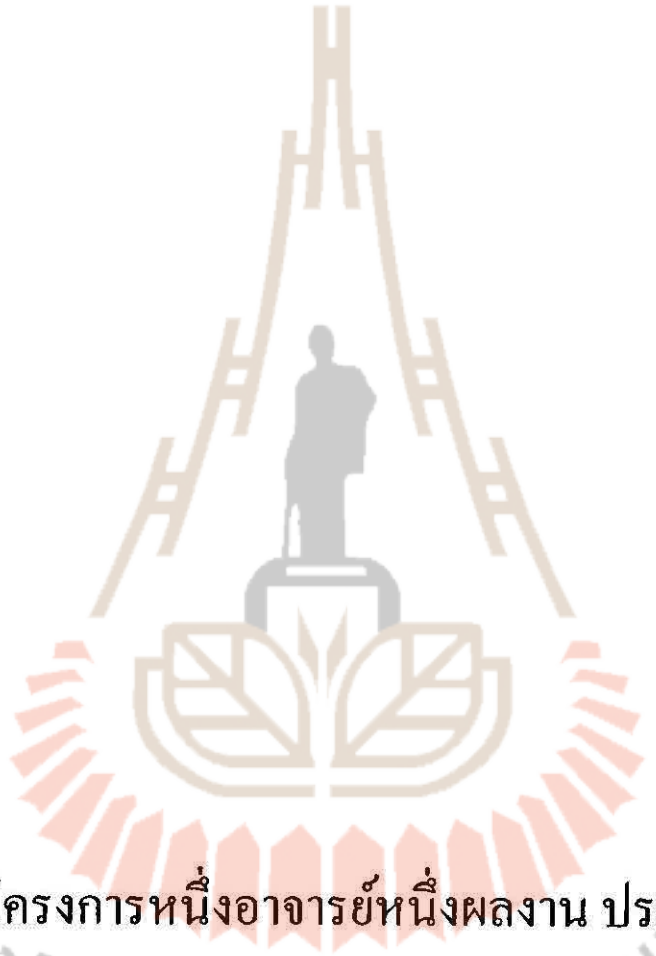
ประมวลสาระวิชาเทคโนโลยีสารสนเทศ 2



ผลงานในโครงการหนึ่งอาจารย์หนึ่งผลงาน ประจำปี 2546

สำนักวิชาเทคโนโลยีสังคม

โดย รองศาสตราจารย์ ดร.คณิต ไข่มุกด์

หน่วยที่

7

ความรู้พื้นฐานในการวิเคราะห์ข้อมูล

มหาวิทยาลัยเทคโนโลยีสุรนารี
ศูนย์บรรณสารและสื่อการศึกษา
มหาวิทยาลัยเทคโนโลยีสุรนารี

ผลงานในโครงการ หนึ่งอาจารย์ หนึ่งผลงาน ประจำปี 2546
สำนักวิชาเทคโนโลยีสังคม

โดย รองศาสตราจารย์ ดร.คณิต ไข่มุกด์

ตอนที่

7.1

ความหมายของข้อมูล และตัวแปร

ข้อมูลเป็นข้อเท็จจริงที่เก็บรวบรวมได้ โดยใช้ตัวแปรเป็นสัญลักษณ์แทนค่าของข้อมูล ระดับของข้อมูลจะเป็นตัวกำหนดวิธีการวิเคราะห์ข้อมูล

1. ข้อมูลคืออะไร

ข้อมูลเป็นข้อเท็จจริงที่สังเกตได้ ในชีวิตประจำวันของเรามีข้อมูลเข้ามาให้สังเกตตลอด ถ้ามนุษย์เราเป็นผู้สังเกต เราสามารถสังเกตได้จากจากประสาทสัมผัสทั้ง 6 คือ หู ตา กาย จมูก ลิ้น และใจ เริ่มจากตื่นขึ้นมาเพราะได้ยินเสียงสุนัขเห่า(หูเป็นเครื่องมือสังเกต) พอลืมตาขึ้นมาก็เห็นทีวีเปิดอยู่(ตาเป็นเครื่องมือสังเกต) เลยใช้มือคลำข้างๆ ตัวหารีโมท(กายเป็นเครื่องมือสังเกต, มือสัมผัส) เมื่อเจอแล้วก็กดปุ่มปิดทีวี ปรากฏว่าขณะเดียวกันก็ได้กลิ่นกาแฟ(จมูกเป็นเครื่องมือสังเกต) จึงลุกจากที่นอนไปตามกลิ่นกาแฟ เมื่อเห็นแล้วก็ลองจิบดู(ลิ้นเป็นเครื่องมือสังเกต) ผลปรากฏว่าขมมาก เพราะไม่ได้ใส่น้ำตาล จึงคิดไปว่าใครลุกขึ้นมาชงกาแฟแล้วลืมทิ้งไว้(ใจเป็นเครื่องมือสังเกต)

จะเห็นว่าข้อมูลที่สังเกตได้และพื้นฐานประสบการณ์จะเป็นตัวตัดสินและอธิบายเหตุการณ์ที่เกิดขึ้นดังนั้นการตัดสินใจที่เกิดขึ้นจะถูกหรือผิดก็ขึ้นอยู่กับข้อมูลที่สังเกตได้และประสบการณ์ของแต่ละคน

ถ้ามองลึกลงไปจะเห็นว่าข้อมูลที่สังเกตได้มีทั้งเป็นลักษณะร่องรอย รูป รส กลิ่น เสียง สิ่งที่สังเกตได้ อาจจะเป็น เสียง ภาพ วัตถุ กลิ่น และรส การบันทึกข้อมูลที่สังเกตได้ อาจจะเป็นข้อความบรรยาย ภาพ เทป ดิสค์ หรือ ตัวเลข ก็ได้ ทั้งหมดนี้เป็นข้อเท็จจริงที่สังเกตได้จากการใช้มนุษย์เป็นผู้สังเกตจากเครื่องมือการรับรู้ที่มีอยู่ในตัวเอง

ความจริงข้อมูลที่สังเกตได้ถ้าเรามีเครื่องมือที่ดีพอหรือเหมาะกับการสังเกตข้อเท็จจริงนั้นก็จะได้ข้อมูลจากการสังเกตมากกว่าเราสังเกตเอง เช่น ใช้สุนัขดมกลิ่นยาเสพติด หรือค้นหาบุคคล ใช้โทรศัพท์เพื่อรับสัญญาณเสียงจากที่ไกลๆ ใช้เครื่องรับโทรทัศน์เพื่อรับสัญญาณภาพและสัญญาณเสียงในอากาศที่คนรับไม่ได้ ใช้เครื่องรับวิทยุเพื่อรับสัญญาณเสียงที่คนรับไม่ได้ ใช้เครื่องวัดกัมมันตภาพรังสี วัดขนาดกัมมันตภาพรังสี แม้แต่ใช้เครื่องรับฟังคลื่นแม่เหล็กไฟฟ้าจากนอกโลก ใช้ดาวเทียมถ่ายภาพความเคลื่อนไหวทั่วโลกทุกวัน

ข้อเท็จจริงที่สังเกตได้เหล่านี้เป็นข้อมูลทั้งสิ้น ถ้าเราจำแนกข้อมูลโดยใช้เกณฑ์ที่สามารถแสดงเป็นตัวเลขได้กับแสดงเป็นตัวเลขไม่ได้ เราจะเรียกข้อมูลที่สามารถแสดงเป็นตัวเลขได้ว่าข้อมูลเชิงปริมาณ และข้อมูลที่ไม่สามารถแสดงเป็นตัวเลขได้ เรียกว่าข้อมูลเชิงคุณลักษณะหรือเชิงคุณภาพ

2. ตัวแปรคืออะไร

จะเห็นว่าข้อมูลนอกจากจะมีมากมายและหลากหลายแล้ว การนำมาจัดเป็นหมวดหมู่แล้วตั้งชื่อเป็นตัวแปรเป็นเรื่องๆ ไป ก็สามารถที่จะนำไปทำการวิเคราะห์ได้ง่ายและเป็นระเบียบ เช่น ข้อมูลปริมาณน้ำฝนแต่ละเดือนในรอบปีที่ผ่านมาที่จังหวัดนครราชสีมา ตั้งเป็นชื่อตัวแปรว่า RAINFALL ก็จะเข้าใจกันได้ง่ายขึ้น ว่าค่าที่สังเกตได้ของตัวแปรนี้คือปริมาณน้ำฝนแต่ละเดือน มีจำนวนข้อมูลเท่ากับ 12 ค่า ข้อมูลผลการสอบกลางภาคของนักศึกษาที่เรียนวิชา เทคโนโลยีสารสนเทศ 2 ตั้งเป็นชื่อตัวแปรว่า SCORE ก็จะเข้าใจว่าคะแนนที่แต่ละคนได้ของตัวแปรนี้คือคะแนนในการสอบกลางภาคที่ผ่านมา จะมีจำนวนข้อมูลเท่ากับจำนวนนักศึกษาที่เข้าสอบวิชานี้

ข้อมูลจากหัวข้อการประเมินการเรียนการสอนจากนักศึกษามีหลายประเด็นอาจจะจำแนกเป็นหลายตัวแปรตามแต่ละข้อคำถาม ข้อคำถามละ 1 ตัวแปร ได้แก่

X1 แทน การแจ้งเค้าโครงรายวิชาต้นภาคการศึกษา

X2 แทน ความถี่ในการแจ้งเนื้อหาก่อนสอนแต่ละครั้ง

X3 แทน ความถี่ในการประเมินความรู้ของนักศึกษาก่อนเรียน

X4 แทน การแนะนำเรื่องที่จะเรียนด้วยการบอกความจำเป็น ชื่อเรื่อง หัวข้อที่จะเรียน และวัตถุประสงค์ที่จะเรียนในแต่ละครั้ง

X5 แทน การนำเข้าสู่บทเรียนด้วยการยกประเด็น เรื่องราว รายกรณีที่เกี่ยวข้องกับเรื่องที่จะสอน

.....

