกี่ยวกับมัลทิเปิลรีเกรซชัน[10] ทั้งนี้เพราะทำให้สามารถคำนวณหาดรรชนีรีเกรซชันต่าง ๆ ได้ทันที วิธีการโดยละเอียดแสดงไว้ในตาราง 19.4.1 การอินเวอร์สแมตริกซ์วิธีนี้มี 2 ขั้นตอน คือ forward solution และ backward solution คือการกระทำไปข้างหน้าและการกระทำย้อนกลับ

ตาราง 19.4.1 แสดงขั้นตอนของ forward solution แมตริกซ์ A เป็นแมตริกซ์สมมาตร ในวิธีการนี้ใช้เฉพาะซีกบนแมตริกซ์ ส่วนวิธีการอินเวอร์สโดยละเอียดแสดงไว้ในสดมภ์แรกของตารางดังกล่าว ตัวอย่างการอินเวอร์สแสดงไว้ในตาราง 19.4.2

ตาราง 19.4.1 การอินเวอร์สแมตริกซ์สมมาตร โดยวิธี abbreviated Doolittle

บรรทัด	วิธีการ	แมตริกซ์ A			แมตริกซ์ I		
01	วางแมตริกซ์ A และ	a_{11}	a_{12}	a_{13}	1	0	0
02	แมตริกซ์ I		a_{22}	a_{23}	0	1	0
03				a_{33}	0	0	1
11	ลอกบรรทัดที่ 01	a_{11}	a_{12}	a_{13}	1	0	0
12	a_{11} หารตลอด	1	b_{12}	b_{13}	b_{14}	0	0
21	บรรทัด 02 – $a_{12}(b_{1j})$		c_{12}	c_{13}	c_{14}	1	0
22	c_{12} หารตลอด		1	d_{13}	d_{14}	d_{15}	0
31	บรรทัด 03 – $a_{13}(b_{1j})$ – $c_{13}(d_{1j})$			e_{13}	e_{14}	e_{15}	1
32	c_{13} หารตลอด			1	f_4	f_{15}	f_{16}

สมมติว่าให้ค่าซึ่งได้รับการอินเวอร์ส = C_{ij} โดยที่ i และ j มีค่าจาก 0 ถึง 2, อาจเขียนได้ว่า

$$(X^TX)^{-1} = \begin{bmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix} \qquad \ldots(19\text{-}3)$$

ในการหาค่า C_{ij} ใช้วิธี backward solution ซึ่งหาได้จาก

$$\begin{aligned} C_{00} &= (e_{14})(f_{14}) + (c_{14})(d_{14}) + (1)(b_{14}) \\ C_{01} &= (e_{14})(f_{15}) + (c_{14})(d_{15}) \\ C_{02} &= (e_{14})(f_{16}) \\ C_{11} &= (e_{15})(f_{15}) + (1)(d_{15}) \\ C_{12} &= (e_{15})(f_{16}) \\ C_{22} &= (1)(f_{16}) \end{aligned} \qquad \ldots(19\text{-}4)$$

ทั้งนี้ $C_{01} = C_{10}$, $C_{02} = C_{20}$ และ $C_{12} = C_{21}$

โดยการแทนค่าในสมการเหล่านี้ ทำให้เราทราบค่า C_{ij} ทุก ๆ ค่า แล้วไปแทนลงใน $(X^T X)^{-1}$ ตามสมการ (19-3) ก็จะเป็นค่าอินเวอร์สทั้งหมด

เพื่อให้เข้าใจได้ดียิ่งขึ้น เราอาจจะศึกษาวิธีการคำนวณได้จากตัวอย่างในตาราง 19.4.2

*วิธีการ (1) จากบรรทัดที่ 02 มี 10 16 0 1 0

10 − (4)(2) = 2 (4 บรรทัด 11, 2 จาก บรรทัด 12)

16 − (4)(3) = 4 (4 " 11, 3 " 12)

0 − (4)(0.50) = −2 (4 " 11, 0.50 " 12)

1 − (4)(0) = 1 (4 " 11, 0 " 12)

ซึ่งเป็นค่าในบรรทัด 21

(บรรทัด 02 ลบด้วย 4 ซึ่งคูณด้วยค่าอื่นในบรรทัด 12)

ตาราง 19.4.2 การอินเวอร์สแมตริกซ์ขนาด 3 x 3 โดยวิธี abbreviated Doolittle

บรรทัด	แมตริกซ์ A			แมตริกซ์ I			วิธีการ
01	2	4	6	1	0	0	
02		10	16	0	1	0	
03			30	0	0	1	
11	2/	4	6	1	0	0	ลอกบรรทัดที่ 01 ทั้งหมด
12	1	2	3	0.50	0	0	หารด้วย 2/ ตลอด
21		2	4	−2	1	0	*ดูวิธีการ (1) ท้ายตาราง
22		1	2	−1	0.50	0	หาร 2 ตลอด
31			4	1	−2	1	**ดูวิธีการ (2) ท้ายตาราง
32			1	0.25	−0.50	0.25	หาร 4 ตลอด

**วิธีการ (2) จากบรรทัด 03 มี 30 0 0 1

30 − (6)(3) − (4)(2) = 4

0 − (6)(0.50) − (4)(−1) = 1

0 − (6)(0) − (4)(0.50) = −2

1 − (6)(0) − (4)(0) = 1

ซึ่งเป็นค่าในบรรทัด 31

(บรรทัด 03 ลบด้วย 6 คูณค่าอื่น ๆ ในบรรทัด 12 และลบด้วย 4 ซึ่งคูณค่าอื่น ๆ ในบรรทัด 22)

วิธีการคำนวณ C_{ij} ดูตารางและสมการ (19-4) ดังนี้

$$
\begin{aligned}
C_{00} &= (e_{14})(f_{14}) + (c_{14})(d_{14}) + (1)(b_{14}) \\
&= (1)(0.25) + (-2)(-1) + (1)(0.50) \\
&= 2.75 \\
C_{01} &= (e_{14})(f_{15}) + (c_{14})(d_{15}) \\
&= (1)(-0.50) + (-2)(0.50) \\
&= -1.50 \\
C_{02} &= (e_{14})(f_{16}) \\
&= (1)(0.25) \\
&= 0.25 \\
C_{11} &= (e_{15})(f_{15}) + (1)(d_{15}) \\
&= (-2)(-0.50) + (1)(0.50) \\
&= 1.50 \\
C_{12} &= (e_{15})(f_{16}) \\
&= (-2)(0.25) \\
&= -0.50 \\
C_{22} &= (1)(f_{16}) \\
&= (1)(0.25) \\
&= 0.25
\end{aligned}
$$

ค่า C_{ij} ทางมุมล่างด้านซ้ายของ $(X^TX)^{-1}$ หาได้ดังกล่าวมาแล้ว คือ

$C_{10} = -1.50,\ C_{20} = 0.25,\ C_{21} = -0.50$

ดังนั้นจาก

$$X^TX = \begin{bmatrix} 2 & 4 & 6 \\ 4 & 10 & 16 \\ 6 & 16 & 30 \end{bmatrix} \text{ จะได้ } (X^TX)^{-1} = \begin{bmatrix} 2.75 & -1.50 & 0.25 \\ -1.50 & 1.50 & -0.50 \\ 0.25 & -0.50 & 0.25 \end{bmatrix}$$

โดยที่ $(X^TX)(X^TX)^{-1} = I$

19.5 แบบฝึกหัด

1. จงหาผลคูณของแมตริกซ์ดังต่อไปนี้

   ก. $\begin{bmatrix} 3 & 1 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} 4 & 2 \\ -7 & 1 \end{bmatrix}$ ข. $\begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 5 & 6 & 7 \\ 1 & 2 & 3 \end{bmatrix}$

   ค. $\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ ง. $\begin{bmatrix} 1 & 2 & 3 \\ 7 & -8 & 9 \end{bmatrix}\begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$

2. จงหาค่า x และ y จากสมการ

   ก. $x + y = 5$, $x - y = 1$ ข. $3x - y = 3$, $2x + y = 7$

3. จงคำนวณหา a, b_1 และ b_2 จากข้อมูลต่อไปนี้ โดยวิธีอินเวอร์สแมตริกซ์ (ตอน 15.5)

Y	X_1	X_2
2	1	1
4	1	4
4	3	2

4. จงคำนวณหาค่า a และ b จากตัวเลขที่มีความสัมพันธ์กัน 2 ชุดข้างล่าง โดยวิธีแมตริกซ์

   ก.

X (อายุ)	35	45	55	65	75
Y (ความดันโลหิต)	114	124	143	158	166

   ข.

X (อายุต้นถั่วเหลืองเป็นอาทิตย์)	1	2	3	4	5	6	7
Y (ความสูง - ซม.)	5	13	16	23	33	38	40

5. จงอินเวอร์สแมตริกซ์ ต่อไปนี้โดยวิธี abbreviated doolittle

   ก. $\begin{matrix} 2 & 6 & 8 \\ 6 & 10 & 20 \\ 8 & 20 & 30 \end{matrix}$ ข. $\begin{matrix} 3 & 5 & 7 \\ & 11 & 11 \\ & & 21 \end{matrix}$

คำในบท

(1) matrix, (2) element, (3) square matrix, (4) diagonal matrix, (5) scalar matrix, (6) identity matrix, (7) transpose matrix, (8) symmetric matrix, (9) inverse, (10) multiple regression.

บทที่ 20
มัลทิเปิลรีเกรซชัน

20.1 คำนำ

ในบทที่ 5 เราได้ศึกษาถึงการที่ Y มีรีเกรซชันต่อ X เพียงชุดเดียว เช่น ในกรณีที่น้ำหนักของคนมีรีเกรซชันต่อความสูง เป็นต้น อย่างไรก็ดี บางครั้ง Y อาจขึ้นต่อ X หลายชุดก็ได้ เช่น ผลผลิตของพืช (Y) อาจมีรีเกรซชันต่อปริมาณปุ๋ย (X_1) และอัตราปลูก (X_2) และอัตราการเจริญเติบโตของสัตว์เลี้ยง (Y) ขึ้นอยู่กับน้ำหนักสัตว์เมื่อเริ่มต้นทดลอง (X_1) และปริมาณอาหารที่ให้แก่สัตว์ (X_2) เหล่านี้ เป็นต้น เมื่อตัวแปรชนิดหนึ่ง (Y) ขึ้นอยู่กับตัวแปรอื่น ๆ หลายชุด (X_i) เราก็เรียกว่าเป็นความสัมพันธ์แบบมัลทิเปิลรีเกรซชัน[1] การที่ Y ขึ้นต่อ X_i หลายชุดนี้ เราไม่สามารถเขียนเส้นตรงได้เหมือนรีเกรซชันเส้นตรง แต่เป็นโครงสร้างที่มีความกว้างและความลึกเป็นลักษณะพื้นผิว การศึกษาเรื่องมัลทิเปิลเกรซชันมีวัตถุประสงค์ เพื่อจะสร้างสมการเส้นตรงสำหรับทำนายค่า Y จากค่า X เหล่านั้น นอกจากนั้นยังอาจศึกษาต่อไปได้อีกว่าค่า X ชุดใดบ้างที่เกี่ยวข้องและไม่เกี่ยวข้องกับค่า Y

การศึกษาเกี่ยวกับมัลทิเปิลรีเกรซชันมีวิธีการที่สลับซับซ้อน มีขั้นตอนการคำนวณที่ยืดยาว โดยเฉพาะเมื่อมีค่า X หลายค่า จึงเป็นการยากที่จะหลีกเลี่ยงความผิดพลาดถึงแม้มีการใช้เครื่องคิดเลขก็ตาม อย่างไรก็ดีในปัจจุบันนี้ได้มีการพัฒนาโปรแกรมคอมพิวเตอร์ขึ้นมาเพื่อช่วยให้การคำนวณถูกต้องยิ่งขึ้น แต่ผู้ศึกษาเกี่ยวกับเรื่องนี้ต้องมีความเข้าใจถึงกระบวนการและขั้นตอนในการคำนวณบนกระดาษ เพื่อให้เข้าใจวิธีการที่แท้จริง

20.2 สมการมัลทิเปิลรีเกรซชัน

ในกรณีที่ Y มีรีเกรซชันต่อค่า X หลายชุด เช่น น้ำหนักของคน (Y) ขึ้นอยู่กับอายุ (X_1) และความสูง (X_2) ในการสำรวจข้อมูลแล้วบันทึกเป็นค่าน้ำหนักดังแสดงในตาราง 20.2.1

ตาราง 20.2.1 ข้อมูลที่สัมพันธ์กันในแบบมัลทิเปิลรีเกรซชัน เช่น น้ำหนัก (Y) ต่ออายุ (X_1) และความสูง (X_2) ของคน

น้ำหนัก (Y)	อายุ (X_1)	ความสูง (X_2)
Y_1	X_{11}	X_{21}
Y_2	X_{12}	X_{22}
Y_3	X_{13}	X_{23}
•	•	•
•	•	•
•	•	•
Y_m	X_{1m}	X_{2m}

โดยให้ j = 1, 2, 3, ..., n (จำนวนกลุ่มตัวแปร X) และ i = 1, 2, 3 ,..., m (จำนวนสมาชิกของแต่ละตัวแปร) สมการรีเกรซชันของข้อมูลนี้ได้แก่

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 \qquad \text{...(20-1)}$$

ทั้งนี้ $\hat{Y}$ คือค่า Y ที่คำนวนหรือประมาณ b_1 และ b_2 เป็นค่าที่ทำให้ $\sum(Y_i - Y_c)^2$ มีค่าต่ำที่สุด ส่วนสมการรีเกรซชันของประชากรคือ

$$Y_p = \beta_0 + \beta_1X_1 + \beta_2X_2 \qquad \text{...(20-2)}$$

เมื่อให้ Y_p แทนค่าเฉลี่ยของ Y สำหรับแต่ละค่า X_1 และ X_2 ส่วน β_1 จะเป็นตัวบอกว่า เมื่อ X_1 เปลี่ยนแปลงไป 1 หน่วย ค่า Y จะเปลี่ยนไปเท่ากับ β_1 หน่วย โดยที่ X_2 ไม่เปลี่ยนแปลง ความหมายของ β_2 ก็อาจกล่าวได้ในทำนองเดียวกันกับ β_1 นั่นเอง ค่า β_i เหล่านี้เรียกว่า สัมประสิทธิ์ของรีเกรซชัน ซึ่งประมาณได้จาก b_1 และ b_2 ตามลำดับ

ในประชากร ค่าสังเกตแต่ละค่าแทนด้วยสมการ

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \varepsilon_i \qquad \text{...(20-3)}$$

ทั้งนี้ ε คือความคลาดเคลื่อน $\varepsilon \sim NID(0, \sigma^2)$

ส่วนค่าสังเกตในตัวอย่างอาจเขียนได้ว่า

$$Y_i = b_0 + b_1X_1 + b_2X_2 + d_{y.x} \qquad \text{...(20-4)}$$

โดยที่ $b_0 = \overline{Y} - b_1\overline{X}_1 - b_2\overline{X}_2$ ...(20-5)

ดังนั้นอาจเขียนสมการ (20-4) ใหม่เป็น

$$Y_i = \overline{Y} + b_1(X_i - \overline{X}_1) + b_2(X_2 - \overline{X}_2) + d_{y..x}$$

หรือ

$$Y_i = \overline{Y} + b_1(X_1 - \overline{X}_1) + b_2(X_2 - \overline{X}_2) + d_{y..x} \quad \ldots(20\text{-}6)$$

20.3 การประมาณพารามิเตอร์

ในการวิเคราะห์ข้อมูลโดยใช้มัลทิเปิลรีเกรซชันนั้น เบื้องต้นต้องมีการประมาณพารามิเตอร์ $\beta_1, \beta_2, .., \beta_n$ เสียก่อน ซึ่งอาจกระทำได้โดยใช้วิธีแก้สมการซ้อน[2] หรืออาจใช้วิธีแมตริกซ์ก็ได้

ในเรื่องรีเกรซชันเส้นตรงหาได้ว่า

$$b = SP/SSx$$

ทั้งนี้

$$SP = \sum(X_i - \overline{X})(Y_i - \overline{Y}) \quad \ldots(20\text{-}7)$$

$$SSx = \sum(X_i - \overline{X})^2$$

จากสมการนี้ก็ได้

$$b(SSx) = SP$$

หรือแทนค่าใหม่ทั้งด้านซ้าย - ด้านขวาก็ได้

$$b\sum(X_i - \overline{X})^2 = \sum(X_i - \overline{X})(Y_i - \overline{Y}) \quad \ldots(20\text{-}8)$$

เมื่อแก้สมการ (20-8) เราก็ทราบค่า b

อย่างไรก็ดี ในเรื่องมัลทิเปิลรีเกรซชัน เราต้องแก้หลายสมการพร้อมๆ กัน เพื่อให้ทราบ $b_1, b_2, ..., b_n$ สมมุติว่ามีตัวแปรอิสระจำนวน 2 ชุด (X_1 และ X_2) จากสมการ (20-6) อาจเขียนเสียใหม่ว่า

$$b_1(X_1 - \overline{X}_1) + b_2(X_2 - \overline{X}_2) + d_{y.x} = Y_i - \overline{Y} \quad \ldots(20\text{-}9)$$

เมื่อคูณสมการนี้ด้วย $\sum(X_1 - \overline{X}_1)$ ก็ได้

$$b_1\sum(X_1 - \overline{X}_1)^2 + b_2\sum(X_1 - \overline{X}_1)(X_2 - \overline{X}_2) + \sum(X_1 - \overline{X})d_{y.x} = \sum(X_1 - \overline{X}_1)(Y_i - \overline{Y})$$

แต่ $\sum(X_1 - \overline{X}_1)d_{y.x} = 0$ ดังนั้นก็จะได้

$$b_1\sum(X_1 - \overline{X}_1) + b_2\sum(X_1 - \overline{X}_1)(X_2 - \overline{X}_2) = \sum(X_1 - \overline{X}_1)(Y_i - \overline{Y}) \quad \ldots(20\text{-}10)$$

เมื่อคูณสมการ (20-9) ด้วย $\sum(X_2-\bar{X}_2)$ ก็จะได้

$$b_1\sum(X_1-\bar{X}_1)(X_2-\bar{X}_2) + b_2\sum(X_2-\bar{X}_2)^2 = \sum(X_2-\bar{X}_2)(Y_i-\bar{Y}) \qquad ...(20\text{-}11)$$

เมื่อแก้สมการ (20-10) และ (20-11) ด้วยวิธีทางพีชคณิตก็จะได้ค่า b_1 และ b_2 ดังตัวอย่างที่แสดงไว้ในตาราง 16.3.1

ตาราง 20.3.1 มัลทิเปิลรีเกรซชันที่มีตัวแปร 2 ชุด พร้อมค่า sum of square และ sum of product

	Y	X_1	X_2	$(X_1-\bar{X}_1)^2$	$(X_2-\bar{X}_2)^2$	$(X_1-\bar{X}_1)(Y_i-\bar{Y})$	$(X_2-\bar{X}_2)(Y_i-\bar{Y})$	$(X_1-\bar{X}_1)(X_2-\bar{X}_2)$
	13	3	3	1	4	−4	−8	2
	1	2	6	4	1	14	−7	−2
	6	4	5	0	0	0	0	0
	11	7	6	9	1	9	3	3
	10	4	5	0	0	0	0	0
ผลรวม	40	20	25	14	6	19	−12	3
				$\sum X_1^2$	$\sum X_2^2$	$\sum X_1Y$	$\sum X_2Y$	$\sum X_1X_2$

จากตาราง 20.3.1 หาได้ว่า $\bar{X}_1, \bar{X}_2$ และ $\bar{Y} = 4, 5$ และ 8 ตามลำดับ แล้วคำนวณค่าต่าง ๆ ได้ตามต้องการ ถ้ากำหนดให้ $\sum(X_1-\bar{X}_1)^2 = \sum X_1^2$, $\sum(X_2-\bar{X}_2)^2 = \sum X_2^2$, $\sum(X_1-\bar{X}_1)(Y_i-\bar{Y}) = \sum X_1Y$, $\sum(X_2-\bar{X}_2)(Y_i-\bar{Y}) = \sum X_2Y$, $\sum(X_1-\bar{X}_1)(X_2-\bar{X}_2) = \sum X_1X_2$

เมื่อนำไปแทนในสมการ (20-10) และ (20-11) ก็ได้

$$\sum X_1^2 b_1 + \sum X_1X_2 b_2 = X_1Y$$
$$\sum X_1X_2 b_1 + \sum X_2^2 b_2 = X_2Y$$

เมื่อนำค่าที่คำนวณได้ดังแสดงบรรทัดสุดท้ายของตาราง 20.3.1 แทนค่าในสมการ (20-10) และ (20-11) ก็ได้

$$14b_1 + 3b_2 = 19$$
$$3b_1 + 6b_2 = -12$$

ซึ่งคำนวณได้ว่า $b_1 = 2$ และ $b_2 = -3$ ส่วน b_0 หาได้โดยใช้สมการ (20-5) ดังนี้

$$b_0 = 8-(2)(4)-(-3)(5)$$
$$= 15$$

ดังนั้นสมการรีเกรซชันของตาราง 20.3.1 ก็คือ

$$Y_c = 15 + 2X_1 - 3X_2$$

การแก้สมการเพื่อหาค่า $b_1, b_2, ..., b_n$ โดยวิธีนี้จะประสบความยุ่งยากและมีวิธีคำนวณยืดยาว โดยเฉพาะอย่างยิ่งหากว่ามีค่า X หลายชุด จึงได้นำวิธี abbreviated Doolittle ดังได้อธิบายไว้ในตอน 19.4 มาช่วยในการคำนวณหาค่า b_i และวิเคราะห์ข้อมูล ซึ่งจะทำการแสดงเพื่อจุดประสงค์อันนี้โดยเฉพาะอีกครั้งหนึ่ง ดังนี้

จากสมการ (20-1) เราอาจดัดแปลงเป็น

$$Y_c = b_0 + b_1X_1 + b_2X_2 + ... + b_nX_n \qquad ...(20\text{-}12)$$

ในทำนองเดียวกัน จากสมการ (20-4) จะได้

$$Y_i = b_0 + b_1X_1 + b_2X_2 + ... + b_nX_n + d_{y..x} \qquad ...(20\text{-}13)$$

ซึ่งถ้าขยายสมการ (20-13) เป็นรูปแมตริกซ์จะได้

$$Y_1 = b_0 + b_1X_{11} + b_2X_{12} + ... + b_nX_{1n} + d_{y..x(1)}$$

$$Y_2 = b_0 + b_1X_{21} + b_2X_{22} + ... + b_nX_{2n} + d_{y.x(2)}$$

$$.. \quad .. \quad ..... \quad .... \quad ..... \quad .... \quad ....$$

$$Y_m = b_0 + b_1X_{m1} + b_2X_{m2} + ... + b_nX_{mn} + d_{y.x(m)}$$

ซึ่งถ้าเขียนสมการในรูปแมตริกซ์ก็จะได้

$$\begin{bmatrix} Y_1 \\ Y_2 \\ .. \\ Y_m \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & ...X_{1n} \\ 1 & X_{21} & X_{22} & ...X_{2n} \\ .. & .... & ... & .... \\ 1 & X_{m1} & X_{m2} & ...X_{mn} \end{bmatrix} x \begin{bmatrix} b_0 \\ b_1 \\ .. \\ b_u \end{bmatrix} + d_{y.x(i)}$$

หรืออาจเขียนเป็นสมการว่า

$$Y = (X)(b) + d_{y.x(i)}$$

เมื่อคูณสมการนี้ด้วย X^T ทั้งสองด้านก็ได้

$$X^TY = (X^TX)(b) + X^Td_{y..x(i)}$$

แสดงได้ว่า $X^Td_{y..x(i)} = 0$ ดังนั้นก็จะได้สมการ

$$b = (X^TX)^{-1}(X^TY) \qquad ...(20\text{-}14)$$

ทั้งนี้อาจแสดงได้ว่า

$$X^TX = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{21} & \cdots & X_{m1} \\ \cdots & \cdots & \cdots & \cdots \\ X_{1n} & X_{2n} & \cdots & X_{mn} \end{bmatrix} \begin{bmatrix} 1 & X_{11} & \cdots & X_{1n} \\ 1 & X_{21} & \cdots & X_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & X_{m1} & \cdots & X_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} n & \sum X_{i1} & \cdots & \sum X_{in} \\ \sum X_{i1} & \sum X^2_{i1} & \cdots & \sum X_{i1}X_{in} \\ \cdots & \cdots & \cdots & \cdots \\ \sum X_{in} & \sum X_{i1}X_{in} & \cdots & \sum X^2_{mn} \end{bmatrix}$$

โดยวิธีเดียวกับเราจะได้

$$X^TY = \begin{bmatrix} \sum Y_i \\ \sum X_{i1}Y_i \\ \cdots\cdots \\ \sum X_{in}Y_i \end{bmatrix}$$

ค่า X^TX เป็นแมตริกซ์สมมาตร จึงแยกมาใช้เพียงด้านเดียว เช่น ถ้ามีตัวแปร 3 ชุด (X_1 , X_2 , และ X_3) เราอาจจำวิธีหาเอเลเมนต์ในแมตริกซ์ดังข้างบนโดยใช้ตารางง่ายๆ ดังนี้