ดังนั้น โดยสรุปตัวแปรก็คือชื่อของกลุ่มข้อมูลประเภทเดียวกัน ตั้งขึ้นเพื่อประโยชน์ในการอ้างถึงหรือนำไปวิเคราะห์ข้อมูล

3. หลักการแบ่งระดับของข้อมูล

ถ้าเราพิจารณาข้อเท็จจริงที่ที่สังเกตได้แล้วจะเห็นว่าข้อมูลที่ได้รับจะแตกต่างกัน เช่น ชื่อของนักศึกษา 30 คน ไม่ได้ให้ข้อมูลอะไรมากไปกว่าบอกความแตกต่างของแต่ละคนด้วยชื่อเท่านั้น แต่ถ้าเราให้นักศึกษาทั้ง 30 คนมายืนเรียงกันตามลำดับความสูง เราก็จะบอกลำดับที่ตามความสูงได้ แต่ถ้าเราให้นักศึกษาทั้ง 30 คน สอบเราก็จะได้คะแนนของแต่ละคนที่บอกความแตกต่างกันได้ว่า

แต่ละคนได้คะแนนแตกต่างกันเท่าใด หรือถ้าเราให้ทุกคนวัดความสูงเป็นเซ็นติเมตรแล้วบันทึกไว้ ค่าความสูงก็จะบอกความแตกต่างกันได้ว่าแต่ละคนสูงแตกต่างกันเท่าใด จะเห็นว่าข้อมูลจากตัวแปร ชื่อ ลำดับความสูง คะแนนสอบ และความสูง แสดงรายละเอียดหรือความแตกต่างไม่เท่ากัน

ดังนั้นเรายึดหลักความสามารถในการบอกความแตกต่างของข้อมูลในการแบ่งระดับของข้อมูล ดังนี้

ถ้าข้อมูลบอกความแตกต่างได้แต่เพียงชื่อ เช่นชื่อของนักศึกษาทั้ง 30 คน เราเรียกว่าข้อมูลระดับมาตรานามบัญญัติ(Nominal scale)

ถ้าข้อมูลบอกลำดับความแตกต่างได้ เช่นชื่อของนักศึกษาที่เรียงตามลำดับความสูงทั้ง 30 คน เราเรียกว่าข้อมูลระดับมาตราอันดับบัญญัติ(Ordinal scale)

ถ้าข้อมูลบอกความแตกต่างเป็นหน่วยที่เท่ากันได้ เช่น คะแนนสอบของนักศึกษาทั้ง 30 คน เราเรียกว่าข้อมูลระดับมาตราช่วงบัญญัติ(Interval scale)

ถ้าข้อมูลบอกความแตกต่างเป็นหน่วยที่เท่ากันได้และค่าศูนย์เป็นค่าศูนย์ที่แท้จริง (Absolute Zero) เช่น ความสูงเป็นเซ็นติเมตรของนักศึกษาทั้ง 30 คน เราเรียกว่าข้อมูลระดับมาตราอัตราส่วนบัญญัติ(Ratio scale)

สรุป

มหาวิทยาลัยเทคโนโลยีสุรนารี

กิจกรรมต่อเนื่อง ตอนที่ 7.1

1. ศึกษาผ่านสื่อปฏิสัมพันธ์ผ่านจอภาพ การบรรยาย หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.1 ความหมายของข้อมูล และตัวแปร
2. ศึกษาด้วยตนเองผ่านสื่อบทเรียนคอมพิวเตอร์ช่วยสอน หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.1 ความหมายของข้อมูล และตัวแปร
3. ทำกิจกรรมประกอบการเรียนในคู่มือการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.1 ความหมายของข้อมูล และตัวแปร
4. ทำกิจกรรมเสริมประสบการณ์ในแผนกิจกรรมการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.1 ความหมายของข้อมูล และตัวแปร

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตอนที่

7.2

ระดับมาตราของข้อมูล

การแบ่งระดับของข้อมูล จะทำให้สามารถเลือกวิธีวิเคราะห์ทางสถิติที่เหมาะสมได้ โดยแบ่งระดับการวัดข้อมูลออกเป็น 4 ระดับ ได้แก่ (1) ข้อมูลระดับมาตรานามบัญญัติ(Nominal scale) (2) ข้อมูลระดับมาตราอันดับบัญญัติ(Ordinal scale) (3) ข้อมูลระดับมาตราช่วงบัญญัติ(Interval scale) และ (4) ข้อมูลระดับมาตราอัตราส่วนบัญญัติ(Ratio scale)

1. ข้อมูลระดับมาตรานามบัญญัติ

ข้อมูลระดับมาตรานามบัญญัติ เป็นข้อมูลระดับที่หยาบที่สุดเพราะสามารถจำแนกความแตกต่างได้ด้วยชื่อเท่านั้น ตัวอย่างเช่น ตัวแปรการนับถือศาสนา ประกอบด้วยข้อมูล ชื่อศาสนา เช่น พุทธ คริสต์ อิสลาม และอื่นๆ ตัวแปรเพศ ประกอบด้วย ข้อมูล เพศชาย เพศหญิง ตัวแปรรหัสนักศึกษา ข้อมูล B4270911 B4270584 B4270508 และ อื่นๆ ตัวแปรเบอร์เสื้อทีมนักฟุตบอล ตัวเลขบนเสื้อแทนตำแหน่งในการเล่นฟุตบอล เลข 1 ถึง 50 บนเสื้อนักฟุตบอล ไม่ได้มีความหมายในการบอกความแตกต่างมากน้อยแสดงเพียงความแตกต่างโดยชื่อเท่านั้น ถึงแม้ เบอร์ 1 หมายถึงผู้รักษาประตู เบอร์ 7 หมายถึง เบอร์นำโชค เบอร์ 9 หมายถึง กองหน้า เบอร์ 10 หมายถึง เบอร์ทำเกมส์ เป็นต้น แต่เบอร์เสื้อไม่ได้บอกค่าความแตกต่าง เช่น เลข 1 น้อยกว่าเลข 2 หรือ เลข 1 ดีกว่า เลข 2 เป็นความหมายแต่เพียงว่านักฟุตบอลแต่ละคนมีหมายเลขเรียกแทนชื่อเท่านั้น

2. ข้อมูลระดับมาตราอันดับบัญญัติ

ข้อมูลระดับมาตราอันดับบัญญัติ เป็นข้อมูลระดับที่ละเอียดกว่าข้อมูลระดับมาตรานามบัญญัติ เพราะสามารถจำแนกความแตกต่างตามอันดับได้ เช่น การจัดอันดับความสวยของนางงามสามารถบอกความแตกต่างได้ว่าใครสวยกว่าใคร หรือ ให้นักศึกษามายืนเรียงกัน 30 คนตามลำดับความสูง สามารถจำแนกความแตกต่างตามอันดับความสูงได้ หรือจัดอันดับที่ให้นักศึกษาทั้ง 30 คน เรียงตามลำดับคะแนนเฉลี่ยสะสม ข้อมูลที่สามารถบอกลำดับที่ของข้อมูลแต่ไม่สามารถบอกได้ว่าแต่ละอันดับที่นั้นแตกต่างกัน นั้นแตกต่างกันเท่าใด เพียงแต่บอกลำดับที่แตกต่างได้เท่านั้น เราเรียกข้อมูลนี้ว่าข้อมูลระดับมาตราอันดับบัญญัติ

3. ข้อมูลระดับมาตราช่วงบัญญัติ

ข้อมูลระดับมาตราช่วงบัญญัติ เป็นข้อมูลระดับที่ละเอียดกว่าข้อมูลระดับมาตราอันดับบัญญัติ เพราะสามารถบอกขนาดความแตกต่างของแต่ละข้อมูลได้ เช่น ข้อมูลคะแนนสอบของนักศึกษา จำนวน 30 คน สมมติว่านาย ก ได้ 20 คะแนน และ นาย ข ได้ 22 คะแนน ก็หมายความว่านาย ข ได้คะแนนมากกว่านาย ก 2 คะแนน

ดังนั้นใน กรณี ที่เป็นคะแนนความคิดเห็น โดยระบุว่า

เห็นด้วยมากที่สุด ให้คะแนน 5
เห็นด้วยมาก ให้คะแนน 4
เห็นด้วยปานกลาง ให้คะแนน 3
เห็นด้วยน้อย ให้คะแนน 2
เห็นด้วยน้อยที่สุด ให้คะแนน 1

ถ้าตีความว่าตัวเลขที่กำกับ แสดงถึงหน่วยที่มีความแตกต่างเท่ากัน ก็จะถือว่าข้อมูลนี้ เป็นข้อมูลระดับมาตราช่วงบัญญัติ แต่ถ้าตัวเลขที่กำกับเพียงบอกอันดับความแตกต่างของความคิดเห็นเท่านั้น ก็จะถือว่าข้อมูลนี้ เป็นข้อมูลระดับมาตราอันดับบัญญัติ ซึ่งก็ยังเป็นข้อถกเถียงกันจนถึงปัจจุบัน ซึ่งระดับมาตรานำไปสู่ทางเลือกในการใช้สถิติเพื่อการวิเคราะห์ข้อมูล ดังนั้นเรื่องระดับมาตราเป็นเรื่องที่จำเป็นจะต้องเข้าใจอย่างดีจึงจะวิเคราะห์ข้อมูลได้ถูกต้อง

4. ข้อมูลระดับมาตราอัตราส่วนบัญญัติ

ข้อมูลระดับมาตราอัตราส่วนบัญญัติ เป็นข้อมูลระดับที่ละเอียดสูงสุด และแตกต่างกับข้อมูลระดับช่วงบัญญัติ ก็แต่เพียงว่า ค่าศูนย์เป็นค่าศูนย์ที่แท้จริง(absolute zero) เท่านั้น ตัวอย่างเช่น คะแนนสอบเป็นข้อมูลระดับมาตราช่วงบัญญัตินักศึกษาแต่ละคนได้คะแนนแตกต่างกันตามจำนวนคะแนนที่ได้ แต่คะแนน ศูนย์ ไม่ได้เป็นค่าศูนย์ที่แท้จริงเพราะว่า คะแนนศูนย์ คือทำข้อสอบผิดหมดทุกข้อเท่านั้นเอง ในความเป็นจริงเขาอาจจะมีความรู้อีกมากแต่ไม่ตรงกับข้อสอบก็ได้ ดังนั้นคะแนนสอบก็เป็นข้อมูลระดับมาตราช่วงบัญญัติเท่านั้นเพราะคะแนนศูนย์เป็นค่าศูนย์ที่สมมติขึ้น แต่ถ้าเปรียบเทียบกับ ความสูงเป็นเซ็นติเมตรของนักศึกษา เป็นข้อมูลระดับมาตราอัตราส่วนบัญญัติ เพราะว่า นักศึกษาที่มีความสูงเป็น ศูนย์ ก็คือ ผู้ที่ไม่มีความสูงจริงๆ ดังนั้น ค่าศูนย์ ของความสูงเป็น ค่าศูนย์แท้

สถิติที่ใช้ในการวิเคราะห์ข้อมูลที่จำแนกตามระดับของข้อมูล แบ่งออกเป็น 3 กลุ่ม คือ

กลุ่มที่ 1 สถิติที่ใช้ในการวิเคราะห์ข้อมูลระดับมาตรานามบัญญัติ

กลุ่มที่ 2 สถิติที่ใช้ในการวิเคราะห์ข้อมูลระดับมาตราอันดับบัญญัติ

กลุ่มที่ 3 สถิติที่ใช้ในการวิเคราะห์ข้อมูลระดับมาตราช่วงบัญญัติ และ อัตราส่วนบัญญัติ

ดังนั้นในโปรแกรม SPSS ก็แบ่งประเภทของข้อมูลออกเป็น 3 ระดับเช่นเดียวกัน ถ้าดูใน เมนู ก็คือ Nominal, Ordinal และ Scale

Nominal หมายถึง ข้อมูลระดับมาตรานามบัญญัติ

Ordinal หมายถึง ข้อมูลระดับมาตราอันดับบัญญัติ

Scale หมายถึง ข้อมูลระดับมาตราช่วงบัญญัติ และ อัตราส่วนบัญญัติ

ข้อมูลระดับมาตราที่สูงกว่าสามารถใช้ได้กับสถิติ ที่ใช้ได้กับข้อมูลระดับมาตราที่ต่ำกว่า พอจะสรุปสถิติวิเคราะห์จำแนกตามระดับข้อมูลได้เป็นตารางดังนี้