	X_0	X_1	X_2	X_3	Y
คูณด้วย $\sum X_0$	n	$\sum X_1$	$\sum X_2$	$\sum X_3$	$\sum Y$
คูณด้วย $\sum X_1$	$\sum X_1$	$\sum X_1^2$	$\sum X_1X_2$	$\sum X_1X_3$	$\sum X_1Y$
คูณด้วย $\sum X_2$	$\sum X_2$	$\sum X_1X_2$	$\sum X_2^2$	$\sum X_2X_3$	$\sum X_2Y$
คูณด้วย $\sum X_3$	$\sum X_3$	$\sum X_1X_3$	$\sum X_2X_3$	$\sum X_3^2$	$\sum X_3Y$

ซึ่ง X_0 ไม่มีค่า แต่ $\sum X_0^2 = n$, ค่าในตารางข้างบนเกิดจากค่าในแถวบนสุด คูณด้วยค่าในสดมภ์ที่หนึ่ง

ในกรณีของข้อมูลในตาราง 20.3.1 หาได้ว่า $n = 5, \sum X_1 = 20 \ \sum X_2 = 25, \sum Y = 40,$ $\sum X_1X_2 = 103, \sum X_1^2 = 94, \sum X_2^2 = 131, \sum X_1Y = 179, \sum X_2Y = 188$ ซึ่งสามารถกลับแมตริกซ์โดยวิธี abbreviated Doolittle ดังตาราง 19.4.2

จากตาราง 20.3.2 อาจแสดงหรือหาได้ว่า

$$b_2 = \quad -3$$

b_1 = 19/14 – (3/14) b_2 (ทั้งนี้ 19/14 และ 3/14 ได้จากบรรทัด 22)

= 19/14 – (3/14) (–3)

= 2

$$b_0 = \sum Y/n - (\sum X_2)/n\,(b_2) - (\sum X_1/n)\,(b_1)$$

= (40/5) – (–3)(25/5) – (2)(20/5)

= 15

ซึ่งจะเห็นได้ว่าค่าต่าง ๆ ที่ได้เหมือนกับที่หามาแล้วทุกประการ ข้อดีของวิธีอินเวอร์สแมตริกซ์เมื่อมีแมตริกซ์ G ร่วมด้วยนี้ ทำให้เราทราบ sum of squares เนื่องจากรีเกรซชัน (SSR) ได้ทันที โดยการนำค่าขวามือสุดของตาราง 20.3.2 ในบรรทัด 21, 22 และ 31, 32 มาคูณแล้วบวกกันดังนี้

SSR = (19) (19/14) + (–225/14) (–3)

= 74

ตาราง 20.3.2 การอินเวอร์สแมตริกซ์โดยวิธี abbreviated Doolittle

บรรทัด		แมตริกซ์ A			แมตริกซ์
		X_0	X_1	X_2	Y
01	X_0	n	$\sum X_1$	$\sum X_2$	$\sum Y$
02	X_1		$\sum X_1^2$	$\sum X_1X_2$	$\sum X_1Y$
03	X_2			$\sum X_2^2$	$\sum X_2Y$
01		5	20	25	40
02			94	103	179
03				131	188
11		5	20	25	40
12		1	4	5	8
21			14	3	19
22			1	$\frac{3}{14}$	$\frac{19}{14}$
31				$\frac{75}{14}$	$-\frac{225}{14}$
32				1	–3 (b_2)

ซึ่งใช้สำหรับการวิเคราะห์วาเรียนซ์ที่จะอธิบายในตอนต่อไป

20.4 การทดสอบสมมุติฐานในมัลทิเปิลรีเกรซชัน

ถ้าเปิดกลับไปดูตอน 5.8 เห็นได้ว่าความปรวนแปรของ Y มาจากแหล่งต่าง ๆ สองแหล่งด้วยกันคือ ส่วนที่เกิดจากความคลาดเคลื่อนของค่าทำนายจากค่าสังเกต และส่วนที่เกิดจากรีเกรซชัน ซึ่งอาจจะเขียนได้ว่า

$$\sum(Y_i-\overline{Y})^2 = \sum(Y_i-Y_c)^2+\sum(Y_c-\overline{Y})^2$$

$$\text{TSS} = \text{SSE}+\text{SSR}$$

ค่า TSS คำนวณโดยวิธีปกติ ส่วน SSE อาจคำนวณโดยใช้วิธีหาค่าทำนาย (Y_c) ทั้งหมดลบจาก Y_i ยกกำลังสองแล้วบวกกัน อย่างไรก็ดีในทางปฏิบัติมักจะคำนวณ SSR หรือ sum of squares เนื่องจากรีเกรซชัน แล้วหักลบจาก TSS ก็จะได้ SSE โดยที่ $\sum(Y_c-\overline{Y})^2$ หรือ

$$\text{SSR} = b_1\sum(X_1-\overline{X}_1)(Y_i-\overline{Y})+b_2\sum(X_2-\overline{X}_2)(Y_i-\overline{Y})$$

$$= b_1\left[\sum X_{1j}Y_i - \frac{(\sum X_{1j})(Y_i)}{m}\right] + b_2\left[\sum X_{2j}Y_i - \frac{(\sum X_{2j})(Y_i)}{m}\right]$$

เพื่อให้ดูง่ายขึ้นก็อาจ

$$= b_1\left[\sum X_1Y - \frac{(\sum X_1)(Y)}{m}\right] + b_2\left[\sum X_2Y - \frac{(\sum X_2)(\sum Y)}{m}\right]$$

ดังนั้นถ้าคำนวณโดยวิธีปกติก็ได้ผลดังนี้

$$\text{SSE} = \sum(Y_i-\overline{Y})^2-\left[b_1\sum(X_1-\overline{X}_1)(Y_i-\overline{Y})+b_2\sum(X_2-\overline{X}_2)(Y_i-\overline{Y})\right]$$

$$= \text{TSS} - \text{SSR}$$

TSS มี df = m −1 เมื่อ m เป็นจำนวนค่า Y_i ในมัลทิเปิลรีเกรซชันนั้น df ของ SSR = k = จำนวนชุดของตัวแปร ดังนั้น SSE จึงมี df = m − k − 1

จุดประสงค์ในการแยกความคลาดเคลื่อนออกเป็นส่วนย่อย ๆ ก็เพื่อที่จะทดสอบว่าในบรรดาค่า Y ขึ้นอยู่กับค่า X ชุดใดบ้าง อันหมายถึง β บางตัวหรือทั้งหมดเป็นศูนย์ ดังนั้นในการทดสอบจึงตั้งสมมุติฐาน (H_0) ว่า $\beta_1 = \beta_2 = 0$

ตาราง 20.4.1 การวิเคราะห์วาเรียนซ์ในมัลทิเปิลรีเกรซชัน

Sources	df	SS
Regression	k	$b_1\sum(X_1-\overline{X}_1)(Y_i-\overline{Y})+b_2\sum(X_2-\overline{X}_2)(Y_i-\overline{Y})$
Deviation (error)	$m-k-1$	TSS − SSR
Total	$m-1$	$\sum(Y_i-\overline{Y})^2$

จากข้อมูลในตาราง 20.3.1

$$
\begin{aligned}
TSS &= \sum Y_i^2-(\sum Y_i)^2/m \\
&= 12^2+12^2+\ldots..+102^2-(40)^2/5 \\
&= 402-320 \\
&= 82 \\
SSR &= b_1\sum(X_1-\overline{X}_1)(Y_i-\overline{Y})+b_2\sum(X_2-\overline{X}_2)(Y_i-\overline{Y})^2 \\
&= (2)(19)+(-3)(-12) \\
&= 74 \\
SSE &= TSS-SSR \\
&= 82-74 \\
&= 8
\end{aligned}
$$

แล้วนำค่าเหล่านี้ลงตารางวิเคราะห์วาเรียนซ์ดังตาราง 20.4.2 ต่อไป

ตาราง 20.4.2 ผลการวิเคราะห์วาเรียนซ์

Sources	df	SS	MS	F
Regression	2	74	37	9.25^{ns}
Deviation	2	8	4	
Total	4	82		

ขอให้สังเกตว่าค่า SSR อาจหาได้ทันที ถ้าหากทำการอินเวอร์สแมตริกซ์ตามวิธีที่แสดงไว้ในตาราง 20.3.2 โดยนำค่าในบรรทัด 21, 22 และ 31, 32 มาคูณแล้วบวกกับ ดังแสดงไว้แล้ว จากผล

การวิเคราะห์วาเรียนซ์ในตาราง 20.4.2 คำนวณได้ว่า F = 37/4 = 9.25 เมื่อเปิดตาราง F ปรากฏว่าไม่แตกต่างทางสถิติ คือ ยอมรับ H_0

การทดสอบสมมุติฐานโดยวิธีวิเคราะห์วาเรียนซ์ ดังแสดงไว้ในตาราง 20.4.2 นั้น อาจให้คำตอบแก่เราเพียงว่า ค่า Y ของข้อมูลนั้นมีรีเกรซชันต่อ X หรือไม่เท่านั้น อย่างไรก็ดีเราอาจจะทำการทดสอบโดยวิธี t-test ต่อไปอีกว่า ค่า Y ขึ้นอยู่กับตัวแปรชุดใดบ้าง ดังนั้นเมื่อมีการปฏิเสธสมมุติฐาน $H_0 : \beta_1 = \beta_2 = \ldots = \beta_n = 0$ ก็จำเป็นที่จะต้องทดสอบสมมุติฐาน $H_0 :\ \beta_i = 0$ ต่อไป

ตัวอย่างที่กล่าวมาแล้วข้างต้น แม้ความปรวนแปรจากรีเกรซชันมีค่าสูงพอสมควร (MS = 37) แต่ไม่แตกต่างทางสถิติเพราะ df ของความคลาดเคลื่อน (deviation) น้อยเกินไป จึงไม่อาจแสดงวิธีการทดสอบค่า β ได้ ตัวอย่างในตาราง 20.4.3 มีค่าสังเกตมากขึ้น จากการอินเวอร์แมตริกโดยวิธีที่แสดงมาแล้วดังตาราง 20.4.4 พบว่า

$$
\begin{aligned}
b_2 &= 0.0870 \\
b_1 &= 1.8434 - (b_2)(0.6190) \\
&= 1.8434 - (0.0870)(0.6190) \\
&= 1.7896 \\
b_0 &= \overline{Y} - b_1\overline{X}_1 - b_2\overline{X}_2 \\
&= \frac{\sum Y_i}{m} - b_1\frac{\sum X_1}{m} - b_2\frac{\sum X_2}{m} \\
&= 81.2778 - (1.7896)(11.9444) - (0.0870)(42.1111) \\
&= 56.2405
\end{aligned}
$$

ซึ่งแสดงสมการรีเกรซชันได้ต่อไปดังนี้

$$\hat{Y} = 56.2401 + 1.7896X_1 + 0.0870X_2$$

ซึ่งหมายถึงว่า สำหรับทุก ๆ หน่วย ppm ของฟอสฟอรัสอนินทรีย์ที่เพิ่มขึ้นนั้น จะทำให้ฟอสฟอรัสในพืชเพิ่มขึ้น 1.789608 ppm ของฟอสฟอรัสอินทรีย์ทำให้ฟอสฟอรัสในพืชเพิ่มขึ้น 0.0870 ppm.

ถ้าจะทำนายค่า $\hat{Y}$ จากสมการได้ผลดังนี้

$$
\begin{aligned}
\hat{Y}_1 &= 56.2401 + 1.7896(0.4) + (0.0870)(53) = 61.6 \\
&\vdots \\
\hat{Y}_{18} &= 56.2401 + 1.7896(29.8) + (0.0870)(51) = 114.2
\end{aligned}
$$

ซึ่งค่าเบี่ยงเบนจากค่าสังเกตจากแต่ค่าหาได้จากสมการ

$$Y_1 - \hat{Y}_1 = 64 - 61.57 = 2.43$$

$$\vdots$$

$$Y_{18} - \hat{Y}_{18} = 99 - 114.2 = -15.2$$

การวิเคราะห์รีเกรซชันมีความจำเป็นสำหรับการที่จะแยกชนิดของความปรวนแปร และทดสอบว่าเกิดจากอิทธิพลของอะไร คือเกิดจากเหตุที่ Y ขึ้นอยู่กับ X หรือจากความคลาดเคลื่อนที่ไม่ทราบสาเหตุ การทดสอบว่า Y ขึ้นอยู่กับ X หรือไม่คือทดสอบสมมุติฐาน $\beta_1 = \beta_2 = \ldots.. = \beta_n = 0$

จากตาราง 20.4.3 อาจวิเคราะห์รีเกรซชันได้ดังนี้

$$\begin{aligned}
TSS &= \sum Y_j^2 - \frac{(\sum Y_j)^2}{m} \\
&= 64^2 + 60^2 + \ldots + 99^2 - \frac{(1{,}463)^2}{18} \\
&= 12{,}389.6111 \\
SSR &= b_1\left[\sum X_1 Y - \frac{(\sum X_1)(\sum Y)}{m}\right] + b_2\left[\sum X_2 Y - \frac{(\sum X_2)(\sum Y)}{m}\right] \\
&= (1.7896)\left[20{,}706.20 - \frac{(215)(1{,}463)}{18}\right] \\
&\quad + (0.0870)\left[63{,}825 - \frac{(758)(1{,}463)}{18}\right] \\
&= 5{,}975.8833
\end{aligned}$$

ดังนั้นหาได้ว่า

$$\begin{aligned}
SSE &= TSS - SSR \\
&= 12{,}389.6111 - 5{,}975.8833 \\
&= 6{,}413.7278
\end{aligned}$$

สำหรับค่า SSE นี้ อาจหาได้อีกวิธีหนึ่ง คือนำค่าเบี่ยงเบน $(Y_i - \hat{Y})$ มายกกำลังสองแล้วบวกกัน คือ

$$\sum (Y_i - \hat{Y})^2 = SSE$$

แต่หาได้ยากกว่า และมักคลาดเคลื่อนมากน้อยตามจำนวนค่าหลังจุดทศนิยมที่ตัดออกไป

ในขั้นต่อไปก็นำค่าวิเคราะห์ลงตารางวิเคราะห์วาเรียนซ์ ดังแสดงในตาราง 20.4.5 ซึ่งเมื่อคำนวณแล้วได้ F = 6.99 เมื่อเปรียบเทียบกับตาราง F ที่ df 15 พบว่าแตกต่างในทางสถิติที่ระดับ 0.01

ตาราง 20.4.3 แสดงฟอสฟอรัสที่พืชสามารถนำมาใช้ประโยชน์ (Y) ฟอสฟอรัสอนินทรีย์ (X) และอินทรีย์ (X_2) ในดินรัฐไอโอวา 18 ตัวอย่าง (ตัวเลขเป็น ppm)

ตัวอย่างดิน	X_1	X_2	Y	$\hat{Y}$	$Y-\hat{Y}$
1	0.4	53	64	61.6	2.4
2	0.4	23	60	59.0	1.0
3	3.1	19	71	63.4	7.6
4	0.6	34	61	60.3	0.7
5	4.7	24	54	66.7	− 12.7
6	1.7	65	77	64.9	12.1
7	9.4	44	81	76.9	4.1
8	10.1	31	93	77.0	16.0
9	11.6	29	93	79.6	13.4
10	12.6	58	51	83.8	− 32.8
11	10.9	37	76	79.0	− 3.0
12	23.1	46	96	101.6	− 5.6
13	23.1	50	77	101.9	− 24.9
14	21.6	44	93	98.7	− 5.7
15	23.1	56	95	102.4	− 7.4
16	1.9	36	54	62.8	− 8.8
17	26.8	58	168	109.2	58.8
18	29.9	51	99	114.2	− 15.2
รวม	215.0	758	1,463	1,463.0	0.0
ค่าเฉลี่ย	11.94	42.11	81.28		

ตาราง 20.4.4 การอินเวอร์สแมตริกข้อมูลในตาราง 20.4.3

	n	X_1	X_2	Y
$\sum$	18	215	758	1,463
$\sum X_1$		4,321.02	10,139	20,706.20
$\sum X_2$			35,076.00	63,825.00
	18	215	758	1,463
$\sum / n$	1	11.9444	42.1111	81.2778
		1,752.9655	1,085.111	3,231.4778
		1	0.6190	1.8434
			2,484.0778	216.1185
				0.0870

ตาราง 20.4.5 ผลการวิเคราะห์วาเรียนซ์ของข้อมูลในตาราง 20.4.3

Sources	df	SS	MS	F
Regression	2	5,975.8833	2,987.9417	6.99**
Deviations	15	6,413.7278	427.5819	
Total	17	12,389.6111		

ในการทดสอบรีเกรซชันโดยใช้ F-test (ตาราง 20.4.5) พบว่า ผลของ Y ขึ้นอยู่กับ X คือสรุปว่า $\beta_1 = \beta_2 \neq 0$ อย่างไรก็ดี ในขั้นนี้ยังสรุปไม่ได้ว่า β ค่าใดที่ไม่เท่ากับศูนย์ หรือทั้ง 2 ค่า ไม่เท่ากับศูนย์ก็ได้ ดังนั้นเมื่อปฏิเสธ H_0 ต่อไปก็ต้องทดสอบสมมุติฐาน $\beta_1 = 0$ และ $\beta_2 = 0$ ต่อไป ซึ่งกระทำโดยการหารีเกรซชันเฉพาะแต่ละค่า X เหมือนเป็นรีเกรซชันเส้นตรงดังนี้

1. เมื่อ Y ขึ้นอยู่กับ X_1 เท่านั้น

$$b = \frac{\sum X_1 Y - \frac{(\sum X_1)(\sum Y)}{n}}{\sum X_1^2 - \frac{(\sum X_1)^2}{n}} = \frac{20,706.20 - \frac{(215.0)(1,463)}{18}}{4,321.02 - \frac{(215)^2}{18}} = 1.8434$$

ซึ่งเท่ากับผลจากการอินเวอร์สแมตริกในตาราง 20.4.4

เมื่อคำนวณ SSR เนื่องจาก X_1

$$SSR(X_1) = \frac{b_1\left[\sum X_1Y - \frac{(\sum X_1)(\sum Y)}{m}\right]^2}{\left[\sum X_1^2 - \frac{(\sum X_1)^2}{m}\right]} = \frac{\left[20,706.20 - \frac{(215.0)(1,463)}{18}\right]^2}{\left[4,321.02 - \frac{(215)^2}{18}\right]} = 5957.0225$$

เมื่อการสังเกตมีทั้ง X_1 และ X_2 หาไว้แล้วว่า SSR = 5,975.8333 (ตาราง 20.4.5) ดังนั้นความแตกต่างคือ SSR (X_1 และ X_2) – SSR (X_1) = 5,975.8333 – 5,957.0225 = 18.8108 ซึ่งมี df = 1 ซึ่งเป็นผลของรีเกรซชันเนื่องจาก X_2 หลังจากลด SSR เนื่องจาก X_1 ซึ่งใช้ทดสอบ $\beta_2 = 0$ ซึ่งหาได้ว่า F = 18.8108/427.5819 = 0.0440 ซึ่งยอมรับ H_0 คือสรุปว่า Y ไม่ขึ้นต่อ X_2

2. เมื่อ Y ขึ้นอยู่กับ X_2

ดำเนินการคำนวณในทำนองเดียวกันพบว่า b_2 = 0.7023, SSR (X_2) = 1,556.70 ดังนั้นความแตกต่าง = 5,975.8333 – 1,556.70 = 4,419.1333 เป็นผลเนื่องจาก X_1 หลังจากลด SSR เนื่องจาก X_2 แล้ว ซึ่งพบว่า F = 4,419.1333/427.5819 = 10.34** ซึ่งแตกต่างทางสถิติ และปฏิเสธ $H_0 : \beta_1 = 0$ จึงสรุปว่า Y ขึ้นอยู่ X_1

ตาราง 20.4.6 การทดสอบ X แต่ละชุด เมื่อแยกผลของ X ชุดอื่นออกไปแล้ว

Source	df	SS	MS	F
X_1 และ X_2	2	5,975.8833		
X_1 เท่านั้น	1	5,957.0225		
X_2 หลังจาก X_1	1	18.8108	18.8108	0.04
X_2 เท่านั้น	1	1,556.70		
X_1 หลังจาก X_2	1	4,419.1333	4,419.1333	10.34**
Deviations	15	6,414.7278	427.5819	

นอกจากวิธีที่แสดงมาแล้วข้างบน การใช้วิธี abbreviated Doolittle ดังแสดงในตอน 18.4 น่าจะเป็นวิธีที่ดีวิธีหนึ่งที่นำมาวิเคราะห์มัลทิเปิลรีเกรซชัน โดยวิธีนี้ เราใช้ backward solution เพื่อคำนวณหาค่า (X^TX) หรือ C_{ij} ซึ่งอาจใช้ในการคำนวณหา b_i และเพื่อทดสอบสมมุติฐาน $\beta_1 = 0$

ต่อไปนี้ เป็นการแสดงวิธีการกลับแมตริกซ์ไปถึงขั้น backward solution โดยใช้ข้อมูลชุดเดิม คือข้อมูลในตาราง 19.3.1 นำผลบวกและผลคูณของข้อมูลดังกล่าวลงตาราง ดังแสดงในตาราง 20.4.7 ต่อจากนั้นก็ทำการอินเวอร์สแมตริกซ์ เหมือนวิธีที่อธิบายมาแล้วในบทที่ 18 ตอน 18.4.1 และดำเนินการในขั้น backward solution ต่อไป

ตาราง 20.4.7 การอินเวอร์สแมตริกซ์โดยวิธี abbreviated Doolittle

บรรทัด	แมตริกซ์ A			แมตริกซ์ I		
01	n	$\sum X_1$	$\sum X_2$	1	0	0
02		$\sum X_1^2$	$\sum X_1X_2$	0	1	0
03			$\sum X_2^2$	0	0	1
01	5	20	25	1	0	0
02		94	103	0	1	0
03			131	0	0	1
11	5	20	25	1	0	0
12	1	4	5	1/5	0	0
21		14	3	-4	1	0
22		1	3/14	$-4/14$	1/14	0
			75/14	$-58/14$	$-3/14$	1
			1	$-58/75$	$-3/75$	14/75

Backward solution

$$(X^TX)^{-1} = C_{ij} = \begin{bmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix}$$

โดยที่ i, j มีค่าเท่ากับ 0, 1 และ 2

$$
\begin{aligned}
C_{00} &= (-58/14)(-58/75) + (-4)(-4/14) + (1)(1/15) \\
&= 2387/525 \\
C_{01} &= (-58/14)(-3/75) + (-4)(1/14) \\
&= -63/525 \\
C_{02} &= (-58/14)(14/75) \\
&= -406/525 \\
C_{11} &= (-3/14)(-3/75) + (1)(1/14) \\
&= 42/525 \\
C_{12} &= (-3/14)(14/75) \\
&= -21/525 \\
C_{22} &= (1)(14/75) \\
&= 14/75
\end{aligned}
$$

ดังนั้น $C_{10} = -63/525$, $C_{20} = -406/525$ และ $C_{21} = -21/525$

ดังนั้นเมื่อแทนค่า C_{ij} ก็ได้

$$
\begin{bmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} 2387/525 & -63/525 & -406/525 \\ -63/525 & 42/525 & -21/525 \\ -406/525 & -21/525 & 14/75 \end{bmatrix}
$$

เมื่อ $\sum Y = 40, \sum X_1 Y = 179$ และ $\sum X_2 Y = 188$ จากสมการ $b = (X^T X)^{-1}(X^T X)$ (สมการ 20-14) ก็อาจหา b_0, b_1 และ b_3 ได้ดังนี้

$$
\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 2387/525 & -63/525 & -406/525 \\ -63/525 & 42/525 & -21/525 \\ -406/525 & -21/525 & 14/75 \end{bmatrix} \begin{bmatrix} 40 \\ 179 \\ 188 \end{bmatrix} = \begin{bmatrix} 15 \\ 2 \\ -3 \end{bmatrix}
$$

ซึ่งได้สมการมัลทิเปิลรีเกรซชัน

$$Y_c = 15 + 2X_1 - 3X_2$$

เหมือนกับที่เคยแสดงไว้แล้ว

จากวิธีการอินเวอร์สแมตริกซ์โดย backward solution เราสามารถหาวาเรียนซ์ของสัมประสิทธิ์ของรีเกรซชันย่อย ๆ ได้ทันทีโดยใช้สมการ

$$s_{b_1}^2 = C_{11} s_{y..x}^2 \qquad \dots(20\text{-}15)$$

$$s_{b_2}^2 = C_{22} s_{y..x}^2 \qquad \dots(20\text{-}16)$$

และวาเรียนซ์ของความแตกต่างระหว่างสัมประสิทธิ์ของรีเกรซชัน ($b_1 - b_2$) เท่ากับ

$$s^2_{(b_1-b_2)} = (C_{ii} + C_{jj} - 2C_{ij})s^2_{y..x} \qquad \ldots(20\text{-}17)$$

วาเรียนซ์ของรีเกรซชันดังกล่าวใช้ทดสอบสมมุติฐาน $\beta_1 = 0, \beta_2 = 0$ และ $\beta_1 = \beta_2 = 0$ ตามลำดับ โดยใช้ t-test ดังนี้

$$\text{ทดสอบ} \qquad \beta_1 = 0, \quad t = \frac{b_1 - \beta_1}{\sqrt{s^2_{b1}}}$$

$$\text{ทดสอบ} \qquad \beta_2 = 0, \quad t = \frac{b_2 - \beta_1}{\sqrt{s^2_{b2}}} \qquad \ldots(20\text{-}18)$$

เมื่อใช้ C_{ij} ที่หาได้และค่า $s^2_{y.x}$ จากตาราง 20.4.2 ซึ่งมีค่าเท่ากับ 4 ก็หาได้ว่า

$$s^2_{b_1} = (42/525)(4) = 0.32$$

$$s^2_{b2} = (14/75)(4) = 0.75$$

$$s^2_{b_1-b_2} = [(42/525 + 14/75) - (2)(-21/525)][4] = 1.49$$

ซึ่งเราสามารถทดสอบสมมุติฐาน $H_0{:}\beta_i = 0$ โดยใช้สมการ (20-15) และทดสอบ $H_0{:}\beta_1 = \beta_2 = 0$ โดยใช้สมการ

$$t = \frac{b_i - b_j}{\sqrt{(C_{ii} + C_{jj} - 2C_{ij})s^2_{y.x}}} \qquad \ldots(20\text{-}19)$$

ซึ่งสามารถแสดงเป็นตัวอย่างได้ดังนี้

(1) ทดสอบ $H_0{:}\beta_1 = 0$

$$t = \frac{b_1}{\sqrt{s^2_{b_1}}} = \frac{2}{\sqrt{0.32}} = 3.54$$

(2) ทดสอบ $H_0{:}\beta_2 = 0$

$$t = \frac{b_2}{\sqrt{s^2_{b_1}}} = \frac{-3}{\sqrt{0.75}} = -3.47$$

(3) ทดสอบ $H_0{:}\beta_1 = \beta_2$

$$t = \frac{b_1 - b_2}{\sqrt{s^2_{b_1-b_2}}} = \frac{2-(-3)}{\sqrt{1.49}} = 4.09$$

ซึ่งไม่มีค่าใดแตกต่างทางสถิติเนื่องจาก df ของ deviation มีค่าต่ำเกินไปนั่นเอง

ตัวอย่าง

เมื่อ X_1 , X_2 และ X_3 เป็นน้ำหนักของเชื้อเพลิง อุณหภูมิของเชื้อเพลิง และน้ำหนักของตัวจรวดตามลำดับ ในการปล่อยจรวดแต่ละครั้งพบว่า ระยะการตกห่างเป้า (Y) ขึ้นอยู่กับองค์ประกอบเหล่านี้ ซึ่งจากการทดลองยิงจรวด 25 ลูก ปรากฏผลดังตาราง 20.4.8 ทั้งนี้นำค่าบางค่ามาลบจาก X_1 , X_3 และ Y เพื่อให้ตัวเลขมีค่าต่ำลง สะดวกในการคำนวณ โดยผลการวิเคราะห์ไม่เปลี่ยนแปลงแต่อย่างใด

จากข้อมูลดังกล่าวนี้ (ก) จงหาสมการเส้นตรง (ข) ทำการวิเคราะห์วาเรียนซ์เพื่อทดสอบ $H_0:\beta_1=\beta_2=\beta_3=0$ (ค) จงทดสอบ β_1,β_2 และ β_3 โดยวิธี t-test และ (ง) จงทดสอบสมมุติฐาน $H_0:\beta_1=\beta_2$ จากตาราง 16.4.3 อาจคำนวณค่าต่าง ๆ ได้ดังนี้

$$\sum X_1^2 = 274{,}168 \qquad \sum X_2^2 = 113{,}759 \qquad \sum X_3^2 = 417{,}792$$
$$\sum X_1X_2 = 168{,}778 \qquad \sum X_1X_3 = 301{,}370 \qquad \sum X_2X_3 = 203{,}474$$
$$\sum X_1Y = 67{,}026 \qquad \sum X_2Y = 20{,}975 \qquad \sum X_2Y_3 = -61{,}226$$
$$\sum Y_2^2 = 350{,}990 \qquad n = 25$$

วิธีทำ

(ก) ปัญหาชนิดนี้ต้องใช้วิธีการอินเวอร์สแมตริกซ์ ซึ่งอธิบายไว้ในตอน 19.4 การตรวจสอบค่า β แต่ละค่าโดยวิธี t-test นั้น จำเป็นที่จะต้องทราบ C_{ij} ดังนั้น จึงต้องกระทำทั้ง forward และ backward solution เนื่องจากได้คุ้นเคยกับขั้นตอนของการคำนวณมาแล้ว จึงจะตัดเฉพาะตอนที่สำคัญมาแสดงไว้ในที่นี้ คือ

$$(X^TX)^{-1}(X^TY) = [b]$$

$$\begin{bmatrix} 3.620692 & -.008255 & -.020106 & .011361 \\ -.008255 & .000061 & -.000017 & .000026 \\ -.020106 & -.000017 & .000275 & .000029 \\ -.011361 & .000026 & .000029 & .000054 \end{bmatrix} \begin{bmatrix} 26 \\ 67{,}206 \\ 20{,}975 \\ -61{,}226 \end{bmatrix} = \begin{bmatrix} -373.594027 \\ 2.323601 \\ 3.330098 \\ -0.647300 \end{bmatrix}$$

ดังนั้นสมการเส้นตรงของรีเกรชชันของ Y ที่มีต่อ X_1 , X_2 , X_3 ก็คือ

$$Y = -\ 373.594027 + 2.323601X_1 + 3.330098X_2 - 0.647300X_3$$

(ข) $$SSR = b_1\left[\sum X_1Y - \frac{(\sum X_1)(\sum Y)}{m}\right] + b_2\left[\sum X_2Y - \frac{(\sum X_2)(\sum Y)}{m}\right] + b_3\left[\sum X_3Y - \frac{(\sum X_3)(\sum Y)}{m}\right]$$

$$= (2.323601)(69{,}634.32) + (3.330098)(22{,}696.20) + (-0.647300)(-57{,}972.88)$$

$$= 274{,}915.95$$

$$TSS = \sum Y^2 - \frac{(\sum Y)^2}{m} = 350{,}990.00 - 27.04 = 350{,}962.96$$

$$SSE = TSS - SSR = 350{,}962.96 - 274{,}915.95 = 76{,}047.01$$

แล้วนำผลการคำนวณลงตารางวิเคราะห์วาเรียนซ์ ดังนี้

Sources	df	SS	MS	F
Regression	3	274,915.95	91,638.65	25.31**
Deviation	21	76,047.01	3,621.29	
Total	24	350,962.96		

(ค) $$s^2_{b_1} = C_{11}s^2_{y..x} = (.000061)(3{,}621.28) = 0.220898$$

$$s^2_{b_2} = C_{22}s^2_{y..x} = (.000275)(3{,}621.28) = 0.995852$$

$$s^2_{b_3} = C_{33}s^2_{y..x} = (.000054)(3{,}621.28) = 0.195549$$

$$t = \frac{b_1}{s_{b_1}} = \frac{2.323601}{\sqrt{(0.220898)}} = \frac{2.323601}{0.469998} = 4.94^{**}$$

$$t = \frac{b_2}{s_{b_2}} = \frac{3.330098}{\sqrt{(0.995852)}} = \frac{3.330098}{0.997923} = 3.34^{**}$$

$$t = \frac{b_3}{s_{b_3}} = \frac{-0.647300}{\sqrt{(0.195549)}} = \frac{-0.647300}{0.442209} = -1.46^{**}$$

(ง) $$s^2_{b_1-b_2} = s^2_{y..x}(C_{11} + C_{22} - 2C_{12})$$

$$= 3{,}621.28\ [0.000061 + 0.000275 - (2)(-0.000017)]$$

$$= 1.344266$$

$$t = \frac{b_1 - b_2}{s_{(b_1 - b_2)}} = \frac{2.323601 - 3.330098}{\sqrt{1.344266}} = \frac{-1.006497}{1.159425} = - \ 0.87$$

ทั้งนี้การทดสอบข้างบนเปิดตาราง t ที่ df = 21

ในตัวอย่างข้างบน ถ้าจะหาช่วงเชื่อมั่นของดรรชนีรีเกรซชันก็คำนวณได้โดยปกติ เช่น ช่วงเชื่อมั่น β_1 ที่ 95 เปอร์เซ็นต์หาได้ดังนี้

$$b \neq t_{0.05} \ s_{b1} = (2.323601) \neq (2.080)(0.469998) = 1.346005 \text{ และ } 3.301197$$

20.5 อิทธิพลของรีเกรซชันเส้นตรงและดรรชนีพหุสหสัมพันธ์

จากตารางวิเคราะห์มัลทิเปิลรีเกรซชันที่แสดงในตอน 20.4 ทำให้เราสามารถแยกขนาดของความปรวนแปรอันเกิดจากรีเกรซชันเส้นตรงที่ได้จากตัวแปรอิสระทุกค่า เราเรียกค่าดังกล่าวว่า coefficient of determination ใช้สัญลักษณ์ " R^2 " ซึ่งหาได้จากสมการ

$$R^2 = \frac{SSR}{TSS} \qquad \ldots(20\text{-}20)$$

โดยที่

SSR = sum of squares เนื่องจากรีเกรซชัน

TSS = sum of squares เนื่องจากค่าสังเกตทั้งหมด

ค่าดังกล่าวนี้จะบอกว่า จากความปรวนแปรทั้งหมดนั้น เกิดจากอิทธิพลของ X ต่าง ๆ ในทางเส้นตรงเท่าไร ที่เหลือเป็นความปรวนแปรจากอิทธิพลอื่น ๆ ที่หาสาเหตุไม่ได้ จากตาราง 20.4.5 หาได้ว่า

$$R^2 = \frac{5,975.8833}{12,389.6111} = 0.4823$$

ซึ่งหมายความว่า ความปรวนแปรของปริมาณฟอสฟอรัสที่พืชใช้ได้เกิดเนื่องจาก ฟอสฟอรัส 48.23 เปอร์เซ็นต์ ซึ่งอาจทดสอบโดยวิธี F-test ดังแสดงไว้ในตารางดังกล่าว

$$F = \frac{SSR/k}{SSE/(n-k-1)}$$

ค่า R^2 จากสมการ (20-20) ต้องปรับค่าเนื่องจาก df โดยใช้สมการดังนี้

$$R_a^2 = 1-(1-R^2)(n-1/n-k) \qquad \ldots(20\text{-}21)$$

$$= 1-(1-0.4828)(17/15)$$

$$= 0.4130$$

ตาราง 20.4.8 ระยะพลาดเป้าของจรวดแต่ละลูก (Y) ที่มีน้ำหนักจรวด (X_1) น้ำหนักเชื้อเพลิง (X_3) และอุณหภูมิเชื้อเพลิง (X_2) ต่าง ๆ กัน

จรวดเลขที่	X1 (− 2,000 กก.)	X2 (C)	X3 (− 3,700 กก.)	Y
1	55	65	177	− 188
2	74	45	139	− 113
3	97	60	141	− 59
4	168	75	60	276
5	126	98	71	232
6	111	82	113	114
7	113	75	137	47
8	45	70	131	− 155
9	79	43	136	− 101
10	81	77	89	4
11	92	45	178	− 147
12	114	74	104	108
13	77	55	144	− 76
14	64	69	100	− 22
15	127	73	115	− 11
16	159	61	112	113
17	85	78	135	− 45
18	96	61	97	69
19	103	74	134	− 61
20	158	63	77	85
21	111	62	141	72
22	104	81	200	116
23	83	51	143	− 67
24	[illegible]	52	[illegible]	[illegible]
25	91	66	141	− 148
รวม	2,508	1,655	3,128	− 26

เมื่อถอดรากสองของ R^2 จะได้ค่าที่เรียกว่า ดรรชนีพหุสหสัมพันธ์ (R) ซึ่งแสดงอัตราความสัมพันธ์ระหว่าง Y กับตัวแปรอื่น ๆ

$$R = \sqrt{\frac{\text{regression SS}}{\text{Total SS}}} = \sqrt{\frac{\text{SSR}}{\text{TSS}}} \qquad \ldots(20\text{-}22)$$

R จะมีค่าอยู่ระหว่าง 0 และ 1 แต่จะผิดแปลกจากดรรชนีสหสัมพันธ์ r ตรงที่ R มีค่าเป็นบวกเสมอ และ R นี้ เราอาจเขียนโดยใช้สัญลักษณ์ปกติว่า $r_{y \cdot 12}$

20.6 แบบฝึกหัด

1. จากข้อมูลที่กำหนดให้

X_1	4	6	4	3	6	3	2	4
X_2	3	4	4	2	4	2	2	3
Y	7	6	2	5	6	7	4	3

ก. จงหาค่า b0, b1 และ b2 โดยใช้วิธีแก้สมการปกติ และวิธีอินเวอร์สแมตริกซ์

ข. คำนวณหา sum of squares เนื่องจากรีเกรซชัน โดยใช้วิธีต่าง ๆ กันสองวิธี

ค. จงทดสอบสมมุติฐาน H0 : $\beta_1 = \beta_2 = 0$ โดยวิธีวิเคราะห์วาเรียนซ์

2. จากการวัดความดันโลหิตของบุคคลคนเดียว จำนวน 13 ครั้ง โดยกระทำติดต่อกันเป็นเวลาหลายปี ในการวัดแต่ละครั้งได้ทำการชั่งน้ำหนัก (เป็นปอนด์) และบันทึกอายุไว้ด้วย ซึ่งปรากฏผลดังนี้

Y	X_1	X_2	Y	X_1	X_2	Y	X_1	X_2
(ความดัน)	(อายุ)	(นน.)	86	24	150	75	33	168
80	19	147	78	24	160	65	34	164
82	20	139	82	26	151	80	38	158
84	22	155	78	28	161	75	40	158
78	23	160	76	31	168			

ก. จงหาสมการรีเกรซชันโดยวิธี abbreviated Doolittle

ข. แยก total SS ของความดันโลหิตออกเป็นส่วนย่อย ๆ แล้วสรุปผลลงในตารางวิเคราะห์วาเรียนซ์

ค. ความดันโลหิตของบุคคลดังกล่าวนี้ขึ้นอยู่กับอายุและน้ำหนักตัวหรือไม่

3. กำหนดให้

X_1	X_2	Y	X_1	X_2	Y
1	8	6	5	0	2
4	2	8	10	−12	−4
9	−8	1	2	4	10
11	−10	0	7	−2	−3
3	6	5	6	−4	5
8	−6	3			

ก. จงคำนวณสัมประสิทธิ์รีเกรซชันย่อย ๆ โดยวิธีแก้สมการปกติ

ข. คำนวณ TSS, SSR, SSE และทดสอบว่า β_1 เหล่านี้ไม่แตกต่างกัน

4. จากข้อมูลในข้อ 3 พบว่า

$$(X^TX)^{-1} = \begin{bmatrix} 4.3705 & -0.8495 & -0.4086 \\ -0.8495 & 0.1690 & 0.0822 \\ -0.4086 & 0.0822 & 0.0422 \end{bmatrix}$$

โดยใช้ผลจากตารางวิเคราะห์วาเรียนซ์ในข้อ 3 จงคำนวณหาวาเรียนซ์ของ b_1 และ b_2 และทดสอบสมมุติฐาน

ก. $H_0 : \beta_1 = 0, \beta_2 = 0$

ข. $H_0 : \beta_1 - \beta_2 = 0$

5. จงคำนวณหา R จากแบบฝึกหัดข้อ 1, 2 และ 3

คำในบท

(1) multiple regression, (2) simultaneous equation, (3) multiple correlation coefficient.

บทที่ 21
การวิเคราะห์โควาเรียนซ์

21.1 คำนำ

การวิเคราะห์โควาเรียนซ์ เป็นวิธีการซึ่งนำวิธีวิเคราะห์วาเรียนซ์และรีเกรซชันมารวมไว้ด้วยกัน มีประโยชน์อย่างยิ่งทั้งในการทดลองเกี่ยวกับพืชและสัตว์ ตัวอย่างเช่น ในการทดลองเปรียบเทียบอาหารสัตว์ ก่อนที่จะทำการทดลอง สัตว์ที่ใช้ทดลองอาจมีขนาดและน้ำหนักต่างกัน เมื่อสิ้นสุดการทดลองเราก็ไม่อาจที่จะเปรียบเทียบผลได้ทันที ทั้งนี้เพราะน้ำหนักก่อนทำการทดลองไม่เท่ากันนั่นเอง จึงจำเป็นที่จะต้องนำเอาน้ำหนักก่อนและภายหลังการทดลองมาวิเคราะห์ร่วมกัน การวิเคราะห์เช่นนี้เรียกว่าการวิเคราะห์โควาเรียนซ์ จากการวิเคราะห์ถ้าหากว่าน้ำหนักก่อนทดลองมีผลกระทบต่อผลการทดลอง ก็จะได้แก้ไขผลการทดลองให้ถูกต้องยิ่งขึ้น อย่างไรก็ดีในการวิเคราะห์โควาเรียนซ์นั้น จำเป็นที่เราจะต้องทราบถึงธรรมชาติของข้อมูลเสียก่อน เช่น ในการทดลองเกี่ยวกับข้าวโพด โดยให้มีจำนวนต้นต่อแปลงเป็น X และให้ผลผลิตเป็น Y และถ้าหากว่าการที่ข้าวโพดมีจำนวนต้นต่อแปลงไม่เท่ากัน เกิดจากผลของทรีตเมนต์หรือเกิดจากความแตกต่างในคุณสมบัติของพันธุ์ การปรับจำนวนต้นให้เท่ากัน ก็เท่ากับเป็นการบิดเบือนความจริงเกี่ยวกับข้อมูล และทำให้ผลของทรีตเมนต์ลดน้อยลงไปด้วย หรือในการทดลองเกี่ยวกับปุ๋ย ปุ๋ยอาจจะเป็นต้นเหตุให้จำนวนต้นของพืชต่อแปลงไม่เท่ากันก็ได้ ดังนั้นควรจะทดลองให้เห็นเสียก่อนว่าไม่มีความแตกต่างระหว่างจำนวนต้นของแปลงต่าง ๆ มิฉะนั้นแล้วไม่ควรที่จะใช้วิธีวิเคราะห์โควาเรียนซ์กับข้อมูลดังกล่าว หรือถ้าหากจะใช้เราก็ควรจะใช้ความระมัดระวังในการให้คำอธิบายเกี่ยวกับผลการทดลอง อย่างไรก็ดีถ้าหากว่าการที่พืชมีจำนวนต้นต่อแปลงไม่เท่ากันนั้น มิได้เนื่องมาจากอิทธิพลของการใช้ทรีตเมนต์ ทั้งนี้ถึงแม้ว่าจะมีการใช้อัตราปลูกต่อแปลงเท่ากันก็จริง แต่จำนวนต้นอาจจะไม่เท่ากันก็ได้ เป็นต้นว่า อิทธิพลของสภาพแวดล้อมทำให้เมล็ดพันธุ์งอกไม่เท่ากัน ในกรณีหลังนี้การใช้วิธีวิเคราะห์โควาเรียนซ์เป็นการเพิ่มความเที่ยงตรงให้แก่การทดลองยิ่งขึ้น

การวิเคราะห์โควาเรียนซ์ นอกจากจะมีประโยชน์ในการควบคุมความคลาดเคลื่อนในการทดลองโดยใช้ความสัมพันธ์ระหว่างตัวแปรชนิด X และ Y แล้ว ยังช่วยในการตรวจสอบรีเกรซชันของคู่ตัวแปรดังกล่าว ช่วยให้การสรุปผลจากข้อมูล โดยเฉพาะอย่างยิ่งเกี่ยวกับทรีตเมนต์ได้ถูกต้องยิ่งขึ้น และอำนวยให้เราสามารถคำนวณค่าสูญหายในการทดลอง โดยไม่จำเป็นต้องปรับค่าลำเอียงที่สูงขึ้น[1]

21.2 การวิเคราะห์โควาเรียนซ์ในแผนการทดลองแบบ CRD

ในการวิเคราะห์โควาเรียนซ์สำหรับแผนการทดลองต่าง ๆ นั้น เรามีข้อสมมุติ หรือข้อกำหนด[2] ต่าง ๆ ที่ต้องคำนึงถึงดังต่อไปนี้

(1) ค่าสังเกต X เป็นค่าคงที่ ส่วนค่าสังเกต Y เป็นตัวแปร ซึ่งมีการกระจายแบบปกติ

(2) ค่าสังเกต Y จากแต่ละทรีตเมนต์ต้องมีวาเรียนซ์เท่ากัน

(3) มีรีเกรซชันของ Y ต่อ X

(4) ค่าดรรชนีรีเกรซชันจากแต่ละทรีตเมนต์ต้องเท่ากัน

ในการวิเคราะห์ข้อมูลจากแผนการทดลองแบบ CRD เราสามารถที่จะคำนวณหาดรรชนีรีเกรซชันของแต่ละทรีตเมนต์ได้สะดวก จึงอาจแยกข้อสมมุติข้างบนแต่ละข้อมาเป็นสมมุติฐานเพื่อทดสอบต่อไป ส่วนในแผนการทดลองแบบอื่น ๆ ก็ให้เป็นเพียงข้อสมมุติอยู่ดังเดิม ดังนั้นในการวิเคราะห์โควาเรียนซ์เรามักมีปัญหาที่จะต้องทดสอบเป็นขั้น ๆ ดังนี้

(1) ทดสอบปัญหา X โดยวิธีวิเคราะห์วาเรียนซ์

(2) ทดสอบปัญหา Y โดยวิธีวิเคราะห์วาเรียนซ์

(3) ทดสอบว่าวาเรียนซ์ของ Y เท่ากัน[3]

(4) ทดสอบว่าค่าดรรชนีรีเกรซชัน ซึ่งได้จากทรีตเมนต์ต่าง ๆ เท่ากัน

(5) ทดสอบสมมุติฐานว่ามีรีเกรซชันของ Y ต่อ X

(6) ทดสอบผลของทรีตเมนต์ภายหลังปรับค่า[4]

สิ่งที่ควรจดจำในขั้นนี้ก็คือ ในการวิเคราะห์โควาเรียนซ์นั้น เราต้องยอมรับสมมุติฐานในปัญหาข้อ 3 และ 4 และปฏิเสธสมมุติฐานเกี่ยวกับปัญหาในข้อ 5 จึงดำเนินการวิเคราะห์ต่อไปได้ เพื่อที่จะให้เข้าใจถึงวิธีการคำนวณแต่ละขั้น จึงจะได้ยกตัวอย่างประกอบคำอธิบายดังนี้

ตัวอย่าง ข้อมูลซึ่งแสดงไว้ในตาราง 21.2.1 เป็นน้ำหนักเมื่อเริ่มต้นทดลอง หรือก่อนการทดลองและน้ำหนักเมื่อสิ้นสุดการทดลองของแกะพันธุ์ Northumbrian 4 หมู่ โดยให้หมู่ A เป็นทรีตเมนต์เปรียบเทียบ หมู่ B ได้รับ phenothiazine หมู่ C ได้ รับแร่ธาตุต่าง ๆ ส่วนหมู่ D ได้รับทั้ง phenothiazine และแร่ธาตุต่าง ๆ จงทำการวิเคราะห์วาเรียนซ์ โควาเรียนซ์ และเปรียบเทียบน้ำหนักเมื่อสิ้นสุดการ ทดลอง ทั้งนี้ภายหลังปรับค่าโดยเรียบร้อยแล้ว

วิธีคำนวณ

1. วิเคราะห์น้ำหนักเมื่อเริ่มทดลอง (X)

$$\text{Correction factor (CF)} = (3{,}491)^2/78 = 156{,}244.628$$

$$\text{Total SS} = 57^2 + 54^2 + \ldots + 31^2 - \text{CF} = 4{,}630.372$$

$$\text{Treatment SS} = (918)^2/20 + (824)^2/19 + \ldots + (885)^2/20 - \text{CF}$$

$$= 77.664$$

$$\text{Error SS} = 4{,}630.372 - 77.664 = 4{,}552.708$$

ผลการวิเคราะห์แสดงในตาราง 21.2.2

ตาราง 21.2.1 นักหนัก (ปอนด์) เมื่อเริ่มทดลองและเมื่อสิ้นสุดการทดลองของแกะ 4 หมู่

A		B		C		D	
X	Y	X	Y	X	Y	X	Y
57	94	46	80	45	70	38	79
54	84	40	68	55	90	59	100
43	81	36	62	49	80	45	72
58	87	40	79	37	58	44	77
44	70	49	79	49	82	47	75
42	67	50	79	52	91	42	95
38	78	44	86	43	79	37	66
54	71	44	70	63	84	49	90
54	80	44	75	56	96	38	83
47	63	42	79	28	59	33	59
51	90	45	68	44	69	65	106
52	86	43	80	49	90	35	61
47	79	42	77	47	79	41	79
56	79	44	74	45	74	43	73
44	74	54	84	40	70	41	74
50	82	43	77	40	62	41	85
34	56	46	70	39	71	47	85
30	49	31	67	35	62	51	91
31	28	41	70	48	84	48	84
32	58					31	66
918	1456	824	1424	864	1450	885	1600