ข้อมูลระดับมาตรา	สถิติวิเคราะห์เชิงพรรณา	สถิติวิเคราะห์เชิงอนุมาน
นามบัญญัติ	การแจกแจงความถี่ ร้อยละ ฐานนิยม	χ^2 – *test* Binomial test McNemar test (2 popn) Cochran's Q test (3 popn)
อันดับบัญญัติ	มัธยฐาน ส่วนเบี่ยงเบนควอไทล์ Rank correlation	Kolmogorov-Smirnov test One-Sample Runs test (2 popn) - Sign test - Wilcoxon Signed Ranks test - Median test - Mann-Whitney U-test - Kolmogorov-Smirnov Sample test - Wald-Wolfowiz Runs test - Moses test of Extreme Reactions (3 popn) - Friedman Two-way ANOVA - Kendall Coefficient of Concordance - Median test - Kruskal-Wallis One-way ANOVA
ช่วงและอัตราส่วนบัญญัติ	ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน Correlation	Z-test t-test F-test

สรุป

กิจกรรมต่อเนื่อง ตอนที่ 7.2

1. ศึกษาผ่านสื่อปฏิสัมพันธ์ผ่านจอภาพ การบรรยาย หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.2 ระดับมาตราของข้อมูล
2. ศึกษาด้วยตนเองผ่านสื่อบทเรียนคอมพิวเตอร์ช่วยสอน หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.2 ระดับมาตราของข้อมูล
3. ทำกิจกรรมประกอบการเรียนในคู่มือการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.2 ระดับมาตราของข้อมูล
4. ทำกิจกรรมเสริมประสบการณ์ในแผนกิจกรรมการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.2 ระดับมาตราของข้อมูล

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตอนที่

7.3

การวิเคราะห์ข้อมูลทางสถิติ

การวิเคราะห์ข้อมูลทางสถิติสามารถแบ่งออกได้เป็น 2 ประเภท ได้แก่ (1) สถิติเชิงพรรณา เป็นสถิติที่เกี่ยวข้องกับการวิเคราะห์ และการบรรยายถึงลักษณะข้อมูลที่เก็บรวบรวมได้ และ(2) สถิติเชิงอนุมาน เป็นสถิติที่เกี่ยวข้องกับกลุ่มตัวอย่าง ที่ได้มาจากการสุ่มตัวอย่าง(random sample) แล้วนำผลที่ได้ไปอ้างอิงถึงประชากร(population)

1. สถิติเชิงพรรณา

สถิติเชิงพรรณา เป็นสถิติที่แสดงข้อสรุปของข้อมูล โดยไม่ได้นำเอาทฤษฎีความน่าจะเป็นเข้ามามีส่วนในการพยากรณ์ ข้อสรุปของประชากร จึงไม่มีการอ้างอิงถึงตัวพารามิเตอร์(parameter) สถิติเชิงพรรณา ประกอบด้วย เนื้อหา ใหญ่ๆ 3 เรื่อง คือ

ก. การจัดอันดับข้อมูล

ข. การวัดแนวโน้มเข้าสู่ส่วนกลาง

ค. การวัดการกระจาย

การจัดอันดับข้อมูล เริ่มจากการทำการแจกแจงข้อมูลเป็นกลุ่ม ได้แก่ ตารางแจกแจงความถี่ ร้อยละของความถี่ เพื่อจัดอันดับของตำแหน่งข้อมูล ประกอบด้วยตารางแจกแจงความถี่สะสม ร้อยละของความถี่สะสมหรือเปอร์เซ็นไตล์(percentile)

การวัดแนวโน้มเข้าสู่ส่วนกลาง ประกอบด้วยค่าที่เป็นตัวแทนของกลุ่มข้อมูล ได้แก่ ค่าฐานนิยม (mode) ค่ามัธยฐาน(mode) และค่าเฉลี่ย(average หรือ mean)

การวัดการกระจาย เป็นค่าที่บอกความแตกต่างของข้อมูล ว่าห่างออกไปจากค่าที่เป็นตัวแทนของกลุ่มเพียงใด ได้แก่ พิสัย(range) ส่วนเบี่ยงเบนควอไตล์(quatile deviation) ส่วนเบี่ยงเบนเฉลี่ย (average deviation) และส่วนเบี่ยงเบนมาตรฐาน(standard deviation)

สถิติเชิงพรรณาที่จำเป็นและมีที่ใช้กันมาก ก็คือ เปอร์เซ็นต์ความถี่สะสมหรือ เปอร์เซ็นไตล์ ค่าเฉลี่ย และค่าส่วนเบี่ยงเบนมาตรฐาน สำหรับค่าฐานนิยมมีที่ใช้บ้างกรณีข้อมูลระดับมาตรานามบัญญัติ ค่ามัธยฐานมีที่ใช้ในกรณีข้อมูลระดับมาตราอันดับบัญญัติ และกรณีข้อมูลระดับมาตราช่วงบัญญัติหรืออัตราส่วนบัญญัติที่ค่าของข้อมูลสุดโต่ง ไม่ว่าจะเป็นค่าต่ำสุดหรือสูงสุด เพราะถ้าใช้ค่าเฉลี่ยจะทำให้ไม่ได้ค่ากลางที่เป็นตัวแทนข้อมูลที่แท้จริง

ตัวอย่าง จงคำนวณค่าสถิติเชิงพรรณา ของ คะแนนสอบจากนักศึกษา จำนวน 30 คนซึ่งคะแนนผลการสอบเป็นดังนี้ 1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 7, 7, 7, 7, 8, 8, 8, 9, 9, 10

วิธีทำ ค่าเฉลี่ยเลขคณิต $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{165}{30} = 5.5$

ค่าส่วนเบี่ยงเบนมาตรฐาน $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x)^2}{n(n-1)}} = 2.2552$

ค่าฐานนิยม คือค่าคะแนนที่มีความถี่สูงสุด ได้แก่ 5 และ 6

ค่าความเบ้ (Skewness) $= \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{(n-1)s^3}$

$= .000$

ค่าความโด่ง (Kurtosis) $= \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{(n-1)s^4} = 2.2605$

ค่าเปอร์เซ็นไตล์(Percentiles) คือค่าของคะแนนที่ตรงกับอันดับของตำแหน่งที่กำหนด เช่น เปอร์เซ็นไตล์ที่ 10 คือ ค่าของคะแนนที่ตรงกับ ตำแหน่งที่ 10 ใน ร้อย นั่นคือเมื่อเรียงลำดับคะแนนจากน้อยไปมาก ค่าคะแนนนี้จะมีค่ามากกว่าคนอื่นอยู่ 10 เปอร์เซ็นต์ ปกติตำแหน่งของเปอร์เซ็นไตล์จะเขียนเป็นตัวห้อยกับค่า P เช่น เปอร์เซ็นไตล์ที่ 10 เขียนเป็น P_{10}

จากสูตรคำนวณ ค่าเปอร์เซ็นไตล์ จากคะแนนดิบ ให้เรียงคะแนนจากมากไปน้อยและคำนวณตำแหน่งเปอร์เซ็นไตล์ ที่ k ว่าตรงกับคะแนนใด คะแนนนั้นก็จะเป็นค่าของ p_k

ค่าที่ตรงกับเปอร์เซ็นไตล์ที่ k คือค่าคะแนนที่ตรงกับลำดับ คะแนนที่ $(n+1)\frac{k}{100}$

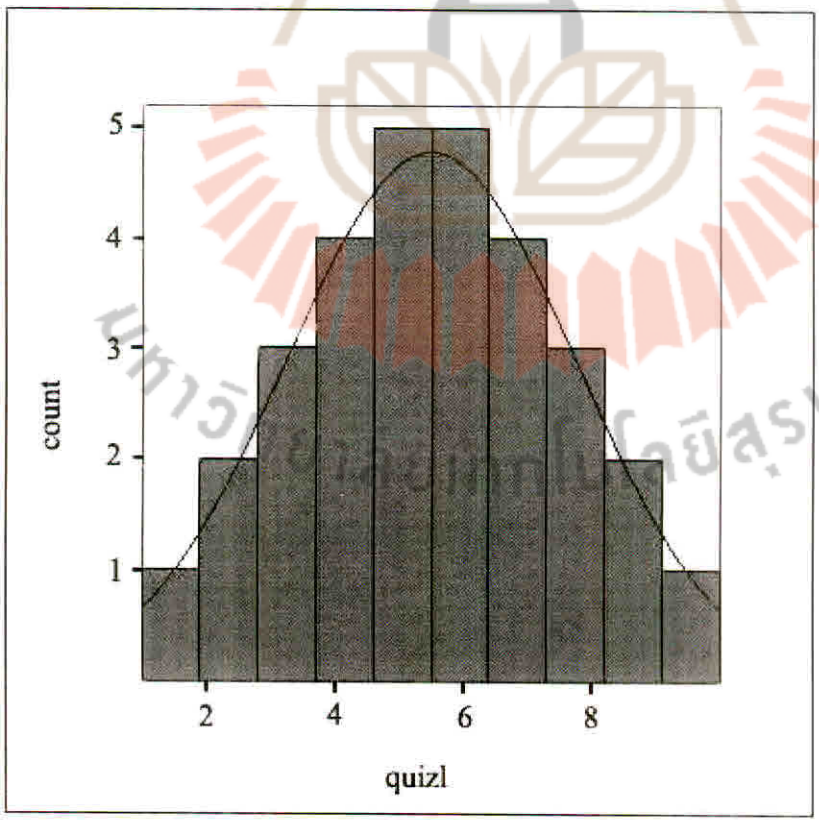
ค่าที่ตรงกับเปอร์เซ็นไตล์ที่ 10 คือค่าคะแนนที่ตรงกับลำดับ คะแนนที่ $(30+1)\frac{10}{100}$

นั่นคือ คะแนนลำดับที่ 3.1 ตรงกับคะแนน 2.1 ฉะนั้น P_{10} =2.1 ทำนองเดียวกัน $P_{20} = 3.2$ $P_{30} = 4$ $P_{40} = 5$ $P_{50} = 5.5$ $P_{60} = 6$ $P_{70} = 7$ $P_{80} = 7.8$ $P_{90} = 8.9$

เมื่อนำข้อมูลมาทำเป็นตารางแจกแจงความถี่จะได้ ดังตารางต่อไปนี้

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	1	3.3	3.3	3.3
	2	2	6.7	6.7	10.0
	3	3	10.0	10.0	20.0
	4	4	13.3	13.3	33.3
	5	5	16.7	16.7	50.0
	6	5	16.7	16.7	66.7
	7	4	13.3	13.3	80.0
	8	3	10.0	10.0	90.0
	9	2	6.7	6.7	96.7
	10	1	3.3	3.3	100.0
	Total	30	100.0	100.0	

เมื่อนำมา แสดงเป็นฮิสโตรแกรม เทียบกับโค้งปกติมาตรฐาน จะได้ดังรูป



ศูนย์บรรณสารและสื่อการศึกษา
มหาวิทยาลัยเทคโนโลยีสุรนารี

2. สถิติเชิงอนุมาน

สถิติเชิงอนุมาน มีจุดเน้นที่จะศึกษาถึงค่าพารามิเตอร์โดยอาศัยทฤษฎีความน่าจะเป็น โดยทั่วไปเราจะแบ่งเนื้อหาและวิธีการทดสอบ ตามจำนวนตัวแปรที่นำมาทดสอบหรือนำมาวิเคราะห์

ก. การทดสอบตัวแปรเดียว ที่สำคัญและใช้กันบ่อย ได้แก่

ก.1 การทดสอบค่าเฉลี่ย โดย z-test, t-test

ก.2 การทดสอบค่าสัดส่วน โดย t-test

ก.3 การทดสอบการแจกแจงแบบปกติ โดย χ^2 -test

ข. การทดสอบสองตัวแปร ที่สำคัญและใช้กันบ่อย ได้แก่

ข.1 การทดสอบความแตกต่างของค่าเฉลี่ย โดย z-test, t-test

ข.2 การทดสอบความแตกต่างของค่าสัดส่วน โดย t-test

ข.3 การทดสอบความเป็นอิสระ(independent test) โดย χ^2 -test

ค. การทดสอบตั้งแต่สามตัวแปรขึ้นไป ที่สำคัญและใช้กันบ่อย ได้แก่

การทดสอบความแตกต่างของค่าเฉลี่ย โดย การวิเคราะห์ความแปรปรวน หรือ F-test

ตัวอย่าง จากข้อมูลเดิมนำมาทดสอบว่าข้อมูลที่ได้เป็นโค้งปกติหรือไม่ ปรากฏว่าไม่สามารถปฏิเสธสมมติฐานว่างที่ว่าการแจกแจงของข้อมูลเป็นโค้งปกติไม่ว่าจะเป็นการทดสอบด้วย Kolmogorov-Smirnov หรือ Shapiro-Wilk

Tests of Normality

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
	Statistic	df	Sig.	Statistic	df	Sig.
quiz1	.088	30	.200	.977	30	.756

* This is a lower bound of the true significance.

a Lilliefors Significance Correction

ดังนั้นการที่เรามองว่าข้อมูลสมมาตรกันอย่างเดียวไม่เพียงพอต่อการสรุปว่าเป็นโค้งปกติ เพราะจุดสำคัญอยู่ที่ ความโด่งด้วย โค้งปกติจะมี ความเบ้เท่ากับ 0 และความโด่งเท่ากับ 3

หรือสามารถทดสอบง่ายๆ ได้โดย หาอัตราส่วนของความเบ้กับค่าคลาดเคลื่อนมาตรฐาน ถ้าอัตราส่วนอยู่ระหว่าง ± 2 ก็ประมาณได้ว่า การแจกแจงนี้เป็นการแจกแจงแบบปกติ

สรุป

กิจกรรมต่อเนื่อง ตอนที่ 7.3

1. ศึกษาผ่านสื่อปฏิสัมพันธ์ผ่านจอภาพ การบรรยาย หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.3 การวิเคราะห์ข้อมูลทางสถิติ
2. ศึกษาด้วยตนเองผ่านสื่อบทเรียนคอมพิวเตอร์ช่วยสอน หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.3 การวิเคราะห์ข้อมูลทางสถิติ
3. ทำกิจกรรมประกอบการเรียนในคู่มือการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.3 การวิเคราะห์ข้อมูลทางสถิติ
4. ทำกิจกรรมเสริมประสบการณ์ในแผนกิจกรรมการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.3 การวิเคราะห์ข้อมูลทางสถิติ

มหาวิทยาลัยเทคโนโลยีสุรนารี

หน่วยที่

8

การวิเคราะห์ข้อมูลทางสถิติด้วย โปรแกรมสำเร็จรูป SPSS

มหาวิทยาลัยเทคโนโลยีสุรนารี

โดย รองศาสตราจารย์ ดร.คณิต ไข่มุกด์

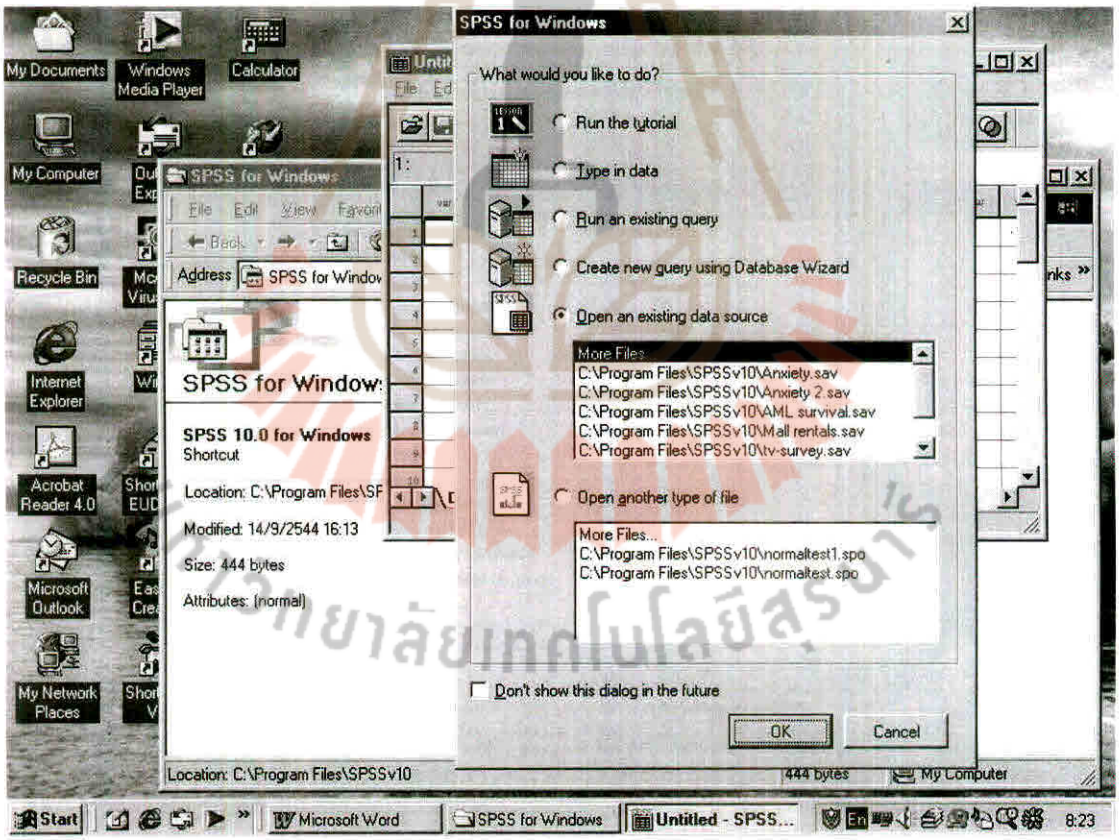
ตอนที่

8.1

หน้าต่างของโปรแกรมสำเร็จรูป SPSS

หน้าต่างของโปรแกรมสำเร็จรูป SPSS จะประกอบด้วยเมนูและคำสั่งในการวิเคราะห์ข้อมูลทางสถิติ ในเบื้องต้นผู้ใช้จะต้องมีความเข้าใจคำสั่งที่เกี่ยวกับตัวแปร ได้แก่ การกำหนดชื่อตัวแปร กำหนดชนิดของตัวแปร กำหนดชื่อของข้อมูล(Data Label) และกำหนดค่า สูญหาย (Missing)

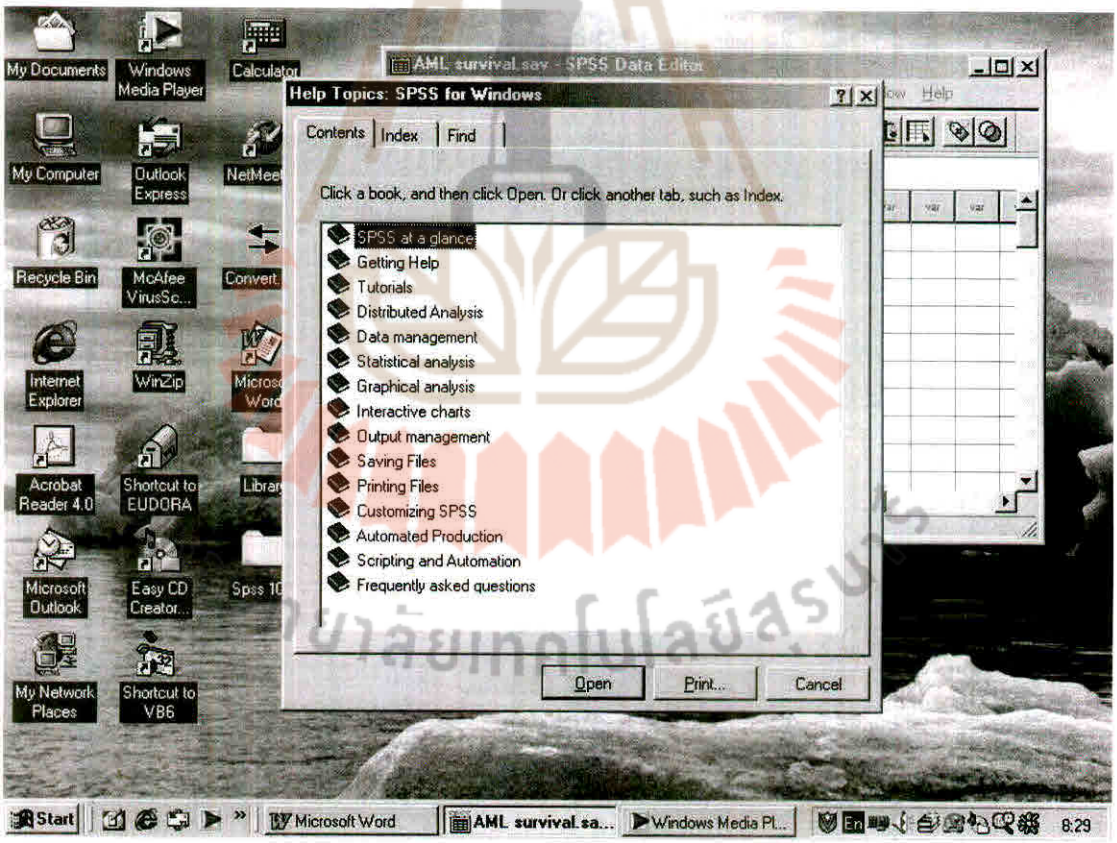
1. เมนูและคำสั่งในเมนู



เมนูแรกที่เปิดได้ มีสาระสำคัญอยู่ 6 หัวข้อ ได้แก่

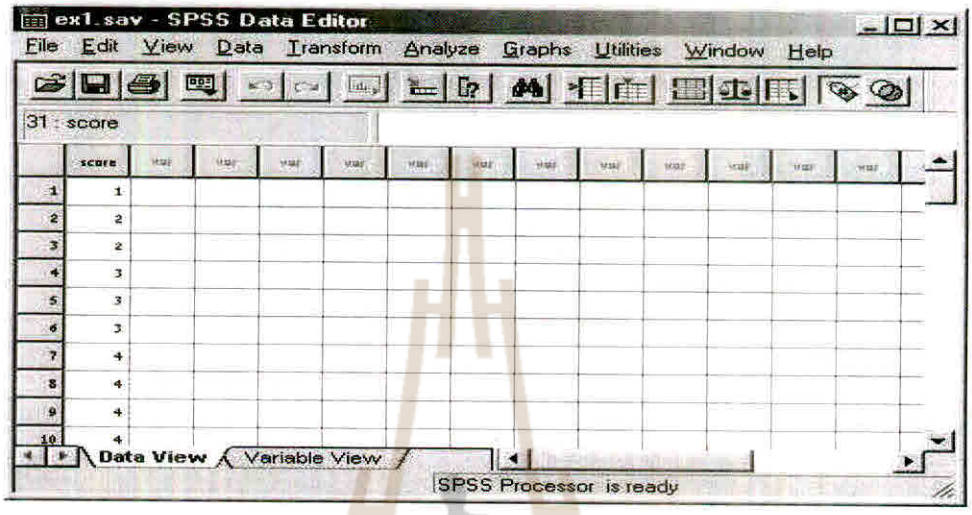
1. Run the tutorial เป็นการสอนวิธีใช้โปรแกรม SPSS เป็น เมนูแรกที่ทุกคนควรจะได้อ่านก่อนที่จะใช้ โปรแกรม SPSS
2. Type in data เป็นการเข้าสู่ตารางเพื่อป้อนข้อมูล
3. Run an existing query เป็นการดึงข้อมูลจากฐานข้อมูลที่ได้กำหนดไว้แล้วเพื่อนำมาทำการวิเคราะห์
4. Create new query using Database Wizard เป็นการดึงข้อมูลจากฐานข้อมูลขึ้นมาใหม่เพื่อนำมาทำการวิเคราะห์
5. Open an existing data source เปิดแฟ้มข้อมูลที่มีอยู่แล้วมาทำการวิเคราะห์
6. Open another type of file เปิดแฟ้มอื่นๆ อาจจะเป็นแฟ้มผลลัพธ์ที่มีอยู่แล้วมาตรวจสอบ