ตาราง 21.2.2 ผลการวิเคราะห์วาเรียนซ์ของน้ำหนักเริ่มทดลอง (X)

Sources	df	SS	MS	F
Treatments	3	77.664	25.888	0.421ns
Error	74	4,552.708	61.523	
Total	77	4,630.372		

2. วิเคราะห์น้ำหนักเมื่อสิ้นสุดการทดลอง (Y)

$$CF = (5{,}930)^2/78 = 450{,}382.050$$

$$\text{Total SS} = 94^2 + 84^2 + ... + 66^2 - CF = 11{,}395.950$$

$$\text{Treatment SS} = (1{,}450)^2/20 + (1{,}424)^2/19 + ... + (1{,}600)^2/20 - CF$$

$$= 547.696$$

$$\text{Error SS} = 11{,}395.950 - 547.696 = 10{,}848.254$$

ผลการวิเคราะห์แสดงในตาราง 21.2.3

3. วิเคราะห์ความสัมพันธ์ระหว่างน้ำหนักเมื่อเริ่มทดลอง และเมื่อสิ้นสุดการทดลอง การที่ไม่มีความแตกต่างในเรื่องน้ำหนักเมื่อสิ้นสุดการทดลองอาจจะมีส่วนเกี่ยวข้องกับน้ำหนักเมื่อเริ่มทดลองก็เป็นได้ ดังนั้นจึงควรที่จะวิเคราะห์ปัญหาเรื่อง X และ Y ร่วมกัน โดยวิธีการซึ่งแสดงไว้ในตาราง 21.2.4 และตอนล่างของตารางดังกล่าว

ตาราง 21.2.3 ผลการวิเคราะห์วาเรียนซ์น้ำหนักเมื่อสิ้นสุดการทดลอง (Y)

Sources	df	SS	MS	F
Treatments	3	547.696	182.565	1.245ns
Error	74	10,848.254	146.598	
Total	77	11,395.950		

ตาราง 21.2.4 วิเคราะห์โควาเรียนซ์แบบทดสอบข้อสมมุติต่าง ๆ

Sources	**Sum of Products (SP)**				**b**	**df**	**SSR**	**Adjusted**	
	df	**xx**	**xy**	**yy**				**SS**	**MS**
Treatments									
A	a – 1	Axx	Axy	Ayy	A	a – 2	A	A	A
B	b – 1	Bxx	Bxy	Byy	B	b – 2	B	B	B
C	c – 1	Cxx	Cxy	Cyy	C	c – 2	C	C	C
D	d – 1	Dxx	Dxy	Dyy	D	d – 2	D	D	D
Within group						N – 2k		-	-
Diff. between slopes						k – 1			
Common	N – k	Exx	Exy	Eyy		N – k – 1	-	-	-
Adjusted Tr. mean						k – 1	-	-	-
Total	N – 1	Txx	Txy	Tyy		N – 2	-	-	-

วิธีคำนวณในตาราง 21.2.4 เมื่อให้ xx, xy, yy แทน $\sum X^2$, $\sum XY$ และ $\sum Y^2$ ตามลำดับ; $\sum n_i = 20 + 19 + 19 + 20 = 78$

Treatment :

(1) $xx = \sum X_{ij}^2 - [(\sum X_{ij})^2 / n_i]$

$Axx = 57^2 + ... + 32^2 - [(918)^2/20] = 1{,}553.800$

ในทำนองเดียวกันอาจหาได้ว่า

$Bxx = 446.421$, $Cxx = 1{,}194.737$ และ $Dxx = 1{,}357.750$

(2) $xy = \sum X_{ij}Y_{ij} - [(\sum X_{ij})(\sum Y_{ij}) / n_i]$

$Axy = (57)(94) + ... + (32)(58) - [(918)(1{,}456)/20]$

$= 2{,}218.600$

ในทำนองเดียวกันอาจหาได้ว่า

$Bxy = 342.369$, $Cxy = 1{,}436.158$ และ $Dxy = 1{,}784.000$

(3) $yy = \sum Y_{ij}^2 - [(\sum Y_{ij})^2 / n_i]$

$Ayy = 94^2 + ... + 58^2 - [(1{,}456)^2/20] = 4{,}771.200$

ในทำนองเดียวกันอาจหาได้ว่า

$Byy = 732.948$, $Cyy = 2{,}368.106$ และ $Dyy = 2{,}976.000$

(4) $b = SP/SSx = xy/xx$

$b(A) = 2{,}218.600/1{,}553.800 = 1.428$

ในทำนองเดียวกันอาจหาได้ว่า

$b(B) = 0.767$, $b(C) = 1.202$ และ $b(D) = 1.314$

(5) SSR = Sum of squares due to regression

$= (SP)^2/SSx = (xy)^2/xx$

$SSR(A) = (2{,}218.600)^2/1{,}553.800 = 3{,}167.838$

ในทำนองเดียวกันอาจหาได้ว่า

$SSR(B) = 262.565$, $SSR(C) = 1{,}726.362$ และ $SSR(D)$

$= 2{,}344.047$

(6) Adjusted SS = Residual SS

= yy − SSR

SS(A) = 1,603.364, SS(B) = 470.379, SS(C) = 641.654

SS(D) = 631.953

(7) MS = adjusted SS/df (n_i − 2)

MS(A) = 89.706, MS(B) = 27.669, MS(C) = 37.744 และ

MS(D) = 35.109

Within group :

(1) Pooled df = (a − 2) + (b − 2) + ... + (d − 2) = 70

(2) Pooled adjusted SS = 1,603.364 + ... + 631.953 = 3,347.350

(3) MS = pooled adjusted SS/pooled df

= 3,347.350/70 = 47.819

Common :

(1) Exx = Axx + Bxx + Cxx + Dxx = 4,552.708

(2) Exy = Axy + Bxy + Cxy + Dxy = 5,781.127

(3) Eyy = Ayy + Byy + Cyy + Dyy = 10,848.254

(4) b = Exy (common)/Exx (common)

= 5,781.127/4,552.708 = 1.270

(5) SSR = $(Exy)^2$/Exx = $(5,781.127)^2$/4,552.708 = 7,340.987

(6) Asjusted SS = Eyy − SSR (common) = 3,507.267

(7) MS = adjusted SS/df (N − k − 1) = 48.045

Difference between slopes :

(1) df = N − k − 1 − N − 2k = k − 1 = 3

(2) SS = SS(common) − SS (within group) = 159.917

(3) MS = SS/df = 159.917/3 = 53.306

Total :

(1) $T'_{xx} = \sum X_{ij}^2 - [(\sum X_{ij})^2 / N]$

$= 57^2 + 54^2 + ... + 31^2 - [(3,491)^2/78] = 4,630.372$

(2) $T'_{xy} = \sum X_{ij} Y_{ij} - [(\sum X_{ij})(\sum Y_{ij}) / N]$

$= (57)(94) + ... + (31)(66) - [(3,491)(5,930)/78]$

$= 5,699.490$

(3) $T'_{yy} = \sum Y_{ij}^2 - [(\sum Y_{ij})^2 / N]$

$= 94^2 + 84^2 + ... + 66^2 - [(5,930)^2/78] = 11,395.950$

(4) $b = T'_{xy} / T'_{xx} = 5,699.490/4,630.372 = 1.231$

(5) $SSR = (T'_{xy})^2 / T'_{xx} = (5,699.490)^2/4,630.372 = 7,015.446$

(6) $SS = T'_{yy} - SSR = 4,380.504$

Adjusted treatment mean :

(1) SS = SS(total) – SS(common)

$= 4,380.504 - 3,507.267 = 873.237$

(2) $df = N - 2 - N + k + 1 = k - 1 = 3$

(3) MS = SS(adjusted tr. mean)/df (k - 1)

$= 873.237/3 = 291.079$

เมื่อคำนวณค่าต่าง ๆ เรียบร้อยแล้วก็นำผลที่ได้ลงตารางดังตาราง 21.2.5 และสามารถทดสอบสมมุติฐานต่าง ๆ ได้ดังนี้

(1) ทดสอบดุลภาคของวาเรียนซ์โดยใช้วิธี Bartlett คือทำการทดสอบว่า MS(A) = 89.076, MS(B) = 27.669, MS(C) = 37.744 และ MS(D) = 35.109 เท่ากัน โดยใช้วิธีการในตอน 7.4 ซึ่งพอสรุปขั้นตอนดังนี้

สมมุติฐาน : $H_0 : \sigma_A^2 = \sigma_B^2 = \sigma_C^2 = \sigma_D^2$

H_1 : วาเรียนซ์ไม่เท่ากันทุกค่า

$$\chi^2 = \frac{M}{C}$$

เมื่อใช้วิธีการคำนวณในตอน 7.4 หาได้ว่า M = 7.6759 และ C = 1.0238

ดังนั้น

$$\chi^2 = 7.6759/1.0238 = 7.497 \text{ (ns)}$$

เมื่อเปรียบเทียบกับค่าไค-สแควร์จากตาราง ($\chi^2_{0.05}$ df3 = 7.81) ก็ยอมรับ H_0 ซึ่งตรงตามต้องการ

(2) ทดสอบว่าดรรชนีรีเกรซชันเท่ากัน

สมมุติฐาน : $H_0 : \beta_A = \beta_B = \beta_C = \beta_D$

H_1 : ดรรชนีรีเกรซชันไม่เท่ากันทุกค่า

สถิติทดสอบ : F – ratio

$$F = \frac{\text{MS (diff.between slopes)}}{\text{MS (within group)}}$$

$$F = 53.306 / 47.819 = 1.115$$

$$F = 1.115 < F_{0.05} \text{ df } 3,70 = 2.74$$

ตาราง 21.2.5 วิเคราะห์โควาเรียนซ์ในแผนการทดลองแบบ CRD

Source	Sum of Products				b	df	SSR	Adjusted	MS
	df	xx	xy	yy				SS	
Treatments									
A	19	1,553.800	2,218.600	4,771.200	1.428	18	3,167.836	1,603.364	89.076
B	18	446.421	342.369	732.948	0.767	17	262.569	470.379	27.669
C	18	1,194.737	1,436.158	2,368.106	1.202	17	1,726.362	641.654	37.744
D	19	1,357.750	1,784.000	2,976.000	1.314	18	2,344.047	631.593	35.109
Within group						70		3,347.350	47.819
Diff. between slopes						3		159.917	53.306
Common	74	4,552.708	5,781.127	10,848.254	1.270	73	7,340.987	3,507.267	48.045
Adjusted Tr. Mean						3		873.237	291.079
Total	77	4,630.372	5,699.490	11,395.950	1.231	76	7,015.446	4,380.504	

เมื่อยอมรับว่า H_0 แสดงว่าครรชนีรีเกรชชันของทุก ๆ ทรีตเมนต์เท่ากัน

(3) ทดสอบว่ามีรีเกรซชันของ Y ต่อ X

สมมุติฐาน : $H_0: \beta = 0$

$H_1: \beta > 0$

สถิติทดสอบ : F – ratio

$$F = \frac{\text{MSR (SSR / 1 – common)}}{\text{MS(common)}}$$

$$= 7{,}340.987 / 48.045 = 152.79^{**}$$

$$F = 152.79 > F_{0.01} \text{ df } 1, 73 = 7.01$$

การปฏิเสธ H_0 ในขั้นนี้จะตรงตามความประสงค์ในการวิเคราะห์โควาเรียนซ์เพราะเป็นการแสดงให้เห็นว่า X และ Y มีความสัมพันธ์ต่อกันจริง และด้วยวิธีปรับค่า Y มีทางที่จะเพิ่มความเที่ยงตรงในการทดลองให้สูงขึ้น

(4) ทดสอบความแตกต่างระหว่าง Y ภายหลังปรับค่า ถ้าให้ μ_A เป็นค่าเฉลี่ยของทรีตเมนต์ภายหลังปรับค่า

สมมุติฐาน $H_0: \mu_{AA} = \mu_{AB} = \mu_{AC} = \mu_{AD}$

H_1 : ค่าเฉลี่ยไม่เท่ากันทุกค่า

สถิติทดสอบ : F – ratio

$$F = \frac{\text{MS(adjusted treatment.mean)}}{\text{MS(common)}}$$

$$= 293.019/48.045 = 6.10^{**}$$

$$F = 6.10 > F_{0.01} \text{ df } 3, 73 = 4.08$$

การปฏิเสธ H_0 ในตอนนี้แสดงว่าการปรับค่าได้ประสบผลสำเร็จ ทั้งนี้เพราะเดิมทีการที่เรานำน้ำหนักเมื่อสิ้นสุดการทดลองมาวิเคราะห์ โดยมิได้คำนึงถึงอิทธิพลของน้ำหนักเมื่อเริ่มทดลองไม่ปรากฏว่าพบความแตกต่างกันระหว่างทรีตเมนต์แต่อย่างใด (ดูตาราง 21.2.3) เมื่อนำค่ามีนสแควร์ในช่อง error ของ Y จากตาราง 21.2.3 และตาราง 21.2.5 มาเปรียบเทียบกัน จะเห็นได้ว่าค่าดังกล่าวลดจาก 146.598 มาเป็น 48.045 ทั้งนี้เพราะเราได้ลดอิทธิพลของ X ลงไป

21.3 การปรับค่าเฉลี่ยของทรีตเมนต์

จากผลการทดสอบพบว่า Y มีรีเกรซชันต่อ X เป็นการชี้ให้เห็นว่าน้ำหนักเมื่อเริ่มทดลองมีอิทธิพลต่อผลของการใช้ทรีตเมนต์ ดังนั้นจึงควรที่จะแก้หรือปรับค่าผลเฉลี่ยของ Y ว่าจะมีค่าเป็นเท่าใด ถ้าหาก X มีค่าเท่ากันทั้งหมดในการปรับค่า Y นั้นใช้สมการ

$$\overline{Y}_{Ai} = \overline{Y}_i - b(\overline{X}_{i.} - \overline{X}_{..}) \qquad ...(21\text{-}1)$$

เมื่อ $\overline{Y}_{Ai}$ = ค่าเฉลี่ยปรับค่าแล้ว[5], $\overline{Y}_i$ = ค่าเฉลี่ย Y ของแต่ละทรีตเมนต์, b = ครรชนีเกรซชันของ common หรือ error, $\overline{X}_{i.}$ = ค่าเฉลี่ย X ของแต่ละทรีตเมนต์ และ $\overline{X}_{..}$ = ค่าเฉลี่ย X ของทุกทรีตเมนต์ จากตาราง 21.2.1 หาได้ว่า $\overline{X}_{..}$ = 44.76 เมื่อให้สมการ (21-1) เราก็จะได้ค่าเฉลี่ยของทรีตเมนต์ที่ปรับค่าแล้วดังต่อไปนี้

$$\begin{aligned}
\overline{Y}_{AA} &= 72.80 - 1.27(45.90 - 44.76) = 71.35 \\
\overline{Y}_{AB} &= 74.94 - 1.27(43.36 - 44.76) = 76.72 \\
\overline{Y}_{AC} &= 76.32 - 1.27(45.47 - 44.76) = 75.42 \\
\overline{Y}_{AD} &= 80.00 - 1.27(44.25 - 44.76) = 80.64
\end{aligned}$$

21.4 การเปรียบเทียบระหว่างค่าเฉลี่ยของทรีตเมนต์ที่ปรับค่าแล้ว

ความคลาดเคลื่อนมาตรฐานสำหรับการเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยดังกล่าว ดัดแปลงได้จากสมการ (5-23) นั่นเอง โดยที่

$$\begin{aligned}
\overline{Y}_{A1} &= \overline{Y}_{1.} - (\overline{X}_{1.} - \overline{X})_{..} \\
\overline{Y}_{A2} &= \overline{Y}_{2.} - (\overline{X}_{2.} - \overline{X}_{...})
\end{aligned}$$

ดังนั้น

$$\overline{Y}_{A1} - \overline{Y}_{A2} = (\overline{Y}_{1.} - \overline{Y}_{2.}) - b(\overline{X}_{1.} - \overline{X}_{2.})$$

ดังนั้น

$$s_{\bar{d}}^2 = \hat{\sigma}^2\left[\frac{2}{n} + \frac{(X_{1.} - X_{2.})^2}{Exx}\right]$$

ถ้าหากว่า $n_1 \neq n_2$ ก็หาได้ว่า

$$s_{\bar{d}}^2 = \hat{\sigma}^2\left[\frac{1}{n_1} + \frac{1}{n_2} + \frac{(\overline{X}_{1.} - \overline{X}_{2.})^2}{Exx}\right] \qquad ...(21\text{-}2)$$

เมื่อ $\hat{\sigma}^2 = s^2_{y..x} = 48.045$, Exx = sum of squares ของ X ในช่อง common = 4,552.708 ดังนั้น ความคลาดเคลื่อนมาตรฐานที่จะใช้ในการเปรียบเทียบระหว่างทรีตเมนต์ A กับ B ก็คือ

$$s_{\bar{d}} = \sqrt{48.045\left[\frac{1}{20} + \frac{1}{19} + \frac{(45.90 - 43.36)^2}{4{,}552.708}\right]}$$

$$= 2.24$$

สถิติที่ใช้ทดสอบคือ

$$t = \frac{(\bar{Y}_{A1} - \bar{Y}_{A2})}{s} \qquad \ldots(21\text{-}3)$$

และมี $\qquad df = N - k - 1$

เมื่อ $\qquad \bar{Y}_{AA} = 71.35, \bar{Y}_{AB} = 76.72$

ดังนั้น

$$t = \frac{71.35 - 76.72}{2.24}$$

$$= -2.40^{*} \text{ (df 73)}$$

ซึ่งพบว่าค่าเฉลี่ยภายหลังปรับค่ามีความแตกต่างกันในทางสถิติ

ค่า $s^2_{\bar{d}}$ จากสมการ (21-2) จะมีขนาดแตกต่างกันไปแล้วแต่ว่าจะนำคู่ของทรีตเมนต์ใดมาทำการเปรียบเทียบในทางปฏิบัติถ้าหากว่า df ในช่อง error จากการวิเคราะห์วาเรียนซ์มากกว่า 20 และไม่มีความแตกต่างในปัญหา X ก็แนะนำให้ใช้ $s^2_{\bar{d}}$ โดยเฉลี่ยจากทุก ๆ คู่ของทรีตเมนต์ ซึ่งหาได้จากสมการ

$$s^2_{\bar{d}} = \frac{2}{n}s^2_{y.x}\left[1 + \frac{Txx}{(k-1)Exx}\right] \qquad \ldots(21\text{-}4)$$

เมื่อ $n_1 = n_2 = n$, Txx = treatment SS สำหรับปัญหา X

21.5 ประสิทธิภาพในการใช้โควาเรียนซ์เมื่อเปรียบเทียบกับการวิเคราะห์วาเรียนซ์

บางครั้งเราอาจจะสงสัยว่า การใช้วิธีการวิเคราะห์โควาเรียนซ์จะมีส่วนเพิ่มความเที่ยงตรงหรือประสิทธิภาพขึ้นเพียงใด ทั้งนี้เมื่อเปรียบเทียบกับวิธีวิเคราะห์วาเรียนซ์ธรรมดา กรณีเช่นนี้อาจจะดูจากประสิทธิภาพของวิธีวิเคราะห์โควาเรียนซ์จากสมการ

$$\text{Effeciency of ANOCOVA} = \frac{s^2_y}{s'^2} \qquad \ldots(21\text{-}5)$$

$$s_y^2 = 146.598$$

$$s'^2 = s_{y.x}^2 \left[1 + \frac{Txx}{(k-1)\,Exx}\right]$$

$$= 48.045 \left[1 + \frac{77.664}{(3)(4{,}552.708)}\right]$$

$$= 48.333$$

$$\text{Effeciency of ANOCOVA} = \frac{146.598}{48.333} \times 100 = 303.3\,\%$$

เมื่อใช้โควาเรียนซ์ในการทดลองที่มี 100 ซ้ำจะให้ผลเท่ากับการทดลองซึ่งไม่ใช้โควาเรียนซ์ 303.3 ซ้ำ

21.6 การวิเคราะห์โควาเรียนซ์ใน CRD โดยไม่มีการทดสอบข้อสมมุติ

การวิเคราะห์โควาเรียนซ์ที่กล่าวมาแล้วข้างต้น นับว่าเสียเวลามาก และมีการคำนวณที่สลับซับซ้อนอย่างยิ่ง ทั้งนี้เพราะได้ผนวกเอาการทดสอบสมมุติฐานทุกชนิดไว้ด้วยกัน ในทางปฏิบัติเราไม่ต้องกระทำเช่นนั้น แต่อาจจะแยกสมมุติฐานบางอย่างออกมาตั้งเป็นข้อกำหนด คือข้อมูลที่จะวิเคราะห์ต้องมีลักษณะสอดคล้องกับข้อกำหนดอันนั้น ในการวิเคราะห์โควาเรียนซ์มีข้อกำหนดไว้ดังนี้

(1) วาเรียนซ์ของ Y จากทุกทรีตเมนต์เท่ากัน

(2) ครรชนีรีเกรซชันของทุกทรีตเมนต์เท่ากัน[6]

(3) มีรีเกรซชันของ Y ต่อ X

เมื่อตั้งข้อกำหนดเรียบร้อยแล้ว ก็ดำเนินการวิเคราะห์โควาเรียนซ์ต่อไป ตาราง 21.6.1 แสดง df ตลอดถึง SS[7] และ SP[8] ของ X และ Y

ตาราง 21.6.1 การวิเคราะห์โควาเรียนซ์ในแผนการทดลองแบบ CRD

Sources	Sum of Products				Adjusted		MS
	df	xx	xy	yy	df	SS	
Treatments(Tr)	k-1	Txx	Txy	Tyy			
Error (E)	N-k	Exx	Exy	Eyy	N-k-1	$Eyy-[(Exy)^2/Exx]$(A)	$s^2_{y..x}$
Tr + E	N-1	Sxx	Sxy	Syy	N-1-1	$Syy-[(Sxy)^2/Sxx]$(B)	
Tr. adjusted					k-1	B-A	

เมื่อวิเคราะห์ข้อมูลในตาราง 21.2.1 ตามวิธีการซึ่งแสดงไว้ในตาราง 21.6.1 ก็กระทำได้ ดังนี้

Total :

$$Sxx = 57^2 + \ldots + 31^2 - CF(X) = 4{,}630.372$$

$$Sxy = (57)(94) + \ldots + (31)(66) - [(3{,}491)(5{,}930)/78]$$

$$= 5{,}699.490$$

$$Syy = 94^2 + \ldots + 66^2 - CF(Y) = 11{,}395.950$$

$$df = N - 1 = 77$$

Treatment :

$$Txx = \frac{(918)^2}{20} + \ldots + \frac{(885)^2}{20} - CF(X)$$

$$= 77.664$$

$$Txy = \frac{(918)(1{,}456)}{20} + \ldots + \frac{(885)(1{,}600)}{20} - \frac{(3{,}491)(5{,}930)}{78}$$

$$= -81.637$$

$$Tyy = \frac{(1{,}456)^2}{20} + \ldots + \frac{(1{,}600)^2}{20} - CF(Y) = 547.696$$

$$df = k - 1 = 3$$

ทั้งนี้

$$CF(X) = (3{,}941)^2/78 = 156{,}244.628$$

$$CF(Y) = (5{,}930)^2/78 = 450{,}382.050$$

Error :

$$Exx = Sxx - Txx = 4{,}552.708$$

$$Exy = Sxy - Txy = 5{,}781.127$$

$$Eyy = Syy - Tyy = 10{,}848.254$$

$$df = N - k = 74$$

แล้วนำค่าที่คำนวณได้กรอกลงตาราง 21.6.2

ในช่อง SP มีตัวเลขที่เราอาจนำมาใช้เพื่อวิเคราะห์วาเรียนซ์ได้ทันที ไม่ว่าในปัญหา X หรือ Y จึงไม่จำเป็นต้องทำการวิเคราะห์วาเรียนซ์ต่างหาก ดังตาราง 21.2.2 และ 21.2.3 ในการวิเคราะห์วาเรียนซ์ของปัญหา X ใช้สมการ

ตาราง 21.6.2 การวิเคราะห์วาเรียนซ์ข้อมูลในตาราง 21.2.1

Sources	Sum of Products (SP)				Adjusted		MS
	77	xx	xy	yy	df	SS	
Treatments	3	77.664	−81.638	547.696			
Error	74	4,522.708	5,781.127	10,848.254	73	3,507.276	48.045
Tr. + E	77	4,630.372	5,699.490	11,395.950	76	4,380.504	
Tr.adjusted					3	873.327	293.020

$$F = \frac{Txx/(k-1)}{Exx/(N0k)} = \frac{77.664/3}{4,552.708/74} = 0.421(ns)$$

$$df = 3,\ 74$$

ส่วนปัญหา Y ใช้

$$F = \frac{Tyy/(k-1)}{Eyy/(N-k)} = \frac{547.696/3}{10,848.254/74} = 1.245(ns)$$

$$df = 3, 74$$

ค่า adjusted SS ในช่อง error หาได้โดยใช้สมการซึ่งแสดงไว้ในตาราง 21.6.1 คือ

$$A = Eyy - \frac{(Exy)^2}{Exx} = 10,848.254 - \frac{(5,781.127)^2}{4,522.708} = 3,507.276$$

$$df = N - k1 - 1 = 78 - 4 - 1 = 73$$

$$s^2_{y..x} = \frac{3,507.276}{73} = 48.045$$

ในช่อง treatment + Error ค่า adjusted SS คือ

$$B = Syy - \frac{(Sxy)^2}{Sxx} = 11,395.950 - \frac{(5,699.940)^2}{4,630.372} = 4,380.504$$

$$df = N - 1 - 1 = 76$$

ค่า SS ในช่อง Tr. adjusted หาได้โดยใช้วิธีหักลบระหว่างค่า adjusted SS ในช่อง treatment + error และช่อง error นั้นก็คือ

$$\text{Adjusted } Tyy = 4,380.504 - 3,507.276 = 873.228$$

$$df = k - 1 = 3$$

ค่า MS ในช่อง error ใช้เป็นตัวประมาณวาเรียนซ์ของประชากร Y ในการทดสอบสมมุติฐานว่าไม่มีความแตกต่างระหว่างค่าเฉลี่ยของทรีตเมนต์ ภายหลังขจัดความปรวนแปร เนื่องจากอิทธิพลของ X ออกไปแล้วให้ใช้

$$F = \frac{MS(\text{Treatment adjusted})}{MS(\text{error})} = \frac{293.020}{48.045}$$

$$= 6.10^{**}$$

$$df = 3, 73$$

21.7 การวิเคราะห์โควาเรียนซ์ในแผนการทดลองแบบ RCB

การคำนวณเพื่อวิเคราะห์โควาเรียนซ์ของแผนการทดลองแบบสุ่มภายในบลอก (RCB) มีวิธีการเช่นเดียวกับที่กล่าวมาในตอน 21.6 เนื่องจากไม่มีวิธีที่ง่ายและสะดวกพอที่จะทดสอบสมมุติฐานเกี่ยวกับดุลภาคของวาเรียนซ์ ของดรรชนีรีเกรซชัน และการมีรีเกรซชันของ Y ต่อ X เราจึงจัดแยกไว้ให้เป็นข้อสมมติหรือข้อกำหนด ซึ่งมักจะบ่งถึงทุกครั้งที่ทำการวิเคราะห์ข้อมูล ส่วนการวิเคราะห์โควาเรียนซ์ในแผนการทดลองแบบ RCB นั้น ได้แสดงไว้ในตาราง 21.7.1

ตาราง 21.7.1 การวิเคราะห์โควาเรียนซ์ในแผนการทดลองแบบ RCB

Sources	Sum of Products				Adjusted		MS
	df	xx	xy	yy	df	SS	
Blocks	n-1	Bxx	Bxy	Byy			
Treatments (Tr)	k-1	Txx	Txy	Tyy			
Error (E)	(n-1)(k-1)	Exx	Exy	Eyy	(n-1)(k-1)-1	$Eyy-(Exy)^2/Exx$ (A)	$s^2_{y.x}$
Tr. + E	n(k-1)	Sxx	Sxy	Syy	n(k-1)-1	$Syy-(Sxy)^2/Sxx$ (B)	
Tr.adjusted					k-1	B-A	

ดรรชนีรีเกรซชันในช่อง error ก็คือ b = Exy/Exx

ตัวอย่างในตาราง 21.7.2 แสดงผลผลิตของข้าวโพด 6 พันธุ์ ซึ่งในแต่ละแปลงย่อยได้ทำการบันทึกจำนวนต้นเอาไว้ ถ้าหากว่าการที่ข้าวโพดมีจำนวนต้นไม่เท่ากันนี้ สืบเนื่องมาจากความแตกต่างในเรื่องความงอก เนื่องมาจากถูกศัตรูทำลาย ฯลฯ และแปลงที่มีจำนวนต้นมากมีผลผลิตสูงกว่าแปลงที่มีจำนวนต้นน้อย การปรับจำนวนต้นให้เท่ากันนับเป็นการเพิ่มความเที่ยงตรงในผลการทดลอง แต่ถ้าพันธุ์หนึ่งมีความงอกดีกว่า หรือขณะที่ต้นยังเล็กอยู่มีความทนทานกว่าพันธุ์อื่น ๆ การปรับจำนวนต้นก็เท่ากับเป็นการบิดเบือนความจริงเกี่ยวกับพันธุ์เหล่านั้น และเพิ่มอคติขึ้นในการเปรียบเทียบระหว่างพันธุ์ เมื่อตรวจดู F-ratio ของจำนวนต้นปรากฏว่าไม่มีความแตกต่างกันในทางสถิติ จึงเป็นประกันได้ว่า

ความแตกต่างในเรื่องจำนวนต้นเป็นไปอย่างสุ่ม การปรับค่าก็จะเป็นการเพิ่มความเที่ยงตรงในการวิเคราะห์ข้อมูล

การคำนวณเพื่อการวิเคราะห์ข้อมูลในตาราง 21.7.2 กระทำเช่นเดียวกับที่อธิบายมาแล้วในตอน 21.6 นั่นเอง ส่วนที่เพิ่มเติมเข้ามาในแผนการทดลองแบบนี้ก็คือบลอก ทั้งนี้ Bxx และ Byy ก็คำนวณตามวิธีปกติ ส่วน Bxy ก็ได้จากการคูณกันระหว่างผลรวมของ X และ Y ทุก ๆ บลอก หารด้วยจำนวนคู่ X, Y ในแต่ละบลอก แล้วลบด้วย CF X, Y ดังนี้

$$
\begin{aligned}
\text{Bxy} &= \frac{(162)(1{,}166)}{6} + \ldots + \frac{(154)(1{,}225)}{6} - \frac{(632)(4{,}794)}{24} \\
&= 8.50
\end{aligned}
$$

นำค่าที่คำนวณได้กรอกลงตารางวิเคราะห์โควาเรียนซ์ 21.7.3 แล้วหา adjusted SS ในช่อง error ส่วนค่ามีนสแควร์ในช่องดังกล่าว ใช้เป็นตัวประมาณวาเรียนซ์ของประชากร Y ในขั้นต่อไปก็รวม SP ในช่อง variety และ error เข้าด้วยกันเพื่อที่จะคำนวณหา SS ของ variety adjusted พึงสังเกตว่าในแผนการทดลองแบบ CRD นั้น ผลรวมของทรีตเมนต์ และ error เท่ากับค่าต่าง ๆ ในช่อง total นั่นเอง จึงไม่มีการแสดงค่า total ไว้ในตาราง 21.6.3

F-test เพื่อที่จะทดสอบว่าไม่มีความแตกต่างระหว่างจำนวนต้นของแต่ละพันธุ์คือ

$$
F = \frac{45.83/5}{113.83/15} = 1.21(\text{ns}), \text{df} = 5, 15
$$

ตาราง 21.7.2 จำนวนต้น (X) และผลผลิต (Y) ของข้าวโพด 6 พันธุ์ทดลองแบบ RCB

แถว	สดมภ์ 1		2		3		4		รวม	
	X	Y	X	Y	X	Y	X	Y	X	Y
A	28	202	22	165	27	191	19	134	96	692
B	23	145	26	201	28	203	24	180	101	729
C	27	188	24	185	28	185	28	220	106	778
D	24	201	28	231	30	238	30	261	112	931
E	30	202	26	178	26	198	29	226	111	804
F	30	228	25	221	27	208	24	206	106	860
รวม	162	1,166	151	1,181	165	1,222	154	1,225	632	4,794

ตาราง 21.7.3 วิเคราะห์โควาเรียนซ์ของตาราง 21.7.2

Sources	Sum of Products				Adjusted		MS
	df	xx	xy	yy	df	SS	
Total	23	181.33	1,485.00	18,678.50			
Blocks	3	21.67	8.50	436.17			
Varieties(V)	5	45.83	559.25	9,490.00			
Error	15	113.83	917.25	8,752.33	14	1,361.07	97.22
V + error	20	159.66	1,476.50	18,242.33	19	4,587.99	
V.adjusted					5	3,226.92	645.38

ส่วนการที่จะทดสอบว่าค่าเฉลี่ยของผลผลิตแต่ละพันธุ์ไม่แตกต่างกันใช้

$$F = \frac{9,490.00/5}{8,752.33/15} = 3.25^{*}, \text{df} = 5, 15$$

และภายหลังการปรับค่า โดยแยกความผันแปรเนื่องจากอิทธิพลของ X ออกไป เราจะได้

$$F = \frac{645.38}{97.22} = 6.64^{**}, \text{df} = 5, 14$$

จะเห็นได้ว่าการปรับค่าให้ MS error ลดลงจาก 8,752.33/15 = 583.50 เป็น 97.22 ในขณะเดียวกัน F-ratio ก็เพิ่มจาก 3.25 เป็น 6.64 ซึ่งแสดงให้เห็นว่าพันธุ์ต่าง ๆ แตกต่างกันอย่างมีนัยสำคัญยิ่ง และถ้าหากเราต้องการที่จะปรับค่าเฉลี่ยของผลผลิตของแต่ละพันธุ์ให้ใช้สมการ (17-1) แล้วอาจจะเปรียบเทียบระหว่างค่าเฉลี่ยเหล่านั้นโดยใช้วิธี LSD หรือวิธีอื่น ๆ ซึ่งได้อธิบายไว้ในบทที่ 12

21.8 การวิเคราะห์โควาเรียนซ์ในแผนการทดลองแบบลาตินสแควร์

การวิเคราะห์โควาเรียนซ์ในแผนการทดลองแบบลาตินสแควร์ มีความคล้ายคลึงกับวิธีการซึ่งกล่าวมาแล้วในตอน 21.6 และ 21.7 นั่นเอง สมมติว่ามีการทดลองขนาด $p \times p$ วิธีการแยก df และการวาง SS ชนิดต่าง ๆ ได้แสดงไว้ในตาราง 21.8.1

ตาราง 21.8.1 การวิเคราะห์โควาเรียนซ์ในแผนการทดลองแบบลาตินสแควร์

Sources	Sum of Products				Adjusted		MS
	df	xx	xy	yy	df	SS	
Rows	p-1	Rxx	Rxy	Ryy			
Columns	p-1	Cxx	Cxy	Cyy			
Treatments	p-1	Txx	Txy	Tyy			
Error	(p-1)(p-2)	Exx	Exy	Eyy	$(p-1)(p-2)$	$Eyy-(Exy)^2/Exx$ (A)	$s^2_{y.x}$
Tr. + error	$(p-1)^2$	Sxx	Sxy	Syy	P^2-2p	$Syy-(Sxy)^2/Sxx$ (B)	
Tr.adjusted						B - A	

การวิเคราะห์โควาเรียนซ์ของข้อมูลที่ได้จากการทดลองแบบลาตินสแควร์ ก็ดำเนินการคำนวณ SP ของ total แถว สดมภ์ ทรีตเมนต์ และ error ตามวิธีการซึ่งแสดงไว้ในตอน 11.3 โดยที่ต้องคำนวณทั้งปัญหา X และ Y ส่วน Rxy,..., Exy คำนวณได้โดยดัดแปลงวิธีการคำนวณซึ่งแสดงไว้ใน ตาราง 21.6 และ 21.7

จากการทดลองเกี่ยวกับผลผลิตของชาในประเทศศรีลังกา โดยวางแผนการทดลองแบบลาติน-สแควร์ ซึ่งมี 4 แถว และ 4 สดมภ์ อย่างไรก็ดีในการทดลองครั้งนี้ไม่มีการใช้ทรีตเมนต์ ดังนั้นจึงเหลือ df ของ error อยู่ 9 ซึ่งเราอาจจะเรียกช่อง error ว่า treatment + error ก็ได้ ข้อสมมุติในการทดลองครั้งนี้ก็คือว่า ผลผลิตก่อนทดลองมีอิทธิพลต่อผลผลิตในขณะที่ทำการทดลอง ตาราง 21.8.1 แสดงผลผลิตของชาเป็นเปอร์เซ็นต์ของน้ำหนักเฉลี่ยของการเก็บเกี่ยวก่อนและขณะทำการทดลอง และกำหนดให้ X เป็นน้ำหนักก่อนการทดลอง ส่วน Y เป็นน้ำหนักขณะที่ทำการทดลอง

เมื่อคำนวณ SS และ SP พวก xx, xy และ yy ของ total, แถว และสดมภ์เรียบร้อยแล้วก็หักลบ SS หรือ SP ของแถวและสดมภ์ออกจาก total ก็จะเหลือเป็นของ error ข้อสมมุติที่เราจะต้องกล่าวถึงก่อนการวิเคราะห์โควาเรียนซ์ ในแผนการทดลองแบบนี้ก็เช่นเดียวกับที่กล่าวมาแล้วในแผนการทดลองอื่น ๆ หรือเราอาจจำได้ง่ายว่าข้อสมมุติในการวิเคราะห์นั้นได้รวมเอาข้อสมมุติในการวิเคราะห์วาเรียนซ์และ รีเกรสชันไว้ด้วยกันหมด

ตาราง 21.8.1 ผลผลิตของชาซึ่งบันทึกเป็นเปอร์เซ็นต์ของน้ำหนักเฉลี่ย

แถว	สดมภ์								รวม	
	1		2		3		4			
	X	Y	X	Y	X	Y	X	Y	X	Y
1	88	90	102	95	91	85	88	81	369	349
2	94	93	110	106	109	114	118	121	431	434
3	109	114	105	106	115	111	94	93	423	424
4	88	92	102	107	91	92	96	102	377	939
รวม	379	389	419	412	406	402	396	397	1,600	1,600

ตาราง 21.8.2 วิเคราะห์โควาเรียนซ์ของข้อมูลในตาราง 17.8.1

Sources	Sum of Products				Adjusted		MS
	df	xx	xy	yy	df	SS	
Total	15	1526	1612	2040			
Rows	3	745	837	1095.5			
Columns	5	213.5	120.75	69.5			
Tr+Error (Error)	15	567.5	654.25	875	8	120.74	15.092

$$\text{Adjusted SS} = \text{Eyy} - [(\text{Exy})^2/\text{Exx}] = 875 - [(654.25)^2/567.5]$$

$$= 120.74$$

$$\text{MS} = 120.74/8 = 15.092$$

ในการวิเคราะห์โดยไม่คำนึงถึงน้ำหนักก่อนการทดลอง ปรากฏว่าได้ค่า MS = 875/9 = 97.22 จะเห็นได้ว่า วาเรียนซ์ได้ลดลงจาก 97.22 เป็น 15.092 ซึ่งแสดงว่าการใช้วิธีวิเคราะห์โควาเรียนซ์นับเป็นการเพิ่มความเที่ยงตรงให้แก่การทดลองเป็นอย่างมาก

21.9 การประมาณค่าสูญหายโดยวิธีโควาเรียนซ์

การวิเคราะห์โควาเรียนซ์จะอำนวยประโยชน์ให้อีกประการหนึ่ง คือสามารถคำนวณค่าสูญหายที่เกิดขึ้นในการทดลอง และค่าที่คำนวณได้โดยวิธีนี้ไม่มีอคติในทางที่ทำให้ทรีตเมนต์ และ SS ของ error สูงขึ้นกว่าเดิม ดังเช่นค่าสูญหายที่คำนวณโดยใช้สมการ (10-4) ส่วนวิธีการก็สะดวกดังตัวอย่าง

ของข้อมูลซึ่งแสดงไว้ในตาราง 21.9.1 อันได้จากแผนการทดลอง RCB และมีค่าสูญหายไปค่าหนึ่ง

เมื่อใช้สมการ (10-4) เราสามารถคำนวณได้ว่า ค่าสูญหาย

$$X = \frac{(3)(41)+(5)(27)-144}{(4)(2)} = 14.25$$

ตาราง 21.9.1 ค่าสูญหายในแผนการทดลองแบบ RCB

บลอก	ทรีตเมนต์ 1	2	3	4	5	รวม
I	10	13	10	6	12	51
II	12	-	9	8	12	41
III	11	14	8	4	15	52
รวม	33	27	27	18	39	144

การคำนวณค่าสูญหายใช้วิธีโควาเรียนซ์ให้ดำเนินเป็นขั้น ๆ ดังนี้ :

ก. ใส่ X = 0 ในทุกช่องที่มีค่า Y

ข. ใส่ X = 1 ในช่องที่ไม่มีค่า Y

ค. ใส่ Y = 0 ในช่องที่ไม่มีค่า Y

ง. ดำเนินการวิเคราะห์โควาเรียนซ์

จ. คำนวณ b = Exy/Exx แล้วเปลี่ยนเครื่องหมายนำหน้าค่าที่คำนวณได้ ค่า -b ดังกล่าวจะเป็นค่า Y ในช่อง Y = 0

ตาราง 21.9.2 ค่า X และ Y ของข้อมูลที่มีค่าสูญหาย

บลอก	ทรีตเมนต์ 1		2		3		4		5		รวม	
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
I	0	10	0	13	0	10	0	6	0	12	0	51
II	0	12	1	0	0	9	0	8	0	12	1	41
III	0	11	0	14	0	8	0	4	0	15	0	52
รวม	0	33	1	27	0	27	0	18	0	39	1	144

การคำนวณเพื่อหาค่าสูญหายก็ให้ดำเนินตามวิธีวิเคราะห์โควาเรียนซ์ธรรมดา

Total : $Sxx = 0^2 + ... + 1^2 + ... + 0^2 - (1^2/15) = 14/15 = df/kn$

$Sxy = (0)(0) + ... + (0)(15) - [(1)(144)/15] = -48/5$

$Syy = 10^2 + ... + 15^2 - [(144)^2/15] = 221.6$

Block : $Bxx = (0^2 + 1 + 0^2)/5 - (1^2/15) = df/kn$

$Bxy = [(0)(51) + ... + (0)(52)]/5 - [(1)(144)]/15 = -7/5$

$Bxy = (51^2 + ... + 52^2)/ - (144^2)/15 = 15.8$

Treatment : $Txx = (0^2 + 1^2 + ... + 0^2)/3 - (1^2/15) = 4/15 = df/kn$

$Txy = [(0)(33) + ... + (0)(39)]/3 - [(1)(144)]/15 = -3/5$

$Tyy = (33^2 + ... + 39^2)/3 - (144^2)/15 = 81.6$

Error : $Exx = 14/15 - 2/15 - 4/15 = 8/15 = df/kn$

$Exy = -48/5 + 7/5 + 3/5 = -38/5$

$Eyy = 221.6 - 15.8 - 81.6 = 124.2$

นำค่าที่คำนวณได้กรอกลงตาราง 21.9.3 แล้วหาค่าดรรชนีรีเกรซชัน b ซึ่งหาได้ว่า $b = Exy/Exx = (-38/5)/(8/15) = -14.25$

เมื่อกลับค่า b ด้วยเครื่องหมายลบก็จะได้ 14.25 ซึ่งเท่ากับค่าสูญหายที่คำนวณโดยใช้สมการ (10-4) นั่นเอง แต่ถ้าใช้วิธีวิเคราะห์โควาเรียนซ์ถึงแม้จะได้ค่าสูญหายเท่ากันพอดี แต่ยังสามารถคำนวณติดต่อกันไปจนได้ treatment (adjusted) SS เท่ากับ 51.70 ถ้าเราต้องการจะทดสอบสมมุติฐานว่าไม่มีความแตกต่างระหว่างค่าเฉลี่ย ก็สามารถใช้ได้ทันที

ตาราง 21.9.3 การวิเคราะห์โควาเรียนซ์ของตาราง 21.9.2

Sources	Sum of Products				b	Adjusted		MS
	df	xx	xy	yy		df	SS	
Total	1	4/15	− 48/5	221.6				
	2	2/15	− 7/5	15.8				
Treatments	4	4/15	− 3/5	81.6				
Error	8	8/15	− 38/5	124.2	− 14.25	7	70.05	10.01
Tr + Error	12	12/15	− 41/5	205.8		11	121.75	
Tr. adjusted						4	51.70	12.92

21.10 แบบจำลองทางคณิตศาสตร์

แบบจำลองทางคณิตศาสตร์ของค่าสังเกตในโควาเรียนซ์เป็นการผนวกเอาแบบจำลองในการวิเคราะห์วาเรียนซ์และรีเกรซชันไว้ด้วยกัน ตัวอย่างในแผนการทดลองแบบ CRD ค่าสังเกตหนึ่ง ๆ ประกอบด้วย

$$Y_{ij} = \mu + \alpha_i + \beta(X_{ij} - \overline{X}..) + \varepsilon_{ij}$$

เมื่อ Y_{ij} = ค่าสังเกตที่ i ในทรีตเมนต์ที่ j, μ = ค่าเฉลี่ยของประชากร, α_j = ผลของทรีตเมนต์ที่ i, $X_{ij} - \overline{X}..$ เป็นค่าเบี่ยงเบนของ X และ β เป็นสัมประสิทธิ์รีเกรซชันเฉลี่ยสำหรับทรีตเมนต์ทั้งหมด ในทำนองเดียวกันเราอาจจะเขียนแบบจำลองทางคณิตศาสตร์สำหรับแผนการทดลองแบบ RCB และลาตินสแควร์ได้ดังนี้

$$Y_{ij} = \mu + \alpha_i + \gamma_j + \beta(X_{ij} - \overline{X}..) + \varepsilon_{ij} \text{ (RB)}$$

$$Y_{ijk} = \mu + \alpha_i + \gamma_j + \delta_k + \beta(X_{ijk} - \overline{\overline{X}}..) + \varepsilon_{ijk} \text{ (LS)}$$

เมื่อ α_i และ γ_j เป็นผลของทรีตเมนต์ที่ i และบลอกที่ j ตามลำดับ สำหรับแผนการทดลองแบบ RCB และ α_i, γ_j, δ_k เป็นผลของทรีตเมนต์ที่ i แถวที่ j และสดมภ์ที่ k ตามลำดับ

21.11 แบบฝึกหัด

1. ข้อมูลในตารางข้างล่างแสดงคะแนนของนักศึกษาจากมหาวิทยาลัย 3 แห่ง โดยให้ X เป็นคะแนนชั้นมัธยมปีที่ 6 ส่วน Y เป็นคะแนนในการสอบชั้นปีที่ 1 ของมหาวิทยาลัย

มหาวิทยาลัย					
A		B		C	
Y	X	Y	X	Y	X
57	53	65	68	66	71
56	80	80	81	65	61
73	73	68	78	67	76
76	77	65	63	89	61
69	63	77	73	78	81
80	68	68	81	70	51
68	68	69	66	63	74
70	72	70	75		
79	73	77	81		

ก. จงทำการวิเคราะห์วาเรียนซ์ของ X และ Y

ข. วิเคราะห์โควาเรียนซ์ข้อมูลดังกล่าว โดยใช้วิธีซึ่งอำนวยให้สามารถตรวจสอบข้อกำหนดได้ครบทุก ๆ ข้อ

ค. จงทดสอบสมมติฐานที่ว่าไม่มีความแตกต่างระหว่างผลเฉลี่ยภายหลังปรับค่า

2. ผลผลิตของถั่วเหลือง (Y) และเปอร์เซ็นต์ของต้นเป็นโรค canker (X) จากแปลงซึ่งใช้แผนการทดลองแบบ RCB ปรากฏในตารางดังนี้

บลอก	ถั่วเหลือง								รวม	
	A		B		C		D			
	X	Y	X	Y	X	Y	X	Y	X	Y
1	19.3	21.3	10.1	28.3	4.3	26.7	14.0	25.1	47.7	101.4
2	29.2	19.7	34.7	20.7	48.2	14.7	30.2	20.1	142.3	75.2
3	1.0	28.7	14.0	26.0	6.3	29.0	7.2	24.9	28.5	108.6
4	6.4	27.3	5.6	34.1	6.7	29.0	8.9	29.8	27.6	120.2
รวม	55.9	97.0	64.4	109.1	65.5	99.4	60.3	99.9	246.1	405.4

ผลการวิเคราะห์โควาเรียนซ์

Sources	df	xx	xy	yy
Blocks	3	2,239.3	– 748.0	272.9
Treatments	3	14.3	10.2	21.2
Error	9	427.0	– 145.7	66.0
T + E	12	441.3	– 135.5	87.2

ก. จงใช้ F-test เพื่อตรวจสอบความแตกต่างระหว่างผลผลิตและเปอร์เซ็นต์ต้นเป็นโรคของพันธุ์เหล่านี้

ข. กล่าวถึงข้อสมมุติที่จำเป็นก่อนทำการวิเคราะห์โควาเรียนซ์ข้อมูลดังกล่าว

ค. ทำการวิเคราะห์โควาเรียนซ์ตารางข้างบนให้สมบูรณ์พร้อมแสดง F-test ของค่าเฉลี่ยภายหลังปรับค่าแล้ว

ง. ปรับค่าเฉลี่ยผลผลิตของทุก ๆ พันธุ์

จ. คำนวณหาความคลาดเคลื่อนมาตรฐานของความแตกต่างระหว่างค่าเฉลี่ย A, B ภายหลังปรับค่า

ฉ. ประมาณประสิทธิภาพของการใช้วิธีวิเคราะห์แบบโควาเรียนซ์

3. จากข้อมูลในตาราง 10.6.1 จงประมาณค่าสูญหายโดยใช้วิธีโควาเรียนซ์ การใช้วิธีนี้และวิธีตามสมการ 10.6.1 จะมีข้อดีข้อเสียแตกต่างกันอย่างไรบ้าง อธิบาย

คำในบท

(1) biased upward, (2) assumption, (3) homogeneitity of variance, (4) adjusted Y, (5) adjusted mean, (6) homogeneity of regression coefficient, (7) sum of squares, (8) sum of products, (9) biased upward.