ในขั้นต้นขอให้เลือก Run the tutorial เพื่อเพื่อเรียนรู้การใช้โปรแกรมเสียก่อน ดังจอภาพ

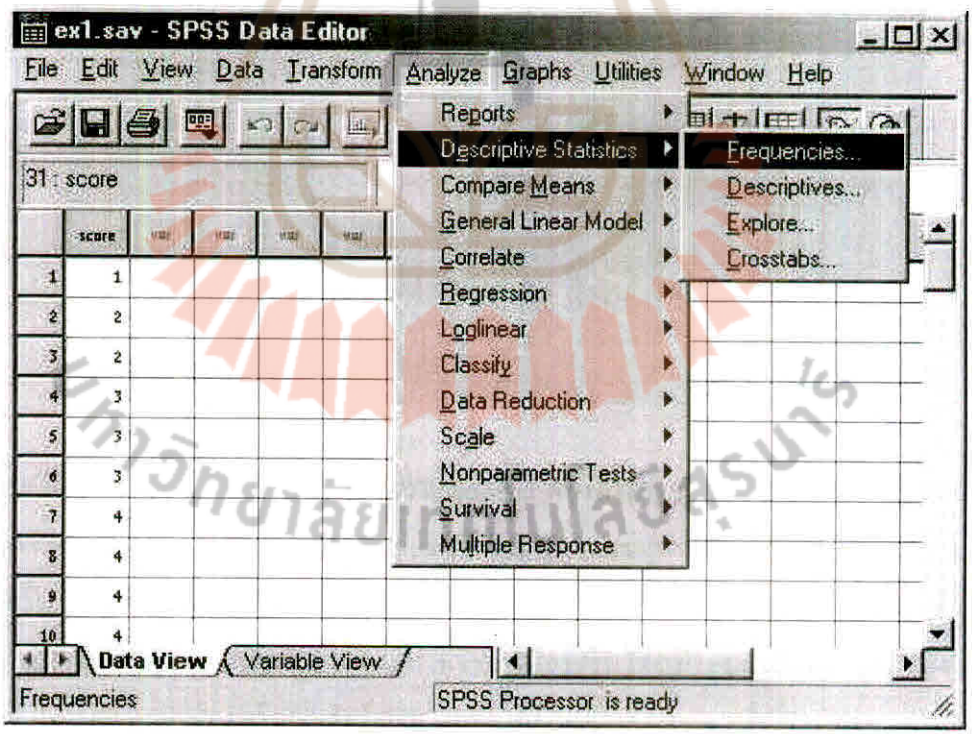


เมื่ออ่านจนเข้าใจ concept เบื้องต้นของโปรแกรม SPSS แล้วจึงจะ close เพื่อเข้าไปทดลองใช้ใบงาน(worksheet) เพื่อทดลองป้อนข้อมูล

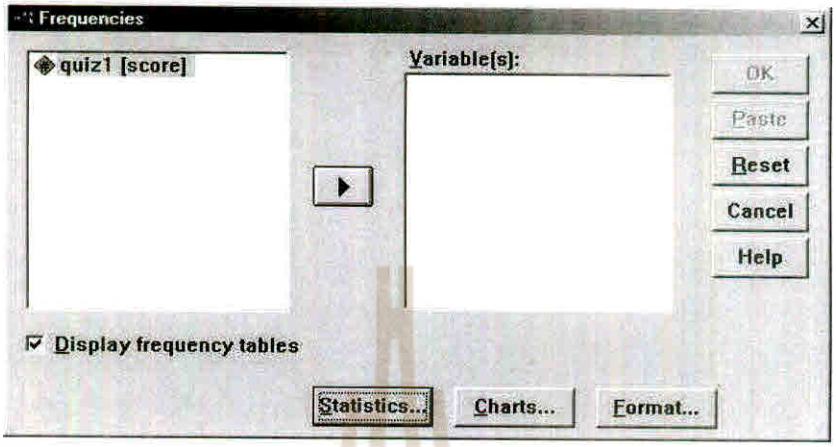
ทดลองป้อนข้อมูล ดังนี้ 1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 7, 7, 7, 7, 8, 8, 8, 9, 9, 10 ลงไปในคอลัมน์ที่ 1 แล้วเปลี่ยนชื่อ ตัวแปรเป็น score แล้ว Save เป็น แฟ้มชื่อ ex1 โปรแกรมจะกำหนดนามสกุลเป็น .sav ซึ่งเป็นนามสกุลของ แฟ้มข้อมูลโปรแกรม SPSS



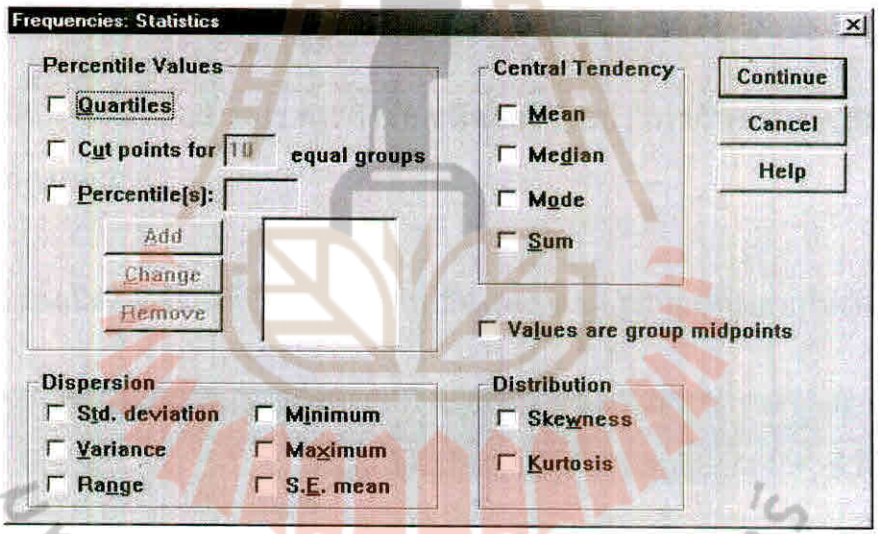
จากนั้นลองใช้คำสั่ง analyze ดู จะเห็นมีฟังก์ชันที่ใช้ได้ตาม pulldown menu ดังนี้



ลองไล่ดูฟังก์ชันทางสถิติ แล้วกลับมาที่ Descriptive statistics และ Frequencies จะเห็นดังนี้

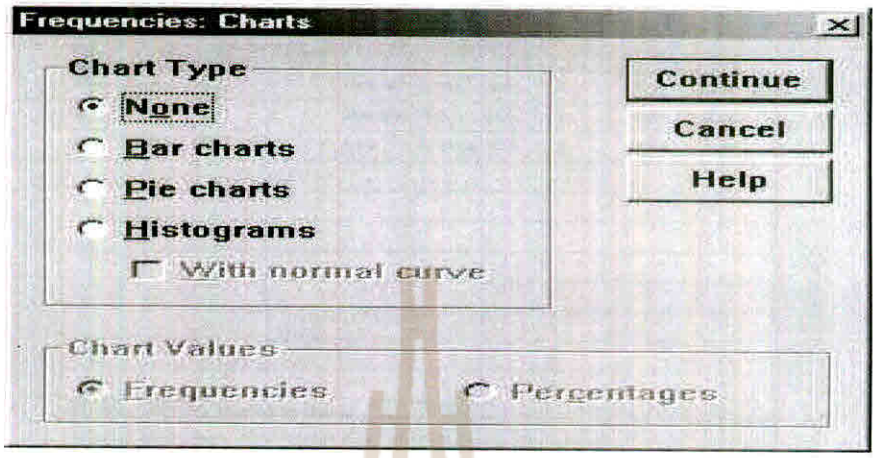


ดึงตัวแปรจากกล่องทางซ้ายมือเข้ามาที่กล่องขวามือเพื่อวิเคราะห์ แล้วเลือก Statistics ดู ปรากฏหน้าจอดังนี้



มีตัวสถิติที่สำคัญให้เลือกคำนวณตามความต้องการ ต้องการค่าสถิติอะไรก็ให้ทำเครื่องหมายในช่องหน้าตัวสถิตินั้นๆ ส่วนการหาเปอร์เซ็นไต์ลสามารถระบุค่าเปอร์เซ็นไตล์ที่ต้องการทราบโดยทำเครื่องหมายที่ช่องหน้าเปอร์เซ็นไตล์ ระบุตัวเลขที่ช่องถัดมา แล้วกดปุ่ม Add เข้าไปในกล่องใหญ่ทางขวามือเพื่อยืนยันว่าต้องการหาค่าของเปอร์เซ็นไตล์ที่ระบุแน่นอน ก่อนสั่งทำงานต่อไป ถ้ามีการเปลี่ยนแปลงหรือยกเลิกก็สามารถทำได้ด้วย ปุ่ม Change หรือ Remove