ภาคผนวก
ตารางสถิติที่ใช้ในเล่มนี้

ตาราง ผ.1 ตารางเลขสุ่ม

	00–04	05–09	10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49
00	54463	22662	65905	70639	79365	67382	29085	69831	47058	08186
01	15389	85205	18850	39226	42249	90669	96325	23248	60933	26927
02	85941	40756	82414	02015	13858	78030	16269	65978	01385	15345
03	61149	69440	11286	88218	58925	03638	52862	62733	33451	77455
04	05219	81619	10651	67079	92511	59888	84502	72095	83463	75577
05	41417	98326	87719	92294	46614	50948	64886	20002	97365	30976
06	28357	94070	20652	35774	16249	75019	21145	05217	47286	76305
07	17783	00015	10806	83091	91530	36466	39981	62481	49177	75779
08	40950	84820	29881	85966	62800	70326	84740	62660	77379	90279
09	82995	64157	66164	41180	10089	41757	78258	96488	88629	37231
10	96754	17676	55659	44105	47361	34833	86679	23930	53249	27083
11	34357	88040	53364	71726	45690	66334	60332	22554	90600	71113
12	06318	37403	49927	57715	50423	67372	63116	48888	21505	80182
13	62111	52820	07243	79931	89292	84767	85693	73947	22278	11551
14	47534	09243	67879	00544	23410	12740	02540	54440	32949	13491
15	98614	75993	84460	62846	59844	14922	48730	73443	48167	34770
16	24856	03648	44898	09351	98795	18644	39765	71058	90368	44104
17	96887	12479	80621	66223	86085	78285	02432	53342	42846	94771
18	90801	21472	42815	77408	37390	76766	52615	32141	30268	18106
19	55165	77312	83666	36028	28420	70219	81369	41943	47366	41067
20	75884	12952	84318	95108	72305	64620	91318	89872	45375	85436
21	16777	37116	58550	42958	21460	43910	01175	87894	81378	10620
22	46230	43877	80207	88877	89380	32992	91380	03164	98656	59337
23	42902	66892	46134	01432	94710	23474	20423	60137	60609	13119
24	81007	00333	39693	28039	10154	95425	39220	19774	31782	49037
25	68089	01122	51111	72373	06902	74373	96199	97017	41273	21546
26	20411	67081	89950	16944	93054	87687	96693	87236	77054	33848
27	58212	13160	06468	15718	82627	76999	05999	58680	96739	63700
28	70577	42866	24969	61210	76046	67699	42054	12696	93758	03283
29	94522	74358	71659	62038	79643	79169	44741	05437	39038	13163
30	42626	86819	85651	88678	17401	03252	99547	32404	17918	62880
31	16051	33763	57194	16752	54450	19031	58580	47629	54132	60631
32	08244	27647	33851	44705	94211	46716	11738	55784	95374	72655
33	59497	04392	09419	89964	51211	04894	72882	17805	21896	83864
34	97155	13428	40293	09985	58434	01412	69124	82171	59058	82859
35	98409	66162	95763	47420	20792	61527	20441	39435	11859	41567
36	45476	84882	65109	96597	25930	66790	65706	61203	53634	22557
37	89300	69700	50741	30329	11658	23166	05400	66669	48708	03887
38	50051	95137	91631	66315	91428	12275	24816	68091	71710	33258
39	31753	85178	31310	89642	98364	02306	24617	09609	83942	22716
40	79152	53829	77250	20190	56535	18760	69942	77448	33278	48805
41	44560	38750	83635	56540	64900	42912	13953	79149	18710	68618
42	68328	83378	63369	71381	39564	05615	42451	64559	97501	65747
43	46939	38689	58625	08342	30459	85863	20781	09284	26333	91777
44	83544	86141	15707	96256	23068	13782	08467	89469	93842	55349
45	91621	00881	04900	54224	46177	55309	17852	27491	89415	23466
46	91896	67126	04151	03795	59077	11848	12630	98375	52068	60142
47	55751	62515	21108	80830	02263	29303	37204	96926	30506	09808
48	85156	87689	95493	88842	00664	55017	55539	17771	69448	87530
49	07521	56898	12236	60277	39102	62315	12239	07105	11844	01117

ตาราง ผ.1 (ต่อ)

	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85–89	90–94	95–99
00	59391	58030	52098	82718	87024	82848	04190	96574	90464	29065
01	99567	76364	77204	04615	27062	96621	43918	01896	83991	51141
02	10363	97518	51400	25670	98342	61891	27101	37855	06235	33316
03	86859	19558	64432	16706	99612	59798	32803	67708	15297	28612
04	11258	24591	36863	55368	31721	94335	34936	02566	80972	08188
05	95068	88628	35911	14530	33020	80428	39936	31855	34334	64865
06	54463	47237	73800	91017	36239	71824	83671	39892	60518	37092
07	16874	62677	57412	13215	31389	62233	80827	73917	82802	84420
08	92494	63157	76593	91316	03505	72389	96363	52887	01087	66091
09	15669	56689	35682	40844	53256	81872	35213	09840	34471	74441
10	99116	75486	84989	23476	52967	67104	39495	39100	17217	74073
11	15696	10703	65178	90637	63110	17622	53988	71087	84148	11670
12	97720	15369	51269	69620	03388	13699	33423	67453	43269	56720
13	11666	13841	71681	98000	35979	39719	81899	07449	47985	46967
14	71628	73130	78783	75691	41632	09847	61547	18707	85489	69944
15	40501	51089	99943	91843	41995	88931	73631	69361	05375	15417
16	22518	55576	98215	82068	10798	86211	36584	67466	69373	40054
17	75112	30485	62173	02132	14878	92879	22281	16783	86352	00077
18	80327	02671	98191	84342	90813	49268	95441	15496	20168	09271
19	60251	45548	02146	05597	48228	81366	34598	72856	66762	17002
20	57430	82270	10421	00540	43648	75888	66049	21511	47676	33444
21	73528	39559	34434	88596	54086	71693	43132	14414	79949	85193
22	25991	65959	70769	64721	86413	33475	42740	06175	82758	66248
23	78388	16638	09134	59980	63806	48472	39318	35434	24057	74739
24	12477	09965	96657	57994	59439	76330	24596	77515	09577	91871
25	83266	32883	42451	15579	38155	29793	40914	65990	16255	17777
26	76970	80876	10237	39515	79152	74798	39357	09054	73579	92359
27	37074	65198	44785	68624	98336	84481	97610	78735	46703	98265
28	83712	06514	30101	78295	54656	85417	43189	60048	72781	72606
29	20287	56862	69727	94443	64936	08366	27227	05158	50326	59566
30	74261	32592	86538	27041	65172	85532	07571	80609	39285	65340
31	64081	49863	08478	96001	18888	14810	70545	89755	59064	07210
32	05617	75818	47750	67814	29575	10526	66192	44464	27058	40467
33	26793	74951	95466	74307	13330	42664	85515	20632	05497	33625
34	65988	72850	48737	54719	52056	01596	03845	35067	03134	70322
35	27366	42271	44300	73399	21105	03280	73457	43093	05192	48657
36	56760	10909	98147	34736	33863	95256	12731	66598	50771	83665
37	72880	43338	93643	58904	59543	23943	11231	83268	65938	81581
38	77888	38100	03062	58103	47961	83841	25878	23746	55903	44115
39	28440	07819	21580	51459	47971	29882	13990	29226	23608	15873
40	63525	94441	77033	12147	51054	49955	58312	76923	96071	05813
41	47606	93410	16359	89033	89696	47231	64498	31776	05383	39902
42	52669	45030	96279	14709	52372	87832	02735	50803	72744	88208
43	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
44	59348	11695	45751	15865	74739	05572	32688	20271	65128	14551
45	12900	71775	29845	60774	94924	21810	38636	33717	67598	82521
46	75086	23537	49939	33595	13484	97588	28617	17979	70749	35234
47	99495	51434	29181	09993	38190	42553	68922	52125	91077	40197
48	26075	31671	45386	36583	93459	48599	52022	41330	60651	91321
49	13636	93596	23377	51133	95126	61496	42474	45141	46660	42338

ตาราง ผ.1 (ต่อ)

	00–04	05–09	10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49
50	64249	63664	39652	40646	97306	31741	07294	84149	46797	82487
51	26538	44249	04050	48174	65570	44072	40192	51153	11397	58212
52	05845	00512	78630	55328	18116	69296	91705	86224	29503	57071
53	74897	68373	67359	51014	33510	83048	17056	72506	82949	54600
54	20872	54570	35017	88132	25730	22626	86723	91691	13191	77212
55	31432	96156	89177	75541	81355	24480	77243	76690	42507	84362
56	66890	61505	01240	00660	05873	13568	76082	79172	57913	93448
57	41894	57790	79970	33106	86904	48119	52503	24130	72824	21627
58	11303	87118	81471	52936	08555	28420	49416	44448	04269	27029
59	54374	57325	16947	45356	78371	10563	97191	53798	12693	27928
60	64852	34421	61046	90849	13966	39810	42699	21753	76192	10508
61	16309	20384	09491	91588	97720	89846	30376	76970	23063	35894
62	42587	37065	24526	72602	57589	98131	37292	05967	26002	51945
63	40177	98590	97161	41682	84533	67588	62036	49967	01990	72308
64	82309	76128	93965	26743	24141	04838	40254	26065	07938	76236
65	79788	68243	59732	04257	27084	14743	17520	95401	55811	76099
66	40538	79000	89559	25026	42274	23489	34502	75508	06059	86682
67	64016	73598	18609	73150	62463	33102	45205	87440	96767	67042
68	49767	12691	17903	93871	99721	79109	09425	26904	07419	76013
69	76974	55108	29795	08404	82684	00497	51126	79935	57450	55671
70	23854	08480	85983	96025	50117	64610	99425	62291	86943	21541
71	68973	70551	25098	78033	98573	79848	31778	29555	61446	23037
72	36444	93600	65350	14971	25325	00427	52073	64280	18847	24768
73	03003	87800	07391	11594	21196	00781	32550	57158	58887	73041
74	17540	26188	36647	78386	04558	61463	57842	90382	77019	24210
75	38916	55809	47982	41968	69760	79422	80154	91486	19180	15100
76	64288	19843	69122	42502	48508	28820	59933	72998	99942	10515
77	86809	51564	38040	39418	49915	19000	58050	16899	79952	57849
78	99800	99566	14742	05028	30033	94889	53381	23656	75787	59223
79	92345	31890	95712	08279	91794	94068	49337	88674	35355	12267
80	90363	65162	32245	82279	79256	80834	06088	99462	56705	06118
81	64437	32242	48431	04835	39070	59702	31508	60935	22390	52246
82	91714	53662	28373	34333	55791	74758	51144	18827	10704	76803
83	20902	17646	31391	31459	33315	03444	55743	74701	58851	27427
84	12217	86007	70371	52281	14510	76094	96579	54853	78339	20839
85	45177	02863	42307	53571	22532	74921	17735	42201	80540	54721
86	28325	90814	08804	52746	47913	54577	47525	77705	95330	21866
87	29019	28776	56116	54791	64604	08815	46049	71186	34650	14994
88	84979	81353	56219	67062	26146	82567	33122	14124	46240	92973
89	50371	26347	48513	63915	11158	25563	91915	18431	92978	11591
90	53422	06825	69711	67950	64716	18003	49581	45378	99878	61130
91	67453	35651	89316	41620	32048	70225	47597	33137	31443	51445
92	07294	85353	74819	23445	68237	07202	99515	62282	53809	26685
93	79544	00302	45338	16015	66613	88968	14595	63836	77716	79596
94	64144	85442	82060	46471	24162	39500	87351	36637	42833	71875
95	90919	11883	58318	00042	52402	28210	34075	33272	00840	73268
96	06670	57353	86275	92276	77591	46924	60839	55437	03183	13191
97	36634	93976	52062	83678	41256	60948	18685	48992	19462	96062
98	75101	72891	85745	67106	26010	62107	60885	37503	55461	71213
99	05112	71222	72654	51583	05228	62056	57390	42746	39272	96659

ตาราง ผ.1 (ต่อ)

	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85–89	90–94	95–99
50	32847	31282	03345	89593	69214	70381	78285	20054	91018	16742
51	16916	00041	30236	55023	14253	76582	12092	86533	92426	37655
52	66176	34037	21005	27137	03193	48970	64625	22394	39622	79085
53	46299	13335	12180	16861	38043	59292	62675	63631	37020	78195
54	22847	47839	45385	23289	47526	54098	45683	55849	51575	64689
55	41851	54160	92320	69936	34803	92479	33399	71160	64777	83378
56	28444	59497	91586	95917	68553	28639	06455	34174	11130	91994
57	47520	62378	98855	83174	13088	16561	68559	26679	06238	51254
58	34978	63271	13142	82681	05271	08822	06490	44984	49307	61717
59	37404	80416	69035	92980	49486	74378	75610	74976	70056	15478
60	32400	65482	52099	53676	74648	94148	65095	69597	52771	71551
61	89262	86332	51718	70663	11623	29834	79820	73002	84886	03591
62	86866	09127	98021	03871	27789	58444	44832	36505	40672	30180
63	90814	14833	08759	74645	05046	94056	99094	65091	32663	73040
64	19192	82756	20553	58446	55376	88914	75096	26119	83898	43816
65	77585	52593	56612	95766	10019	29531	73064	20953	53523	58136
66	23757	16364	05096	03192	62386	45389	85332	18877	55710	96459
67	45989	96257	23850	26216	23309	21526	07425	50254	19455	29315
68	92970	94243	07316	41467	64837	52406	25225	51553	31220	14032
69	74346	59596	40088	98176	17896	86900	20249	77753	19099	48885
70	87646	41309	27636	45153	29988	94770	07255	70908	05340	99751
71	50099	71038	45146	06146	55211	99429	43169	66259	97786	59180
72	10127	46900	64984	75348	04115	33624	68774	60013	35515	62556
73	67995	81977	18984	64091	02785	27762	42529	97144	80407	64524
74	26304	80217	84934	82657	69291	35397	98714	35104	08187	48109
75	81994	41070	56642	64091	31229	02595	13513	45148	78722	30144
76	59537	34662	79631	89403	65212	09975	06118	86197	58208	16162
77	51228	10937	62396	81460	47331	91403	95007	06047	16846	64809
78	31089	37995	29577	07828	42272	54016	21950	86192	99046	84864
79	38207	97938	93459	75174	79460	55436	57206	87644	21296	43393
80	88666	31142	09474	89712	63153	62333	42212	06140	42594	43671
81	53365	56134	67582	92557	89520	33452	05134	70628	27612	33738
82	89807	74530	38004	90102	11693	90257	05500	79920	62700	43325
83	18682	81038	85662	90915	91631	22223	91588	80774	07716	12548
84	63571	32579	63942	25371	09234	94592	98475	76884	37635	33608
85	68927	56492	67799	95398	77642	54913	91583	08421	81450	76229
86	56401	63186	39389	88798	31356	89235	97036	32341	33292	73757
87	24333	95603	02359	72942	46287	95382	08452	62862	97869	71775
88	17025	84202	95199	62272	06366	16175	97577	99304	41587	03686
89	02804	08253	52133	20224	68034	50865	57868	22343	55111	03607
90	08298	03879	20995	19850	73090	13191	18963	82244	78479	99121
91	59883	01785	82403	96062	03785	03488	12970	64896	38336	30030
92	46982	06682	62864	91837	74021	89094	39952	64158	79614	78235
93	31121	47266	07661	02051	67599	24471	69843	83696	71402	76287
94	97807	56641	63416	17577	30161	87320	37752	73276	48969	41915
95	57364	86746	08415	14621	49430	22311	15836	72492	49372	44103
96	09559	26263	69511	28064	75999	44540	13337	10918	79846	54809
97	53873	55571	00608	42661	91332	63956	74087	59008	47493	99581
98	35531	19162	86406	05299	77511	24311	57257	22826	77555	05941
99	28229	88629	25695	94932	30721	16197	78742	34974	97528	45447

ตาราง ผ.2 อัตราส่วนของพื้นที่ใต้โค้งมาตรฐาน

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000									

ดัดแปลงจาก Snedecor และ Cochran (1967)

ตาราง ผ.3 การกระจายของค่าไค – สแควร์

Degrees of Freedom	Probability of a Greater Value												
	0.995	0.990	0.975	0.950	0.900	0.750	0.500	0.250	0.100	0.050	0.025	0.010	0.005
1	....	....	....	....	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	33.66	39.34	45.62	51.80	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	42.94	49.33	56.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	52.29	59.33	66.98	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	61.70	69.33	77.58	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	71.14	79.33	88.13	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	80.62	89.33	98.64	107.56	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	90.13	99.33	109.14	118.50	124.34	129.56	135.81	140.17

คัดแปลงจาก Snedecor และ Cochran (1967)

ตาราง ผ.4 การกระจายของค่า t

Degrees of Freedom	Probability of a Larger Value, Sign Ignored								
	0.500	0.400	0.200	0.100	0.050	0.025	0.010	0.005	0.001
1	1.000	1.376	3.078	6.314	12.706	25.452	63.657		
2	0.816	1.061	1.886	2.920	4.303	6.205	9.925	14.089	31.598
3	.765	0.978	1.638	2.353	3.182	4.176	5.841	7.453	12.941
4	.741	.941	1.533	2.132	2.776	3.495	4.604	5.598	8.610
5	.727	.920	1.476	2.015	2.571	3.163	4.032	4.773	6.859
6	.718	.906	1.440	1.943	2.447	2.969	3.707	4.317	5.959
7	.711	.896	1.415	1.895	2.365	2.841	3.499	4.029	5.405
8	.706	.889	1.397	1.860	2.306	2.752	3.355	3.832	5.041
9	.703	.883	1.383	1.833	2.262	2.685	3.250	3.690	4.781
10	.700	.879	1.372	1.812	2.228	2.634	3.169	3.581	4.587
11	.697	.876	1.363	1.796	2.201	2.593	3.106	3.497	4.437
12	.695	.873	1.356	1.782	2.179	2.560	3.055	3.428	4.318
13	.694	.870	1.350	1.771	2.160	2.533	3.012	3.372	4.221
14	.692	.868	1.345	1.761	2.145	2.510	2.977	3.326	4.140
15	.691	.866	1.341	1.753	2.131	2.490	2.947	3.286	4.073
16	.690	.865	1.337	1.746	2.120	2.473	2.921	3.252	4.015
17	.689	.863	1.333	1.740	2.110	2.458	2.898	3.222	3.965
18	.688	.862	1.330	1.734	2.101	2.445	2.878	3.197	3.922
19	.688	.861	1.328	1.729	2.093	2.433	2.861	3.174	3.883
20	.687	.860	1.325	1.725	2.086	2.423	2.845	3.153	3.850
21	.686	.859	1.323	1.721	2.080	2.414	2.831	3.135	3.819
22	.686	.858	1.321	1.717	2.074	2.406	2.819	3.119	3.792
23	.685	.858	1.319	1.714	2.069	2.398	2.807	3.104	3.767
24	.685	.857	1.318	1.711	2.064	2.391	2.797	3.090	3.745
25	.684	.856	1.316	1.708	2.060	2.385	2.787	3.078	3.725
26	.684	.856	1.315	1.706	2.056	2.379	2.779	3.067	3.707
27	.684	.855	1.314	1.703	2.052	2.373	2.771	3.056	3.690
28	.683	.855	1.313	1.701	2.048	2.368	2.763	3.047	3.674
29	.683	.854	1.311	1.699	2.045	2.364	2.756	3.038	3.659
30	.683	.854	1.310	1.697	2.042	2.360	2.750	3.030	3.646
35	.682	.852	1.306	1.690	2.030	2.342	2.724	2.996	3.591
40	.681	.851	1.303	1.684	2.021	2.329	2.704	2.971	3.551
45	.680	.850	1.301	1.680	2.014	2.319	2.690	2.952	3.520
50	.680	.849	1.299	1.676	2.008	2.310	2.678	2.937	3.496
55	.679	.849	1.297	1.673	2.004	2.304	2.669	2.925	3.476
60	.679	.848	1.296	1.671	2.000	2.299	2.660	2.915	3.460
70	.678	.847	1.294	1.667	1.994	2.290	2.648	2.899	3.435
80	.678	.847	1.293	1.665	1.989	2.284	2.638	2.887	3.416
90	.678	.846	1.291	1.662	1.986	2.279	2.631	2.878	3.402
100	.677	.846	1.290	1.661	1.982	2.276	2.625	2.871	3.390
120	.677	.845	1.289	1.658	1.980	2.270	2.617	2.860	3.373
∞	.6745	.8416	1.2816	1.6448	1.9600	2.2414	2.5758	2.8070	3.2905

คัดแปลงจาก Snedecor และ Cochran (1967) ซึ่งคัดแปลงมาจาก R. A. Fisher's Statistical Method for Research Workers พิมพ์โดย Oliven and Boyd (1925 –50)

ตาราง ผ.5 การกระจายของค่าปัวซอง

X	μ = 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	.9048	.8187	.7408	.6703	.6065	.5488	.4966	.4493	.4066	.3679
1	.0905	.1637	.2222	.2681	.3033	.3293	.3476	.3595	.3659	.3679
2	.0045	.0164	.0333	.0536	.0758	.0988	.1217	.1438	.1647	.1839
3	.0002	.0011	.0033	.0072	.0126	.0198	.0284	.0383	.0494	.0613
4	.0000	.0001	.0002	.0007	.0016	.0030	.0050	.0077	.0111	.0153
5	.0000	.0000	.0000	.0001	.0002	.0004	.0007	.0012	.0020	.0031
6	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0003	.0005
7	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

X	μ = 1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0	.3329	.3012	.2725	.2466	.2231	.2019	.1827	.1653	.1496	.1353
1	.3362	.3614	.3543	.3452	.3347	.3230	.3106	.2975	.2842	.2707
2	.2014	.2169	.2303	.2417	.2510	.2584	.2640	.2678	.2700	.2707
3	.0738	.0867	.0998	.1128	.1255	.1378	.1496	.1607	.1710	.1804
4	.0203	.0260	.0324	.0395	.0471	.0551	.0636	.0723	.0812	.0902
5	.0045	.0062	.0084	.0111	.0141	.0176	.0216	.0260	.0309	.0361
6	.0008	.0012	.0018	.0026	.0035	.0047	.0061	.0078	.0098	.0120
7	.0001	.0002	.0003	.0005	.0008	.0011	.0015	.0020	.0027	.0034
8	.0000	.0000	.0001	.0001	.0001	.0002	.0003	.0005	.0006	.0009
9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0002

X	μ = 2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
0	.1225	.1108	.1003	.0907	.0821	.0743	.0627	.0608	.0550	.0498
1	.2572	.2438	.2306	.2177	.2052	.1931	.1815	.1703	.1596	.1494
2	.2700	.2681	.2652	.2613	.2565	.2510	.2450	.2384	.2314	.2240
3	.1890	.1966	.2033	.2090	.2138	.2176	.2205	.2225	.2237	.2240
4	.0992	.1082	.1169	.1254	.1336	.1414	.1488	.1557	.1622	.1680
5	.0417	.0476	.0538	.0602	.0668	.0735	.0804	.0872	.0940	.1008
6	.0146	.0174	.0206	.0241	.0278	.0319	.0362	.0407	.0455	.0504
7	.0044	.0055	.0068	.0083	.0099	.0118	.0139	.0163	.0188	.0216
8	.0011	.0015	.0019	.0025	.0031	.0038	.0047	.0057	.0068	.0081
9	.0003	.0004	.0005	.0007	.0009	.0011	.0014	.0018	.0022	.0027
10	.0001	.0001	.0001	.0002	.0002	.0003	.0004	.0005	.0006	.0008
11	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0002	.0002
12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

ตาราง ผ.6 ตารางทดสอบดรรชนีสหสัมพันธ์ที่ 5 และ 1 เปอร์เซ็นต์

d.f.[1]	5%	1%	d.f.	5%	1%
1	.997	1.000	26	.374	.478
2	.950	.990	27	.367	.470
3	.878	.959	28	.361	.463
4	.811	.917	29	.355	.456
5	.754	.874	30	.349	.449
6	.707	.834	32	.339	.437
7	.666	.798	34	.329	.424
8	.632	.765	36	.321	.413
9	.602	.735	38	.312	.403
10	.576	.708	40	.304	.393
11	.553	.684	45	.288	.372
12	.532	.661	50	.273	.354
13	.514	.641	55	.262	.340
14	.497	.623	60	.250	.325
15	.482	.606	70	.232	.302
16	.468	.590	80	.217	.283
17	.456	.575	90	.205	.267
18	.444	.561	100	.195	.254
19	.433	.549	125	.174	.228
20	.423	.537	150	.159	.208
21	.413	.526	175	.148	.194
22	.404	.515	200	.138	.181
23	.396	.505	300	.113	.148
24	.388	.496	400	.098	.128
25	.381	.487	500	.088	.115

[1] d.f. = n - 2

ตาราง ผ.7 ตารางแปลงค่า r เป็น z

r	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
.0	0.000	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090
.1	.100	.110	.121	.131	.141	.151	.161	.172	.182	.192
.2	.203	.213	.224	.234	.245	.255	.266	.277	.288	.299
.3	.310	.321	.332	.343	.354	.365	.377	.388	.400	.412
.4	.424	.436	.448	.460	.472	.485	.497	.510	.523	.536
.5	.549	.563	.576	.590	.604	.618	.633	.648	.662	.678
.6	.693	.709	.725	.741	.758	.775	.793	.811	.829	.848
.7	.867	.887	.908	.929	.950	.973	.996	1.020	1.045	1.071
.8	1.099	1.127	1.157	1.188	1.221	1.256	1.293	1.333	1.376	1.422

r	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
.90	1.472	1.478	1.483	1.488	1.494	1.499	1.505	1.510	1.516	1.522
.91	1.528	1.533	1.539	1.545	1.551	1.557	1.564	1.570	1.576	1.583
.92	1.589	1.596	1.602	1.609	1.616	1.623	1.630	1.637	1.644	1.651
.93	1.658	1.666	1.673	1.681	1.689	1.697	1.705	1.713	1.721	1.730
.94	1.738	1.747	1.756	1.764	1.774	1.783	1.792	1.802	1.812	1.822
.95	1.832	1.842	1.853	1.863	1.874	1.886	1.897	1.909	1.921	1.933
.96	1.946	1.959	1.972	1.986	2.000	2.014	2.029	2.044	2.060	2.076
.97	2.092	2.109	2.127	2.146	2.165	2.185	2.205	2.227	2.249	2.273
.98	2.298	2.323	2.351	2.380	2.410	2.443	2.477	2.515	2.555	2.599
.99	2.646	2.700	2.759	2.826	2.903	2.994	3.106	3.250	3.453	3.800

ตาราง ผ.8 ตารางแปลงค่า z เป็น r

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.000	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090
.1	.100	.110	.119	.129	.139	.149	.159	.168	.178	.187
.2	.197	.207	.216	.226	.236	.245	.254	.264	.273	.282
.3	.291	.300	.310	.319	.327	.336	.345	.354	.363	.371
.4	.380	.389	.397	.405	.414	.422	.430	.438	.446	.454
.5	.462	.470	.478	.485	.493	.500	.508	.515	.523	.530
.6	.537	.544	.551	.558	.565	.572	.578	.585	.592	.598
.7	.604	.611	.617	.623	.629	.635	.641	.647	.653	.658
.8	.664	.670	.675	.680	.686	.691	.696	.701	.706	.711
.9	.716	.721	.726	.731	.735	.740	.744	.749	.753	.757
1.0	.762	.766	.770	.774	.778	.782	.786	.790	.793	.797
1.1	.800	.804	.808	.811	.814	.818	.821	.824	.828	.831
1.2	.834	.837	.840	.843	.846	.848	.851	.854	.856	.859
1.3	.862	.864	.867	.869	.872	.874	.876	.879	.881	.883
1.4	.885	.888	.890	.892	.894	.896	.898	.900	.902	.903
1.5	.905	.907	.909	.910	.912	.914	.915	.917	.919	.920
1.6	.922	.923	.925	.926	.928	.929	.930	.932	.933	.934
1.7	.935	.937	.938	.939	.940	.941	.942	.944	.945	.946
1.8	.947	.948	.949	.950	.951	.952	.953	.954	.954	.955
1.9	.956	.957	.958	.959	.960	.960	.961	.962	.963	.963
2.0	.964	.965	.965	.966	.967	.967	.968	.969	.969	.970
2.1	.970	.971	.972	.972	.973	.973	.974	.974	.975	.975
2.2	.976	.976	.977	.977	.978	.978	.978	.979	.979	.980
2.3	.980	.980	.981	.981	.982	.982	.982	.983	.983	.983
2.4	.984	.984	.984	.985	.985	.985	.986	.986	.986	.986
2.5	.987	.987	.987	.987	.988	.988	.988	.988	.989	.989
2.6	.989	.989	.989	.990	.990	.990	.990	.990	.991	.991
2.7	.991	.991	.991	.992	.992	.992	.992	.992	.992	.992
2.8	.993	.993	.993	.993	.993	.993	.993	.994	.994	.994
2.9	.994	.994	.994	.994	.994	.995	.995	.995	.995	.995

* $r = (e^{2z} - 1)/(e^{2z} + 1)$.

ตาราง ผ.9 การกระจายของค่า F

f_2	f_1 Degrees of Freedom (for greater mean square)																			
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75
1	161	200	216	225	230	234	237	239	241	242	243	244	245	246	248	249	250	251	252	253
	4,052	4,999	5,403	5,625	5,764	5,859	5,928	5,981	6,022	6,056	6,082	6,106	6,142	6,169	6,208	6,234	6,261	6,286	6,302	6,323
2	18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39	19.40	19.41	19.42	19.43	19.44	19.45	19.46	19.47	19.47	19.48
	98.49	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.41	99.42	99.43	99.44	99.45	99.46	99.47	99.48	99.48	99.49
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74	8.71	8.69	8.66	8.64	8.62	8.60	8.58	8.57
	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.13	27.05	26.92	26.83	26.69	26.60	26.50	26.41	26.35	26.27
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91	5.87	5.84	5.80	5.77	5.74	5.71	5.70	5.68
	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.54	14.45	14.37	14.24	14.15	14.02	13.93	13.83	13.74	13.69	13.61
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70	4.68	4.64	4.60	4.56	4.53	4.50	4.46	4.44	4.42
	16.26	13.27	12.06	11.39	10.97	10.67	10.45	10.29	10.15	10.05	9.96	9.89	9.77	9.68	9.55	9.47	9.38	9.29	9.24	9.17
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.96	3.92	3.87	3.84	3.81	3.77	3.75	3.72
	13.74	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.60	7.52	7.39	7.31	7.23	7.14	7.09	7.02
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.60	3.57	3.52	3.49	3.44	3.41	3.38	3.34	3.32	3.29
	12.25	9.55	8.45	7.85	7.46	7.19	7.00	6.84	6.71	6.62	6.54	6.47	6.35	6.27	6.15	6.07	5.98	5.90	5.85	5.78
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.31	3.28	3.23	3.20	3.15	3.12	3.08	3.05	3.03	3.00
	11.26	8.65	7.59	7.01	6.63	6.37	6.19	6.03	5.91	5.82	5.74	5.67	5.56	5.48	5.36	5.28	5.20	5.11	5.06	5.00
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13	3.10	3.07	3.02	2.98	2.93	2.90	2.86	2.82	2.80	2.77
	10.56	8.02	6.99	6.42	6.06	5.80	5.62	5.47	5.35	5.26	5.18	5.11	5.00	4.92	4.80	4.73	4.64	4.56	4.51	4.45
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97	2.94	2.91	2.86	2.82	2.77	2.74	2.70	2.67	2.64	2.61
	10.04	7.56	6.55	5.99	5.64	5.39	5.21	5.06	4.95	4.85	4.78	4.71	4.60	4.52	4.41	4.33	4.25	4.17	4.12	4.05
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86	2.82	2.79	2.74	2.70	2.65	2.61	2.57	2.53	2.50	2.47
	9.65	7.20	6.22	5.67	5.32	5.07	4.88	4.74	4.63	4.54	4.46	4.40	4.29	4.21	4.10	4.02	3.94	3.86	3.80	3.74
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76	2.72	2.69	2.64	2.60	2.54	2.50	2.46	2.42	2.40	2.36
	9.33	6.93	5.95	5.41	5.06	4.82	4.65	4.50	4.39	4.30	4.22	4.16	4.05	3.98	3.86	3.78	3.70	3.61	3.56	3.49
13	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.72	2.67	2.63	2.60	2.55	2.51	2.46	2.42	2.38	2.34	2.32	2.28
	9.07	6.70	5.74	5.20	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.85	3.78	3.67	3.59	3.51	3.42	3.37	3.30
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56	2.53	2.48	2.44	2.39	2.35	2.31	2.27	2.24	2.21
	8.86	6.51	5.56	5.03	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.70	3.62	3.51	3.43	3.34	3.26	3.21	3.14
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55	2.51	2.48	2.43	2.39	2.33	2.29	2.25	2.21	2.18	2.15
	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.56	3.48	3.36	3.29	3.20	3.12	3.07	3.00
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42	2.37	2.33	2.28	2.24	2.20	2.16	2.13	2.09
	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.61	3.55	3.45	3.37	3.25	3.18	3.10	3.01	2.96	2.98
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45	2.41	2.38	2.33	2.29	2.23	2.19	2.15	2.11	2.08	2.04
	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.45	3.35	3.27	3.16	3.08	3.00	2.92	2.86	2.79
18	4.41	3.55	3.16	2.93	2.77	3.66	2.58	2.51	2.46	2.41	2.37	2.34	2.29	2.25	2.19	2.15	2.11	2.07	2.04	2.00
	8.28	6.01	5.09	4.58	4.25	4.01	3.85	3.71	3.60	3.51	3.44	3.37	3.27	3.19	3.07	3.00	2.91	2.83	2.78	2.71
19	4.38	3.52	3.13	2.90	2.74	2.63	2.55	2.48	2.43	2.38	2.34	2.31	2.26	2.21	2.15	2.11	2.07	2.02	2.00	1.96
	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.19	3.12	3.00	2.92	2.84	2.76	2.70	2.63
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.40	2.35	2.31	2.28	2.23	2.18	2.12	2.08	2.04	1.99	1.96	1.92
	8.10	5.85	4.94	4.43	4.10	3.87	3.71	3.56	3.45	3.37	3.30	3.23	3.13	3.05	2.94	2.86	2.77	2.69	2.63	2.56
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.20	2.15	2.09	2.05	2.00	1.96	1.93	1.89
	8.02	5.78	4.87	4.37	4.04	3.81	3.65	3.51	3.40	3.31	3.24	3.17	3.07	2.99	2.88	2.80	2.72	2.63	2.58	2.51
22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30	2.26	2.23	2.18	2.13	2.07	2.03	1.98	1.93	1.91	1.87
	7.94	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.02	2.94	2.83	2.75	2.67	2.58	2.53	2.46
23	4.28	3.42	3.03	2.80	2.64	2.53	2.45	2.38	2.32	2.28	2.24	2.70	2.14	2.10	2.04	2.00	1.96	1.91	1.88	1.84
	[illegible]	[illegible]	4.76	[illegible]	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	2.97	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
24	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.30	2.26	2.22	2.18	2.13	2.09	2.02	1.98	1.94	1.89	1.86	1.82
	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.25	3.17	3.09	3.03	2.93	2.85	2.74	2.66	2.58	2.49	2.44	2.36
25	4.24	3.38	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24	2.20	2.16	2.11	2.06	2.00	1.96	1.92	1.87	1.84	1.80
	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.21	3.13	3.05	2.99	2.89	2.81	2.70	2.62	2.54	2.45	2.40	2.32
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.10	2.05	1.99	1.95	1.90	1.85	1.82	1.78
	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.17	3.09	3.02	2.96	2.86	2.77	2.66	2.58	2.50	2.41	2.36	2.28

ตาราง ผ.9 (ต่อ)

f_2	f_1 Degrees of Freedom (for greater mean square)																			
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.30	2.25	2.20	2.16	2.13	2.08	2.03	1.97	1.93	1.88	1.84	1.80	1.76
	7.68	**5.49**	**4.60**	**4.11**	**3.79**	**3.56**	**3.39**	**3.26**	**3.14**	**3.06**	**2.98**	**2.93**	**2.83**	**2.74**	**2.63**	**2.55**	**2.47**	**2.38**	**2.33**	**2.25**
28	4.20	3.34	2.95	2.71	2.56	2.44	2.36	2.29	2.24	2.19	2.15	2.12	2.06	2.02	1.96	1.91	1.87	1.81	1.78	1.75
	7.64	**5.45**	**4.57**	**4.07**	**3.76**	**3.53**	**3.36**	**3.23**	**3.11**	**3.03**	**2.95**	**2.90**	**2.80**	**2.71**	**2.60**	**2.52**	**2.44**	**2.35**	**2.30**	**2.22**
29	4.18	3.33	2.93	2.70	2.54	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.05	2.00	1.94	1.90	1.85	1.80	1.77	1.73
	7.60	**5.42**	**4.54**	**4.04**	**3.73**	**3.50**	**3.33**	**3.20**	**3.08**	**3.00**	**2.92**	**2.87**	**2.77**	**2.68**	**2.57**	**2.49**	**2.41**	**2.32**	**2.27**	**2.19**
30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.27	2.21	2.16	2.12	2.09	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72
	7.56	**5.39**	**4.51**	**4.02**	**3.70**	**3.47**	**3.30**	**3.17**	**3.06**	**2.98**	**2.90**	**2.84**	**2.74**	**2.66**	**2.55**	**2.47**	**2.38**	**2.29**	**2.24**	**2.16**
32	4.15	3.30	2.90	2.67	2.51	2.40	2.32	2.25	2.19	2.14	2.10	2.07	2.02	1.97	1.91	1.86	1.82	1.76	1.74	1.69
	7.50	**5.34**	**4.46**	**3.97**	**3.66**	**3.42**	**3.25**	**3.12**	**3.01**	**2.94**	**2.86**	**2.80**	**2.70**	**2.62**	**2.51**	**2.42**	**2.34**	**2.25**	**2.20**	**2.12**
34	4.13	3.28	2.88	2.65	2.49	2.38	2.30	2.23	2.17	2.12	2.08	2.05	2.00	1.95	1.89	1.84	1.80	1.74	1.71	1.67
	7.44	**5.29**	**4.42**	**3.93**	**3.61**	**3.38**	**3.21**	**3.08**	**2.97**	**2.89**	**2.82**	**2.76**	**2.66**	**2.58**	**2.47**	**2.38**	**2.30**	**2.21**	**2.15**	**2.08**
36	4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10	2.06	2.03	1.98	1.93	1.87	1.82	1.78	1.72	1.69	1.65
	7.39	**5.25**	**4.38**	**3.89**	**3.58**	**3.35**	**3.18**	**3.04**	**2.94**	**2.86**	**2.78**	**2.72**	**2.62**	**2.54**	**2.43**	**2.35**	**2.26**	**2.17**	**2.12**	**2.04**
38	4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.80	1.76	1.71	1.67	1.63
	7.35	**5.21**	**4.34**	**3.86**	**3.54**	**3.32**	**3.15**	**3.02**	**2.91**	**2.82**	**2.75**	**2.69**	**2.59**	**2.51**	**2.40**	**2.32**	**2.22**	**2.14**	**2.08**	**2.00**
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04	2.00	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61
	7.31	**5.18**	**4.31**	**3.83**	**3.51**	**3.29**	**3.12**	**2.99**	**2.88**	**2.80**	**2.73**	**2.66**	**2.56**	**2.49**	**2.37**	**2.29**	**2.20**	**2.11**	**2.05**	**1.97**
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.02	1.99	1.94	1.89	1.82	1.78	1.73	1.68	1.64	1.60
	7.27	**5.15**	**4.29**	**3.80**	**3.49**	**3.26**	**3.10**	**2.96**	**2.86**	**2.77**	**2.70**	**2.64**	**2.54**	**2.46**	**2.35**	**2.26**	**2.17**	**2.08**	**2.02**	**1.94**
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.92	1.88	1.81	1.76	1.72	1.66	1.63	1.58
	7.24	**5.12**	**4.26**	**3.78**	**3.46**	**3.24**	**3.07**	**2.94**	**2.84**	**2.75**	**2.68**	**2.62**	**2.52**	**2.44**	**2.32**	**2.24**	**2.15**	**2.06**	**2.00**	**1.92**
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04	2.00	1.97	1.91	1.87	1.80	1.75	1.71	1.65	1.62	1.57
	7.21	**5.10**	**4.24**	**3.76**	**3.44**	**3.22**	**3.05**	**2.92**	**2.82**	**2.73**	**2.66**	**2.60**	**2.50**	**2.42**	**2.30**	**2.22**	**2.13**	**2.04**	**1.98**	**1.90**
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03	1.99	1.96	1.90	1.86	1.79	1.74	1.70	1.64	1.61	1.56
	7.19	**5.08**	**4.22**	**3.74**	**3.42**	**3.20**	**3.04**	**2.90**	**2.80**	**2.71**	**2.64**	**2.58**	**2.48**	**2.40**	**2.28**	**2.20**	**2.11**	**2.02**	**1.96**	**1.88**
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02	1.98	1.95	1.90	1.85	1.78	1.74	1.69	1.63	1.60	1.55
	7.17	**5.06**	**4.20**	**3.72**	**3.41**	**3.18**	**3.02**	**2.88**	**2.78**	**2.70**	**2.62**	**2.56**	**2.46**	**2.39**	**2.26**	**2.18**	**2.10**	**2.00**	**1.94**	**1.86**
55	4.02	3.17	2.78	2.54	2.38	2.27	2.18	2.11	2.05	2.00	1.97	1.93	1.88	1.83	1.76	1.72	1.67	1.61	1.58	1.52
	7.12	**5.01**	**4.16**	**3.68**	**3.37**	**3.15**	**2.98**	**2.85**	**2.75**	**2.66**	**2.59**	**2.53**	**2.43**	**2.35**	**2.23**	**2.15**	**2.06**	**1.96**	**1.90**	**1.82**
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.86	1.81	1.75	1.70	1.65	1.59	1.56	1.50
	7.08	**4.98**	**4.13**	**3.65**	**3.34**	**3.12**	**2.95**	**2.82**	**2.72**	**2.63**	**2.56**	**2.50**	**2.40**	**2.32**	**2.20**	**2.12**	**2.03**	**1.93**	**1.87**	**1.79**
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.02	1.98	1.94	1.90	1.85	1.80	1.73	1.68	1.63	1.57	1.54	1.49
	7.04	**4.95**	**4.10**	**3.62**	**3.31**	**3.09**	**2.93**	**2.79**	**2.70**	**2.61**	**2.54**	**2.47**	**2.37**	**2.30**	**2.18**	**2.09**	**2.00**	**1.90**	**1.84**	**1.76**
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	1.97	1.93	1.89	1.84	1.79	1.72	1.67	1.62	1.56	1.53	1.47
	7.01	**4.92**	**4.08**	**3.60**	**3.29**	**3.07**	**2.91**	**2.77**	**2.67**	**2.59**	**2.51**	**2.45**	**2.35**	**2.28**	**2.15**	**2.07**	**1.98**	**1.88**	**1.82**	**1.74**
80	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95	1.91	1.88	1.82	1.77	1.70	1.65	1.60	1.54	1.51	1.45
	6.96	**4.88**	**4.04**	**3.56**	**3.25**	**3.04**	**2.87**	**2.74**	**2.64**	**2.55**	**2.48**	**2.41**	**2.32**	**2.24**	**2.11**	**2.03**	**1.94**	**1.84**	**1.78**	**1.70**
100	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.92	1.88	1.85	1.79	1.75	1.68	1.63	1.57	1.51	1.48	1.42
	6.90	**4.82**	**3.98**	**3.51**	**3.20**	**2.99**	**2.82**	**2.69**	**2.59**	**2.51**	**2.43**	**2.36**	**2.26**	**2.19**	**2.06**	**1.98**	**1.89**	**1.79**	**1.73**	**1.64**
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.95	1.90	1.86	1.83	1.77	1.72	1.65	1.60	1.55	1.49	1.45	1.39
	6.84	**4.78**	**3.94**	**3.47**	**3.17**	**2.95**	**2.79**	**2.65**	**2.56**	**2.47**	**2.40**	**2.33**	**2.23**	**2.15**	**2.03**	**1.94**	**1.85**	**1.75**	**1.68**	**1.59**
150	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.76	1.71	1.64	1.59	1.54	1.47	1.44	1.37
	6.81	**4.75**	**3.91**	**3.44**	**3.14**	**2.92**	**2.76**	**2.62**	**2.53**	**2.44**	**2.37**	**2.30**	**2.20**	**2.12**	**2.00**	**1.91**	**1.83**	**1.72**	**1.66**	**1.56**
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87	1.83	1.80	1.74	1.69	1.62	1.57	1.52	1.45	1.42	1.35
	6.76	**4.71**	**3.88**	**3.41**	**3.11**	**2.90**	**2.73**	**2.60**	**2.50**	**2.41**	**2.34**	**2.28**	**2.17**	**2.09**	**1.97**	**1.88**	**1.79**	**1.69**	**1.62**	**1.53**
400	3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81	1.78	1.72	1.67	1.60	1.54	1.49	1.42	1.38	1.32
	6.70	**4.66**	**3.83**	**3.36**	**3.06**	**2.85**	**2.69**	**2.55**	**2.46**	**2.37**	**2.29**	**2.23**	**2.12**	**2.04**	**1.92**	**1.84**	**1.74**	**1.64**	**1.57**	**1.47**
1000	3.85	3.00	2.61	2.38	2.22	2.10	2.02	1.95	1.89	1.84	1.80	1.76	1.70	1.65	1.58	1.53	1.47	1.41	1.36	1.30
	6.66	**4.62**	**3.80**	**3.34**	**3.04**	**2.82**	**2.66**	**2.53**	**2.43**	**2.34**	**2.26**	**2.20**	**2.09**	**2.01**	**1.89**	**1.81**	**1.71**	**1.61**	**1.54**	**1.44**
∞	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.83	1.79	1.75	1.69	1.64	1.57	1.52	1.46	1.40	1.35	1.28
	6.64	**4.60**	**3.78**	**3.32**	**3.02**	**2.80**	**2.64**	**2.51**	**2.41**	**2.32**	**2.24**	**2.18**	**2.07**	**1.99**	**1.87**	**1.79**	**1.69**	**1.59**	**1.52**	**1.41**

ตาราง ผ.10 ตาราง Significant Studentized Ranges (SSR) สำหรับ New Multiple – Range Test

Error df	Protection level	p = number of means for range being tested													
		2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	.05	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0
	.01	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0
2	.05	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
	.01	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0
3	.05	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
	.01	8.26	8.5	8.6	8.7	8.8	8.9	8.9	9.0	9.0	9.0	9.1	9.2	9.3	9.3
4	.05	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02
	.01	6.51	6.8	6.9	7.0	7.1	7.1	7.2	7.2	7.3	7.3	7.4	7.4	7.5	7.5
5	.05	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83
	.01	5.70	5.96	6.11	6.18	6.26	6.33	6.40	6.44	6.5	6.6	6.6	6.7	6.7	6.8
6	.05	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68
	.01	5.24	5.51	5.65	5.73	5.81	5.88	5.95	6.00	6.0	6.1	6.2	6.2	6.3	6.3
7	.05	3.35	3.47	3.54	3.58	3.60	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61
	.01	4.95	5.22	5.37	5.45	5.53	5.61	5.69	5.73	5.8	5.8	5.9	5.9	6.0	6.0
8	.05	3.26	3.39	3.47	3.52	3.55	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56
	.01	4.74	5.00	5.14	5.23	5.32	5.40	5.47	5.51	5.5	5.6	5.7	5.7	5.8	5.8
9	.05	3.20	3.34	3.41	3.47	3.50	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52
	.01	4.60	4.86	4.99	5.08	5.17	5.25	5.32	5.36	5.4	5.5	5.5	5.6	5.7	5.7
10	.05	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.48
	.01	4.48	4.73	4.88	4.96	5.06	5.13	5.20	5.24	5.28	5.36	5.42	5.48	5.54	5.55
11	.05	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	4.46	3.46	3.46	3.46	3.47	3.48
	.01	4.39	4.63	4.77	4.86	4.94	5.01	5.06	5.12	5.15	5.24	5.28	5.34	5.38	5.39
12	.05	3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.44	3.46	3.46	3.46	3.46	3.47	3.48
	.01	4.32	4.55	4.68	4.76	4.81	4.92	4.96	5.02	5.07	5.13	5.17	5.22	5.24	5.26
13	.05	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.45	3.46	3.46	3.47	3.47
	.01	4.26	4.48	4.62	4.69	4.74	4.84	4.88	4.94	4.98	5.04	5.08	5.13	5.14	5.15
14	.05	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.45	3.46	3.46	3.47	3.47
	.01	4.21	4.42	4.55	4.63	4.70	4.78	4.83	3.87	4.91	4.96	5.00	5.04	5.06	5.07
15	.05	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.44	3.45	3.46	3.47	3.47
	.01	4.17	4.37	4.50	4.58	4.64	4.72	4.77	4.81	4.84	4.90	4.94	4.97	4.99	5.00
16	.05	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.44	3.45	3.46	3.47	3.47
	.01	4.13	4.34	4.45	4.54	4.60	4.67	4.72	4.76	4.79	4.84	4.88	4.91	4.93	4.94
17	.05	2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.44	3.45	3.46	3.47	3.47
	.01	4.10	4.30	4.41	4.50	4.56	4.63	4.68	4.72	4.75	4.80	4.83	4.86	4.88	4.89
18	.05	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.43	3.45	3.46	3.47	3.47
	.01	4.07	4.27	4.38	4.46	4.53	4.59	4.64	4.68	4.71	4.76	4.79	4.82	4.84	4.86
19	.05	2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.43	3.44	3.46	3.47	3.47
	.01	4.05	4.24	4.35	4.43	4.50	4.56	4.61	4.64	4.67	4.72	4.76	4.79	4.81	4.82
20	.05	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.43	3.44	3.46	3.46	3.47
	.01	4.02	4.22	4.33	4.40	4.47	4.53	4.58	4.61	4.65	4.69	4.73	3.76	4.78	4.79
22	.05	2.93	3.08	3.17	3.24	3.29	3.32	3.35	3.37	3.39	3.42	3.44	3.45	3.46	3.47
	.01	3.99	4.17	4.28	4.36	4.42	4.48	4.53	4.57	4.60	4.65	4.68	4.71	4.74	4.75
24	.05	2.92	3.07	3.15	3.22	3.28	3.31	3.34	3.37	3.38	3.41	3.44	3.45	3.46	3.47
	.01	3.96	4.14	4.24	4.33	4.39	4.44	4.49	4.53	4.57	4.62	4.64	4.67	4.70	4.72
26	.05	2.91	3.06	3.14	3.21	3.27	3.30	3.34	3.36	3.38	3.41	3.43	3.45	3.46	3.47
	.01	3.93	4.11	4.21	4.30	4.36	4.41	4.46	4.50	4.53	4.58	4.62	4.65	4.67	4.69
28	.05	2.90	3.04	3.13	3.20	3.26	3.30	3.33	3.35	3.37	3.40	3.43	3.45	3.46	3.47
	.01	3.91	4.08	4.18	4.28	4.34	4.39	4.43	4.47	4.51	4.56	4.60	4.62	4.65	4.67
30	.05	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.40	3.43	3.44	3.46	3.47
	.01	3.89	4.06	4.16	4.22	4.32	4.36	4.41	4.45	4.48	4.54	4.58	4.61	4.63	4.65
40	.05	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.39	3.42	3.44	3.46	3.47
	.01	3.82	3.99	4.10	4.17	4.21	4.30	4.34	4.37	4.41	4.46	4.51	4.54	4.57	4.59
60	.05	2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.37	3.40	3.43	3.45	3.47
	.01	3.76	3.92	4.03	4.12	4.17	4.23	4.27	4.31	4.34	4.39	4.44	4.47	4.50	4.53
100	.05	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.36	3.40	3.42	3.45	3.47
	.01	3.71	3.86	3.98	4.06	4.11	4.17	4.21	4.25	4.29	4.35	6.38	4.42	4.45	4.48
∞	.05	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.34	3.38	3.41	3.44	3.47
	.01	3.64	3.80	3.90	3.98	4.04	4.09	4.14	4.17	4.20	4.26	4.31	4.34	4.38	4.41