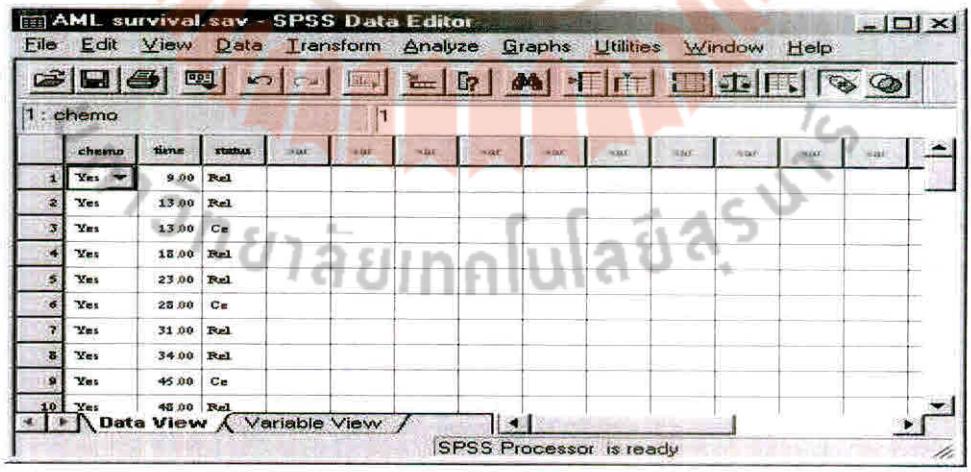
ลองเลือก Charts ดู ปรากฏหน้าจอดังนี้



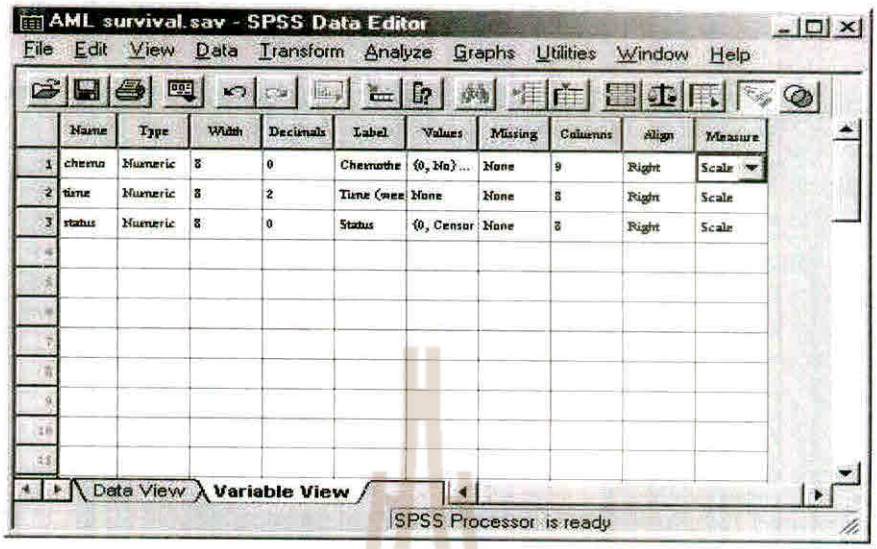
ท่านสามารถเลือกแสดง แผนภูมิ ตามทางเลือกที่ระบุไว้ในจอ ได้แก่ กราฟแท่ง กราฟวงกลม และกราฟแผนภูมิ ที่มีโค้งปกติให้ด้วย และตัวเลขที่แสดงกำกับกราฟ จะเป็นจำนวนความถี่หรือร้อยละก็ได้

2. การดำเนินการเกี่ยวกับตัวแปร การตั้งชื่อ ชนิดของตัวแปร การกำหนด LABEL และกำหนดค่าที่ MISSING

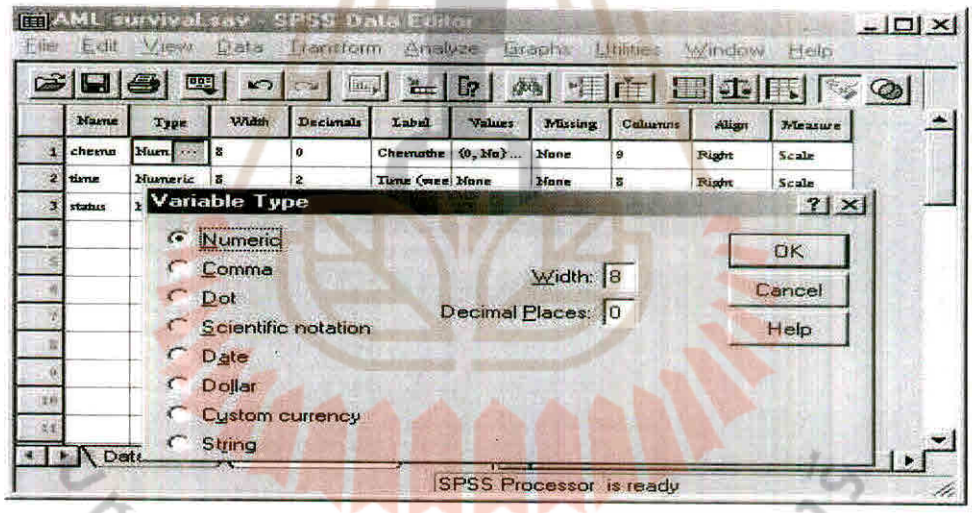
ไปที่ file ลองเปิดแฟ้มข้อมูลที่มีอยู่แล้วดู เลือก แฟ้ม survival.sav จะเห็นข้อมูลตามใบงานข้างล่าง



ลองเลื่อน cursor ไปที่ Variable View จะเห็นข้อมูลในอีกรูปแบบหนึ่งดังภาพ

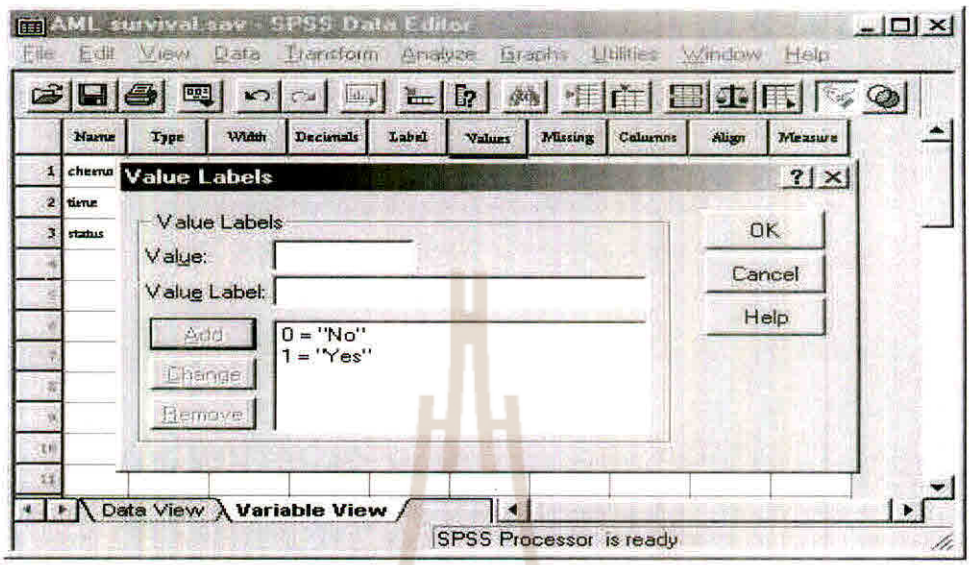


Name หมายถึงชื่อตัวแปร Label หมายถึง ชื่อของข้อมูล Type หมายถึง ชนิดและรูปแบบของ ของค่าข้อมูลของตัวแปรนั้นๆ มีรายละเอียดดังนี้

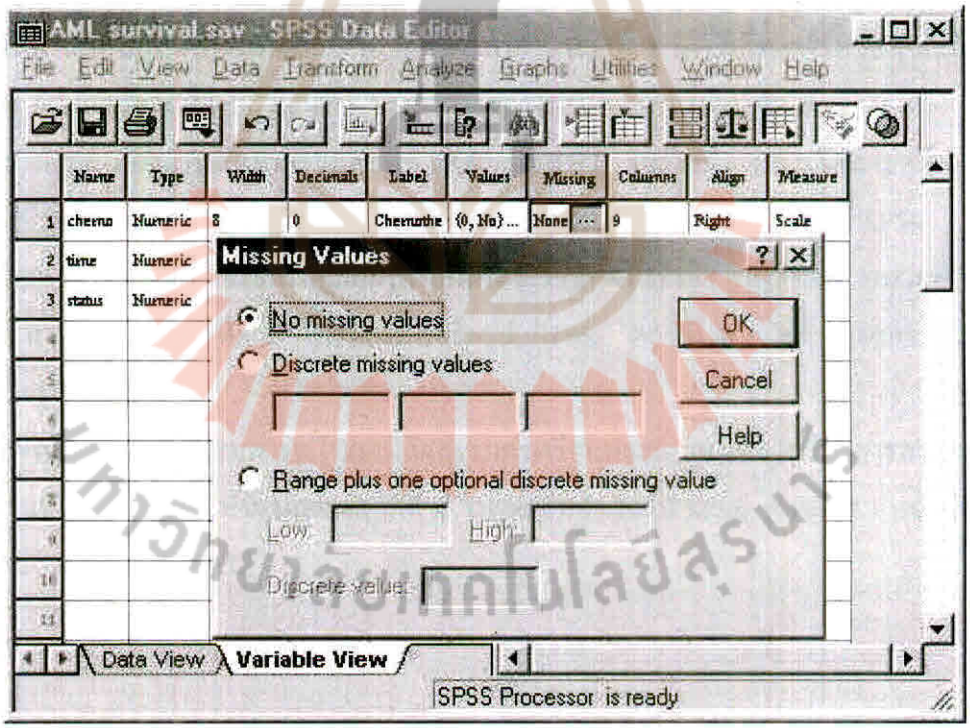


มหาวิทยาลัยเทคโนโลยีสุรนารี

และเราสามารถกำหนดค่าของ Label ให้เป็นชื่อใดๆ ที่สื่อความหมายก็ได้ ดังภาพ



และสามารถกำหนดค่าสูญหายหรือไม่ก็ได้ ดังแสดงดังภาพ



ค่าสูญหายกำหนดได้ทั้งกราฟเป็นแบบไม่ต่อเนื่องและต่อเนื่องก็ได้

ถึงจุดนี้เข้าใจว่านักศึกษาคงจะสามารถใช้คำสั่ง double click ที่ตำแหน่งที่ต้องการแก้ไขหรือกำหนดได้แล้ว ไม่ว่าจะเป็นการกำหนดชื่อตัวแปร ชนิดของตัวแปร กำหนดชื่อของข้อมูลและกำหนดค่าที่สูญหาย

สรุป

กิจกรรมต่อเนื่อง ตอนที่ 8.1

1. ศึกษาผ่านสื่อปฏิสัมพันธ์ผ่านจอภาพ การบรรยาย หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วยโปรแกรมสำเร็จรูป SPSS ตอนที่ 8.1 หน้าต่างของโปรแกรมสำเร็จรูป SPSS
2. ศึกษาด้วยตนเองผ่านสื่อบทเรียนคอมพิวเตอร์ช่วยสอน หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วย โปรแกรมสำเร็จรูป SPSS ตอนที่ 8.1 หน้าต่างของโปรแกรมสำเร็จรูป SPSS
3. ทำกิจกรรมประกอบการเรียนในคู่มือการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วย โปรแกรมสำเร็จรูป SPSS ตอนที่ 8.1 หน้าต่างของโปรแกรมสำเร็จรูป SPSS
4. ทำกิจกรรมเสริมประสบการณ์ในแผนกิจกรรมการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วย โปรแกรมสำเร็จรูป SPSS ตอนที่ 8.1 หน้าต่างของโปรแกรมสำเร็จรูป SPSS

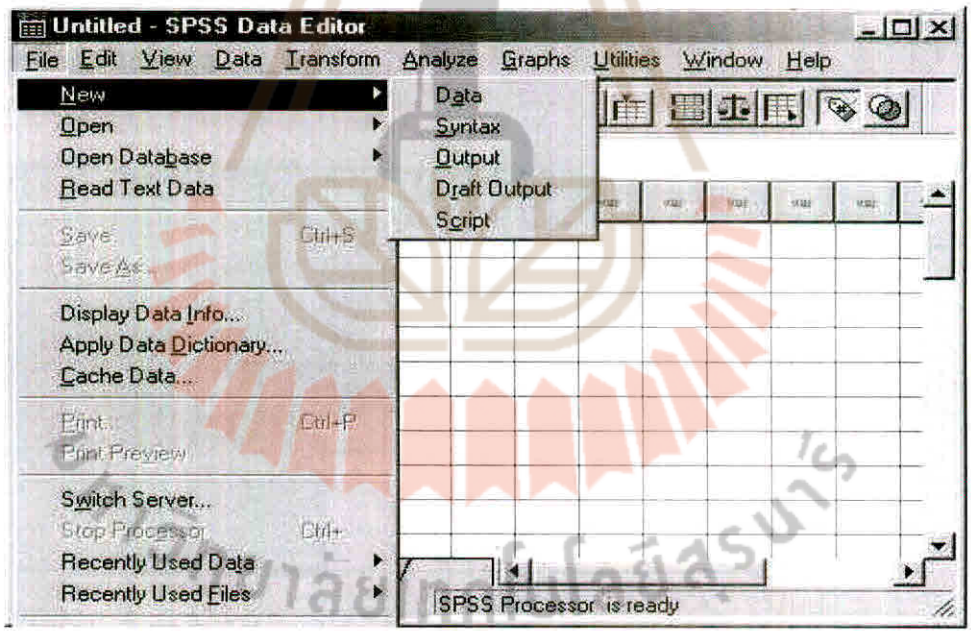
ตอนที่

8.2

การดำเนินการเกี่ยวกับแฟ้มข้อมูล

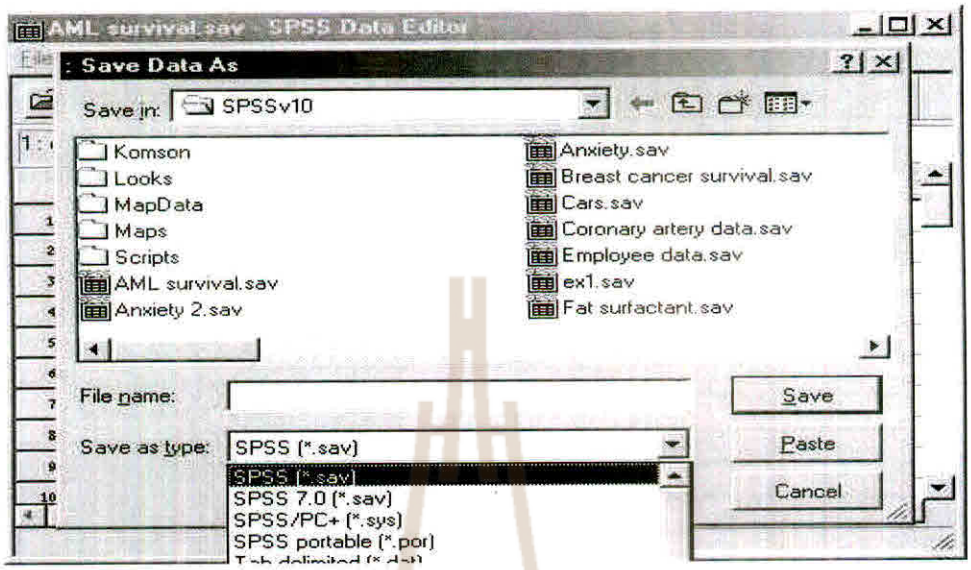
คำสั่งที่เกี่ยวข้องกับแฟ้มข้อมูลที่จะนำมาใช้ในการวิเคราะห์ด้วยโปรแกรมสำเร็จรูป SPSS ประกอบด้วยคำสั่ง (1) การสร้างแฟ้มข้อมูล (2) การบันทึกแฟ้มข้อมูล (3) การแก้ไขแฟ้มข้อมูล และ (4) การปิดแฟ้มข้อมูล

1. การสร้างแฟ้มข้อมูล



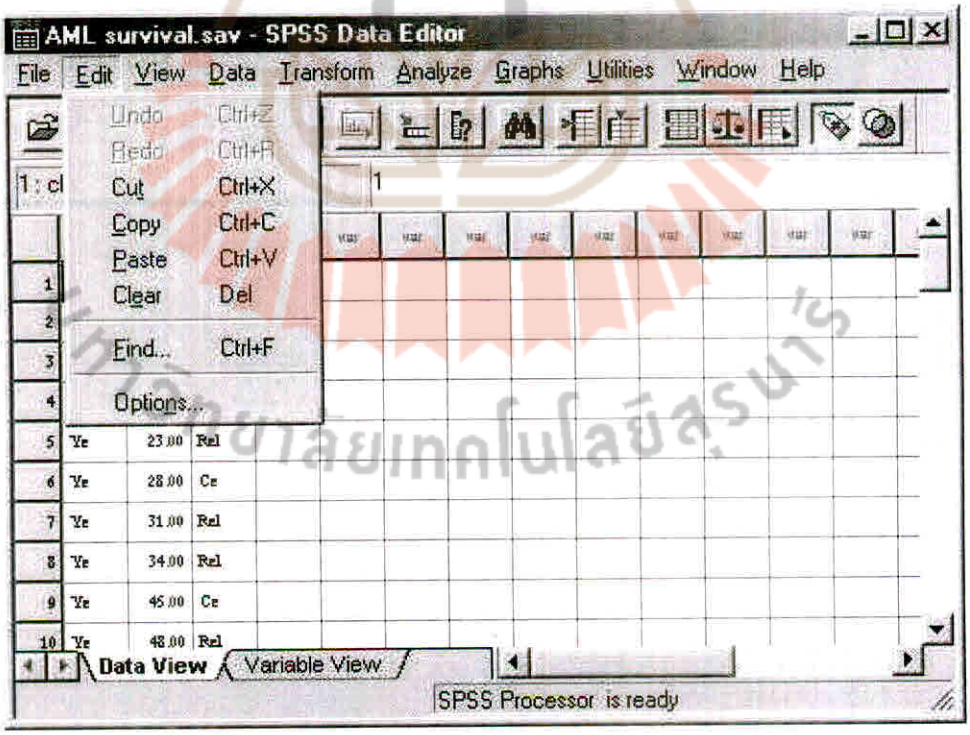
เข้าไปที่ File ⟶ New ⟶ Data แล้วทำการป้อนข้อมูล และเปลี่ยนชื่อตัวแปรและตั้งชื่อข้อมูล กำหนดชนิดและค่าสูญหาย ตามข้อมูลที่มี

2. การบันทึกแฟ้มข้อมูล



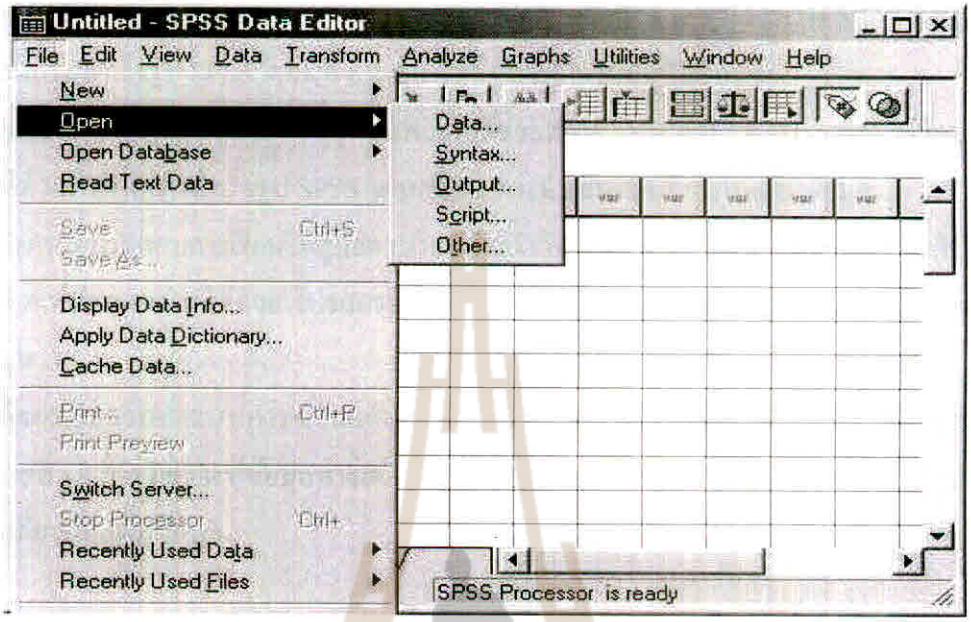
เข้าไปที่ File ⟶ Save As ระบุชื่อแฟ้มข้อมูลที่ต้องการบันทึก

3. การแก้ไขแฟ้มข้อมูล



เข้าไปที่ Edit ⟶ จัดการแก้ไข ตัด ปะ ลบ ได้เหมือนกับโปรแกรมตระกูลไมโครซอฟท์ทั่วไป

4. การเปิดแฟ้มข้อมูล



เข้าไป File ⟶ Open ⟶ Data ระบุชื่อแฟ้มข้อมูล หรือ เลือกชื่อแฟ้มข้อมูลที่มีมาเปิดใช้งาน

ให้ทดลองสร้างแฟ้มข้อมูลดู แล้วบันทึกไว้ ลองเรียกมาแก้ไขแล้วบันทึกไว้แล้วเปิดดูใหม่จนสามารถทำได้คล่อง

สรุป

มหาวิทยาลัยเทคโนโลยีสุรนารี

กิจกรรมต่อเนื่อง ตอนที่ 8.2

1. ศึกษาผ่านสื่อปฏิสัมพันธ์ผ่านจอภาพ การบรรยาย หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วยโปรแกรมสำเร็จรูป SPSS ตอนที่ 8.2 การดำเนินการเกี่ยวกับแฟ้มข้อมูล
2. ศึกษาด้วยตนเองผ่านสื่อบทเรียนคอมพิวเตอร์ช่วยสอน หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วย โปรแกรมสำเร็จรูป SPSS ตอนที่ 8.2 การดำเนินการเกี่ยวกับแฟ้มข้อมูล
3. ทำกิจกรรมประกอบการเรียนในคู่มือการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วย โปรแกรมสำเร็จรูป SPSS ตอนที่ 8.2 การดำเนินการเกี่ยวกับแฟ้มข้อมูล
4. ทำกิจกรรมเสริมประสบการณ์ในแผนกิจกรรมการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วย โปรแกรมสำเร็จรูป SPSS ตอนที่ 8.2 การดำเนินการเกี่ยวกับแฟ้มข้อมูล