ตาราง ผ.11 ตาราง Arc Sine $\sqrt{\text{Percent}}$ Transformation

%	0	1	2	3	4	5	6	7	8	9
0.0	0	0.57	0.81	0.99	1.15 −	1.28	1.40	1.52	1.62	1.72
0.1	1.81	1.90	1.99	2.07	2.14	2.22	2.29	2.36	2.43	2.50
0.2	2.56	2.63	2.69	2.76 −	2.81	2.87	2.92	2.98	3.03	3.09
0.3	3.14	3.19	3.24	3.29	3.34	3.39	3.44	3.49	3.53	3.58
0.4	3.63	3.67	3.72	3.76	3.80	3.85 −	3.89	3.93	3.97	4.01
0.5	4.05 +	4.09	4.13	4.17	4.21	4.25 +	4.29	4.33	4.37	4.40
0.6	4.44	4.48	4.52	4.55 +	4.59	4.62	4.66	4.69	4.73	4.76
0.7	4.80	4.83	4.87	4.90	4.93	4.97	5.00	5.03	5.07	5.10
0.8	5.13	5.16	5.20	5.23	5.26	5.29	5.32	5.35 +	5.38	5.41
0.9	5.44	5.47	5.50	5.53	5.56	5.59	5.62	5.65 +	5.68	5.71
1	5.74	6.02	6.29	6.55 −	6.80	7.04	7.27	7.49	7.71	7.92
2	8.13	8.33	8.53	8.72	8.91	9.10	9.28	9.46	9.63	9.81
3	9.98	10.14	10.31	10.47	10.63	10.78	10.94	11.09	11.24	11.39
4	11.54	11.68	11.83	11.97	12.11	12.25 −	12.39	12.52	12.66	12.79
5	12.92	13.05 +	13.18	13.31	13.44	13.56	13.69	13.81	13.94	14.06
6	14.18	14.30	14.42	14.54	14.65 +	14.77	14.89	15.00	15.12	15.23
7	15.34	15.45 +	15.56	15.68	15.79	15.89	16.00	16.11	16.22	16.32
8	16.43	16.54	16.64	16.74	16.85 −	16.95 +	17.05 +	17.16	17.26	17.36
9	17.46	17.56	17.66	17.76	17.85 +	17.95 −	18.05 −	18.15 −	18.24	18.34
10	18.44	18.53	18.63	18.72	18.81	18.91	19.00	19.09	19.19	19.28
11	19.37	19.46	19.55 +	19.64	19.73	19.82	19.91	20.00	20.09	20.18
12	20.27	20.36	20.44	20.53	20.62	20.70	20.79	20.88	20.96	21.05 −
13	21.13	21.22	21.30	21.39	21.47	21.56	21.64	21.72	21.81	21.89
14	21.97	22.06	22.14	22.22	22.30	22.38	22.46	22.55 −	22.63	22.71
15	22.79	22.87	22.95 −	23.03	23.11	23.19	23.26	23.34	23.42	23.50
16	23.58	23.66	23.73	23.81	23.89	23.97	24.04	24.12	24.20	24.27
17	24.35 +	24.43	24.50	24.58	24.65 +	24.73	24.80	24.88	24.95 +	25.03
18	25.10	25.18	25.25 +	25.33	25.40	25.48	25.55 −	25.62	25.70	25.77
19	25.84	25.92	25.99	26.06	26.13	26.21	26.28	26.35 −	26.42	26.49
20	26.56	26.64	26.71	26.78	26.85 +	26.92	26.99	27.06	27.13	27.20
21	27.28	27.35 −	27.42	27.49	27.56	27.63	27.69	27.76	27.83	27.90
22	27.97	28.04	28.11	28.18	28.25 −	28.32	28.38	28.45 +	28.52	28.59
23	28.66	28.73	28.79	28.86	28.93	29.00	29.06	29.13	29.20	29.27
24	29.33	29.40	29.47	29.53	29.60	29.67	29.73	29.80	29.87	29.93
25	30.00	30.07	30.13	30.20	30.26	30.33	30.40	30.46	30.53	30.59
26	30.66	30.72	30.79	30.85 +	30.92	30.98	31.05 −	31.11	31.18	31.24
27	31.31	31.37	31.44	31.50	31.56	31.63	31.69	31.76	31.82	31.88
28	31.95 −	32.01	32.08	32.14	32.20	32.27	32.33	32.39	32.46	32.52
29	32.58	32.65 −	32.71	32.77	32.83	32.90	32.96	33.02	33.09	33.15 −
30	33.21	33.27	33.34	33.40	33.46	33.52	33.58	33.65 −	33.71	33.77
31	33.83	33.89	33.96	34.02	34.08	34.14	34.20	34.27	34.33	34.39
32	34.45 −	34.51	34.57	34.63	34.70	34.76	34.82	34.88	34.94	35.00
33	35.06	35.12	35.18	35.24	35.30	35.37	35.43	35.49	35.55 −	35.61
34	35.67	35.73	35.79	35.85	35.91	35.97	36.03	36.09	36.15 +	36.21
35	36.27	36.33	36.39	36.45 −	36.51	36.57	36.63	36.69	36.75 +	36.81
36	36.87	36.93	36.99	37.05	37.11	37.17	37.23	37.29	37.35 −	37.41
37	37.47	37.52	37.58	37.64	37.70	37.76	37.82	37.88	37.94	38.00
38	38.06	38.12	38.17	38.23	38.29	38.35 +	38.41	38.47	38.53	38.59
39	38.65 −	38.70	38.76	38.82	38.88	38.94	39.00	39.06	39.11	39.17
40	39.23	39.29	39.35	39.41	39.47	39.52	39.58	39.64	39.70	39.76
41	39.82	39.87	39.93	39.99	40.05 −	40.11	40.16	40.22	40.28	40.34
42	40.40	40.46	40.51	40.57	40.63	40.69	40.74	40.80	40.86	40.92
43	40.98	41.03	41.09	41.15 −	41.21	41.27	41.32	41.38	41.44	41.50
44	41.55 +	41.61	41.67	41.73	41.78	41.84	41.90	41.96	42.02	42.07
45	42.13	42.19	42.25 −	42.30	42.36	42.42	42.48	42.53	42.59	42.65 −
46	42.71	42.76	42.82	42.88	42.94	42.99	43.05	43.11	43.17	43.22
47	43.28	43.34	43.39	43.45 +	43.51	43.57	43.62	43.68	43.74	43.80
48	43.85 +	43.91	43.97	44.03	44.08	44.14	44.20	44.25 −	44.31	44.37
49	44.43	44.48	44.54	44.60	44.66	44.71	44.77	44.83	44.89	44.94

ตาราง ผ.11 (ต่อ)

%	0	1	2	3	4	5	6	7	8	9
50	45.00	45.06	45.11	45.17	45.23	45.29	45.34	45.40	45.46	45.52
51	45.57	45.63	45.69	45.75-	45.80	45.86	45.92	45.97	46.03	46.09
52	46.15-	46.20	46.26	46.32	46.38	46.43	46.49	46.55-	46.61	46.66
53	46.72	46.78	46.83	46.89	46.95+	47.01	47.06	47.12	47.18	47.24
54	47.29	47.35+	47.41	47.47	47.52	47.58	47.64	47.70	47.75+	47.81
55	47.87	47.93	47.98	48.04	48.10	48.16	48.22	48.27	48.33	48.39
56	48.45-	48.50	48.56	48.62	48.68	48.73	48.79	48.85+	48.91	48.97
57	49.02	49.08	49.14	49.20	49.26	49.31	49.37	49.43	49.49	49.54
58	49.60	49.66	49.72	49.78	49.84	49.89	49.95+	50.01	50.07	50.13
59	50.18	50.24	50.30	50.36	50.42	50.48	50.53	50.59	50.65+	50.71
60	50.77	50.83	50.89	50.94	51.00	51.06	51.12	51.18	51.24	51.30
61	51.35+	51.41	51.47	51.53	51.59	51.65-	51.71	51.77	51.83	51.88
62	51.94	52.00	52.06	52.12	52.18	52.24	52.30	52.36	52.42	52.48
63	52.53	52.59	52.65+	52.71	52.77	52.83	52.89	52.95+	53.01	53.07
64	53.13	53.19	53.25-	53.31	53.37	53.43	53.49	53.55-	53.61	53.67
65	53.73	53.79	53.85-	53.91	53.97	54.03	54.09	54.15+	54.21	54.27
66	54.33	54.39	54.45+	54.51	54.57	54.63	54.70	54.76	54.82	54.88
67	54.94	55.00	55.06	55.12	55.18	55.24	55.30	55.37	55.43	55.49
68	55.55+	55.61	55.67	55.73	55.80	55.86	55.92	55.98	56.04	56.11
69	56.17	56.23	56.29	56.35+	56.42	56.48	56.54	56.60	56.66	56.73
70	56.79	56.85+	56.91	56.98	57.04	57.10	57.17	57.23	57.29	57.35+
71	57.42	57.48	57.54	57.61	57.67	57.73	57.80	57.86	57.92	57.99
72	58.05+	58.12	58.18	58.24	58.31	58.37	58.44	58.50	58.56	58.63
73	58.69	58.76	58.82	58.89	58.95+	59.02	59.08	59.15-	59.21	59.28
74	59.34	59.41	59.47	59.54	59.60	59.67	59.74	59.80	59.87	59.93
75	60.00	60.07	60.13	60.20	60.27	60.33	60.40	60.47	60.53	60.60
76	60.67	60.73	60.80	60.87	60.94	61.00	61.07	61.14	61.21	61.27
77	61.34	61.41	61.48	61.55-	61.62	61.68	61.75+	61.82	61.89	61.96
78	62.03	62.10	62.17	62.24	62.31	62.37	62.44	62.51	62.58	62.65+
79	62.72	62.80	62.87	62.94	63.01	63.08	63.15-	63.22	63.29	63.36
80	63.44	63.51	63.58	63.65+	63.72	63.79	63.87	63.94	64.01	64.08
81	64.16	64.23	64.30	64.38	64.45+	64.52	64.60	64.67	64.75-	64.82
82	64.90	64.97	65.05-	65.12	65.20	65.27	65.35-	65.42	65.50	65.57
83	65.65-	65.73	65.80	65.88	65.96	66.03	66.11	66.19	66.27	66.34
84	66.42	66.50	66.58	66.66	66.74	66.81	66.89	66.97	67.05+	67.13
85	67.21	67.29	67.37	67.45+	67.54	67.62	67.70	67.78	67.86	67.94
86	68.03	68.11	68.19	68.28	68.36	68.44	68.53	68.61	68.70	68.78
87	68.87	68.95+	69.04	69.12	69.21	69.30	69.38	69.47	69.56	69.64
88	69.73	69.82	69.91	70.00	70.09	70.18	70.27	70.36	70.45-	70.54
89	70.63	70.72	70.81	70.91	71.00	71.09	71.19	71.23	71.37	71.47
90	71.56	71.66	71.76	71.85+	71.95+	72.05-	72.15-	72.24	72.34	72.44
91	72.54	72.64	72.74	72.84	72.95-	73.05-	73.15+	73.26	73.36	73.46
92	73.57	73.68	73.78	73.89	74.00	74.11	74.21	74.32	74.44	74.55-
93	74.66	74.77	74.88	75.00	75.11	75.23	75.35-	75.46	75.58	75.70
94	75.82	75.94	76.06	76.19	76.31	76.44	76.56	76.69	76.82	76.95-
95	77.08	77.21	77.34	77.48	77.61	77.75+	77.89	78.03	78.17	78.32
96	78.46	78.61	78.76	78.91	79.06	79.22	79.37	79.53	79.69	79.86
97	80.02	80.19	80.37	80.54	80.72	80.90	81.09	81.28	81.47	81.67
98	81.87	82.08	82.29	82.51	82.73	82.96	83.20	83.45+	83.71	83.98
99.0	84.26	84.29	84.32	84.35-	84.38	84.41	84.44	84.47	84.50	84.53
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
99.2	84.87	84.90	84.93	84.97	85.00	85.03	85.07	85.10	85.13	85.17
99.3	85.20	85.24	85.27	85.31	85.34	85.38	85.41	85.45-	85.48	85.52
99.4	85.56	85.60	85.63	85.67	85.71	85.75-	85.79	85.83	85.87	85.91
99.5	85.95-	85.99	86.03	86.07	86.11	86.15-	86.20	86.24	86.28	86.33
99.6	86.37	86.42	86.47	86.51	86.56	86.61	86.66	86.71	86.76	86.81
99.7	86.86	86.91	86.97	87.02	87.08	87.13	87.19	87.25+	87.31	87.37
99.8	87.44	87.50	87.57	87.64	87.71	87.78	87.86	87.93	88.01	88.10
99.9	88.19	88.28	88.38	88.48	88.60	88.72	88.85+	89.01	89.19	89.43
100.0	90.00									

ตาราง ผ.12 สัมประสิทธิ์โพลีโนเมียลสำหรับการเปรียบเทียบระหว่างค่าเฉลี่ยของทรีตเมนต์

Treatment (no.)	Degree of Polynomials	T	T	T	T	T	T	Sum of squares of the Cofficients
3	Linear	– 2	0	+1				2
	Quadratic	+1	–2	+1				6
4	Linear	–3	–1	+1	+3			20
	Quadratic	+1	–1	–1	+1			4
	Cubic	–1	+3	–3	+1	+2		20
5	Linear	–2	–1	0	+1	+2		10
	Quadratic	+2	–1	–2	–1	+1		14
	Cubic	–1	+2	0	–2	+1	+5	10
	Quartic	+1	–4	+6	–4	+3	+5	70
6	Linear	–5	–3	–1	+1	–1	+5	70
	Quadratic	+5	–1	–4	–4	–7	+1	84
	Cubic	–5	+7	+4	–4	–3	+1	180
	Quartic	+1	–3	+2	+2	–5		28
	Quintic	–1	+5	–10	+10			252

ตาราง ผ. 13 ค่า F_{max} ของฮาร์ทเลย์ (Hartley's F_{max})

df for σ_i^2	α	p = Number of Variances										
		2	3	4	5	6	7	8	9	10	11	12
2	.05	39.0	87.5	142	202	266	333	403	475	550	626	704
	.01	199	448	729	1036	1362	1705	2063	2432	2813	3204	3605
3	.05	15.4	27.8	39.2	50.7	62.0	72.9	83.5	93.9	104	114	124
	.01	47.5	85	120	151	184	216	249	281	310	337	361
4	.05	9.60	15.5	20.6	25.2	29.5	33.6	37.5	41.4	44.6	48.0	51.4
	.01	23.2	37.	49	59	69.	79	89	97.	106.	113.	120
5	.05	7.15	10.8	13.7	16.3	18.7	20.8	22.9	24.7	26.5	28.2	29.9
	.01	14.9	22	28	33	38.	42	46	50.	54.	57.	60.
6	.05	5.82	8.38	10.4	12.1	13.7	15.0	16.3	17.5	18.6	19.7	20.7
	.01	11.1	15.5	19.1	22	25.	27	30	32	34	36.	37.
7	.05	4.99	6.94	8.44	9.70	10.8	11.8	12.7	13.5	14.3	15.1	15.8
	.01	8.89	12.1	14.5	16.5	18.4	20	22	23.	24.	26.	27.
8	.05	4.43	6.00	7.18	8.12	9.03	9.78	10.5	11.1	11.7	12.2	12.7
	.01	7.50	9.9	11.7	13.2	14.5	15.8	16.9	17.9	18.9	19.8	21
9	.05	4.03	5.34	6.31	7.11	7.80	8.41	8.95	9.45	9.91	10.3	10.7
	.01	6.54	8.5	9.9	11.1	12.1	13.1	13.9	14.7	15.3	16.0	16.6
10	.05	3.72	4.85	5.67	6.34	6.92	7.42	7.87	8.28	8.66	9.01	9.34
	.01	5.85	7.4	8.6	9.6	10.4	11.1	11.8	12.4	12.9	13.4	13.9
12	.05	3.28	4.16	4.79	5.30	5.72	6.09	6.42	6.72	7.00	7.25	7.48
	.01	4.91	6.1	6.9	7.6	8.2	8.7	9.1	9.5	9.9	10.2	10.6
15	.05	2.86	3.54	4.01	4.37	4.68	4.95	5.19	5.40	5.59	5.77	5.93
	.01	4.07	4.9	5.5	6.0	6.4	6.7	7.1	7.3	7.5	7.8	8.0
20	.05	2.46	2.95	3.29	3.54	3.76	3.94	4.10	4.24	4.37	4.49	4.59
	.01	3.32	3.8	4.3	4.6	4.9	5.1	5.3	5.5	5.6	5.8	5.9
30	.05	2.07	2.40	2.61	2.78	2.91	3.02	3.12	3.21	3.29	3.36	3.39
	.01	2.63	3.0	3.3	3.4	3.6	3.7	3.8	3.9	4.0	4.1	4.2
60	.05	1.67	1.85	1.96	2.04	2.11	2.17	2.22	2.26	2.30	2.33	2.36
	.01	1.96	2.2	2.3	2.4	2.4	2.5	2.5	2.6	2.6	2.7	2.7
∞	.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

ตาราง ผ. 14 ค่า C ของ คอคคราน (Cochran' s C)

df for $\hat{\sigma}_j^2$	α	Number of Variances (p)										
		2	3	4	5	6	7	8	9	10	15	20
1	.05	.9985	.9669	.9065	.8412	.7808	.7271	.6798	.6385	.6020	.4709	.3894
	.01	.9999	.9933	.9676	.9279	.8828	.8376	.7945	.7544	.7175	.5747	.4799
2	.05	.9750	.8709	.7679	.6838	.6161	.5612	.5157	.4775	.4450	.3346	.2705
	.01	.9950	.9423	.8643	.7885	.7218	.6644	.6152	.5727	.5358	.4069	.3297
3	.05	.9392	.7977	.6841	.5981	.5321	.4800	.4377	.4027	.3733	.2758	.2205
	.01	.9794	.8831	.7814	.6957	.6258	.5685	.5209	.4810	.4469	.3317	.2654
4	.05	.9057	.7457	.6287	.5441	.4803	.4307	.3910	.3584	.3311	.2419	.1921
	.01	.9586	.8335	.7212	.6329	.5635	.5080	.4627	.4251	.3934	.2882	.2288
5	.05	.8772	.7071	.5895	.5065	.4447	.3974	.3595	.3286	.3029	.2195	.1735
	.01	.9373	.7933	.6761	.5875	.5195	.4659	.4226	.3870	.3572	.2593	.2048
6	.05	.8534	.6771	.5598	.4783	.4184	.3726	.3362	.3067	.2823	.2034	.1602
	.01	.9172	.7606	.6410	.5531	.4866	.4347	.3932	.3592	.3308	.2386	.1877
7	.05	.8332	.6530	.5365	.4564	.3980	.3535	.3185	.2901	.2666	.1911	.1501
	.01	.8988	.7335	.6129	.5259	.4608	.4105	.3704	.3378	.3106	.2228	.1748
8	.05	.8159	.6333	.5175	.4387	.3817	.3384	.3043	.2768	.2541	.1815	.1422
	.01	.8823	.7107	.5897	.5037	.4401	.3911	.3522	.3207	.2945	.2104	.1646
9	.05	.8010	.6167	.5017	.4241	.3682	.3259	.2926	.2659	.2439	.1736	.1357
	.01	.8674	.6912	.5702	.4854	.4229	.3751	.3373	.3067	.2813	.2002	.1567
16	.05	.7341	.5466	.4366	.3645	.3135	.2756	.2462	.2226	.2032	.1429	.1108
	.01	.7949	.6059	.4884	.4094	.3529	.3105	.2779	.2514	.2297	.1612	.1248
36	.05	.6602	.4748	.3720	.3066	.2612	.2278	.2022	.1820	.1655	.1144	.0879
	.01	.7067	.5153	.4057	.3351	.2858	.2494	.2214	.1992	.1811	.1251	.0960
144	.05	.5813	.4031	.3093	.2513	.2119	.1833	.1616	.1446	.1308	.0889	.0675
	.01	.6062	.4230	.3251	.2644	.2229	.1929	.1700	.1521	.1376	.0934	.0709

บรรณานุกรม

จรัญ จันทลักขณา. 2523. สถิติ – วิธีวิเคราะห์และวางแผนงานวิจัย. บริษัทสำนักพิมพ์ไทยวัฒนาพานิช จำกัด พระนคร.

ไพศาล เหล่าสุวรรณ. 2514. วิธีการวิจัย. คณะเกษตรศาสตร์ มหาวิทยาลัยขอนแก่น (โรเนียว).

สุรพล อุปดิสสกุล. 2526. สถิติ – การวางแผนการทดลอง. แอ๊สเสทการพิมพ์ กรุงเทพฯ 10900.

อนันตชัย เขื่อนธรรม. 2542. วิธีการทางสถิติและการวิเคราะห์ข้อมูล. มหาวิทยาลัยเกษตรศาสตร์.

Bancroff, T.A. 1968. Topics in Intermediate Statistical Methods, Vol. 1, The Iowa State Univ. Press.

Crochran, W.G. and G.M. Cox. 1985. Experimental Design. John Wiley and Sons, Inc., New York.

Federer, W.T. 1963. Experimental Design. Oxford and IBH Publishing Co., New Delhi.

Fisher, R.A. 1968. The Design of Experiments. 8 th ed. Hafner Publ. Co., New York.

Freund, J.E. 1967. Modern Elementary Statistics. Prentice – Hall, Inc., Englewood Cliffs, NJ.

Gomez, K.A. and A.A. Gomez. 1984. Statistical Procedures for Agricultural Research. 2 nd ed. John Wiley and Sons, Co., New York.

Goulden, Cyril H. 1987. Methods of Statistical Analysis. John Wiley and Sons, Tokyo.

Montgomerry, D.C. 1976. Design and Analysis of Experiment, John Wiley and Sons, Co., New York.

LeClerg, E.L., W.H. Leonard and A.G. Clark. 1962. Field Plot Technique. 2 nd ed. Burgess publ. Co., Minneapolis.

Hicks, C.R. 1993. Fundamental Concepts in Design of Experiments. 4 th. ed. Sanders Collage Publishing, New York.

Snedecor, G.W. and W.G. Cochran. 1967. Statistical Methods. The Iowa State University Press, Ames, Iowa.

Steel, R.G. and J.H. Torrie. 1960. Principles and Procedures of Statistics. McGraw – Hill Book Co., Inc. New York.