มหาวิทยาลัยเทคโนโลยีสุรนารี

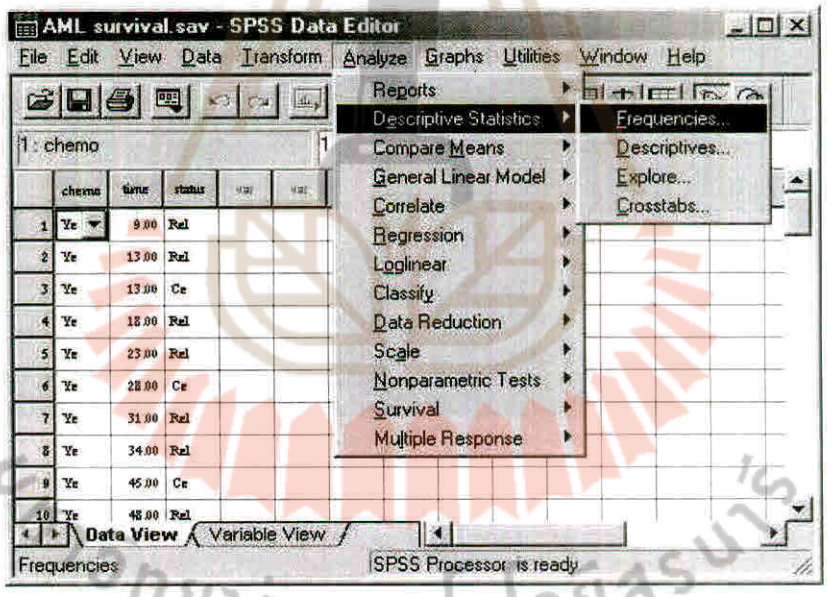
ตอนที่

8.3

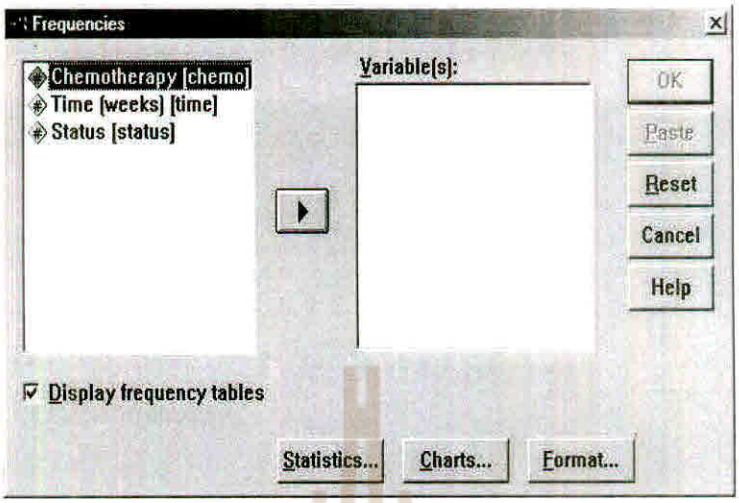
คำสั่งในการวิเคราะห์สถิติเชิงพรรณา

คำสั่งในการวิเคราะห์สถิติเชิงพรรณา ประกอบด้วยชุดคำสั่ง (1) Frequencies เพื่อสร้างตารางแจกแจงความถี่ (2) Descriptive เพื่อหาสถิติแนวโน้มเข้าสู่ส่วนกลาง และ (3) Correlation เพื่อหาค่าสัมประสิทธิ์สหสัมพันธ์

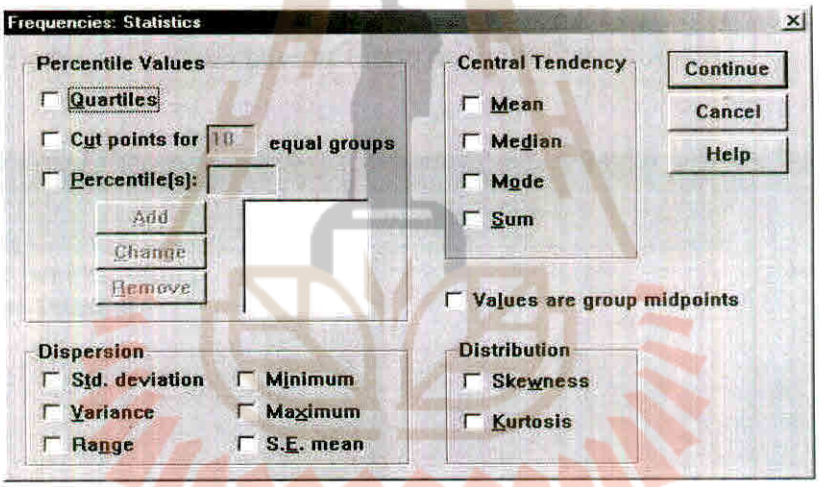
1. คำสั่ง Frequencies



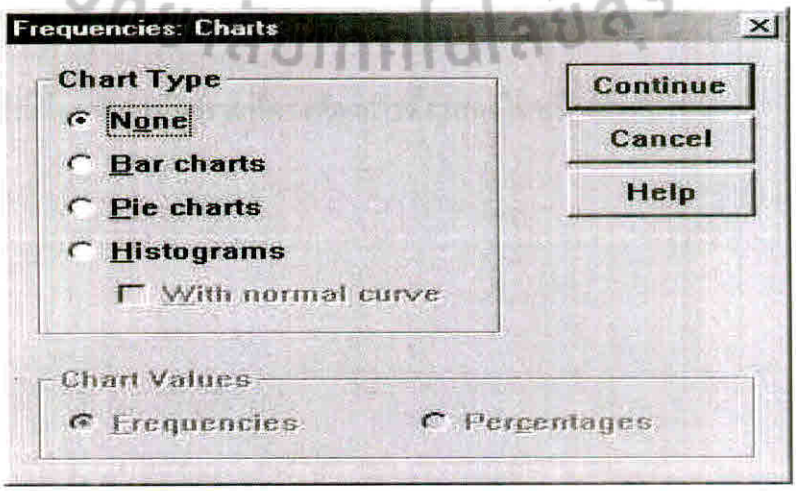
เมื่อเปิดแฟ้มข้อมูลแล้ว เลือก Analyze ⟶ Descriptive Statistics ⟶ Frequencies จะได้หน้าต่างใหม่ให้เลือกตัวแปรที่จะนำมาสร้างตารางแจกแจงหาความถี่



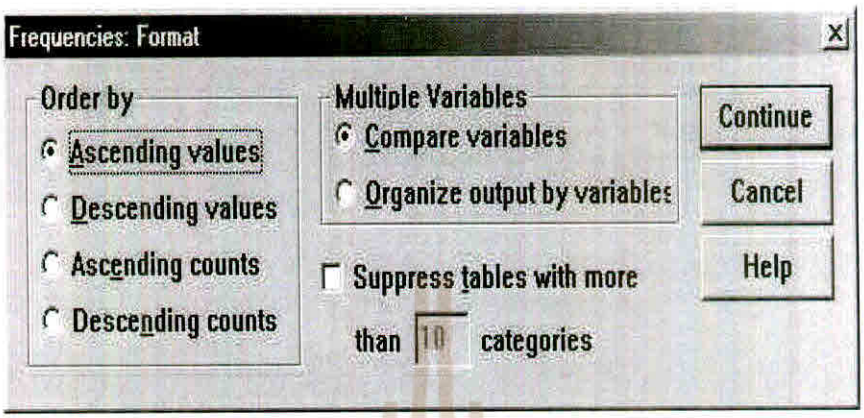
เมื่อเลือกตัวแปรที่จะนำมาทำตารางแจกแจงความถี่แล้วให้เลือก Statistics จะได้ หน้าต่างถัดไป เลือกตัวสถิติเพื่อประกอบตารางแจกแจงความถี่ที่ต้องการ โดยคลิ๊กที่หน้าค่าที่ต้องการ



เลือก Chart หรือ แผนภูมิที่ต้องการ

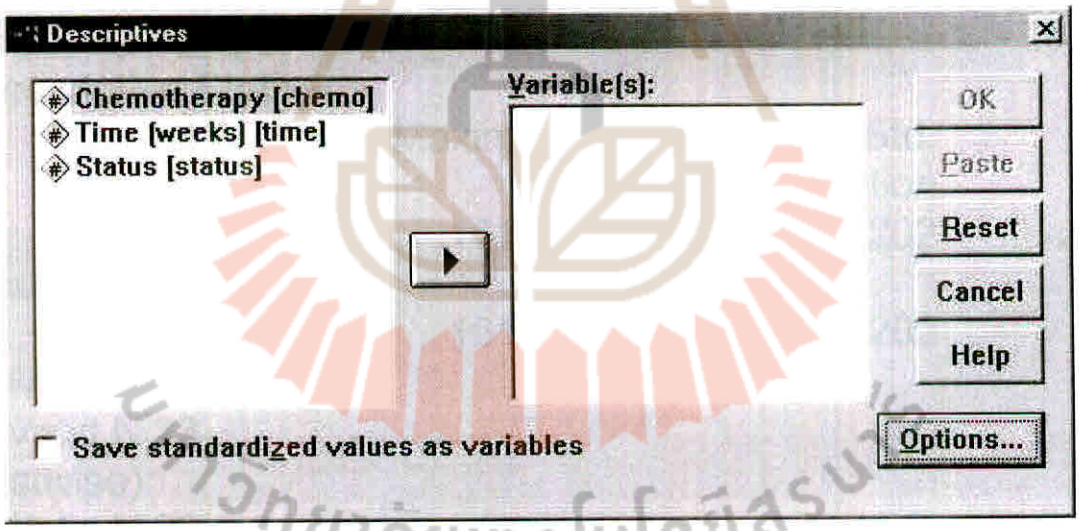


เลือก Format หรือรูปแบบในการจัดลำดับค่า หรือ จัดลำดับตามค่าความถี่

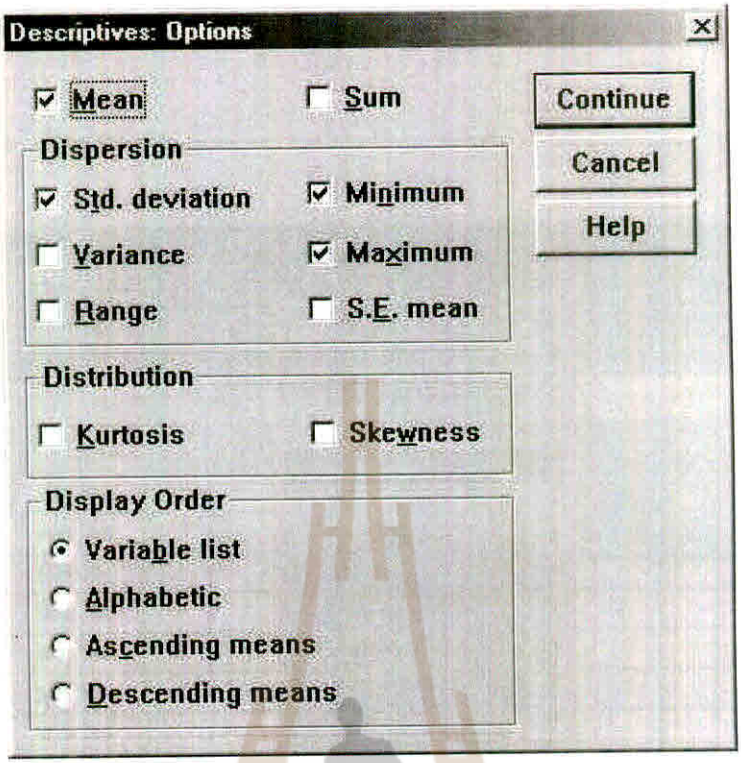


2. คำสั่ง Descriptive

เลือก Descriptive ⟶ Descriptive จะได้จอภาพให้เลือกตัวแปรที่ต้องการ



มี Options ให้เลือกทำการวิเคราะห์ตามต้องการดังแสดงในหน้าต่างต่อไปนี้



ตัวอย่าง ผลลัพธ์

Descriptive Statistics

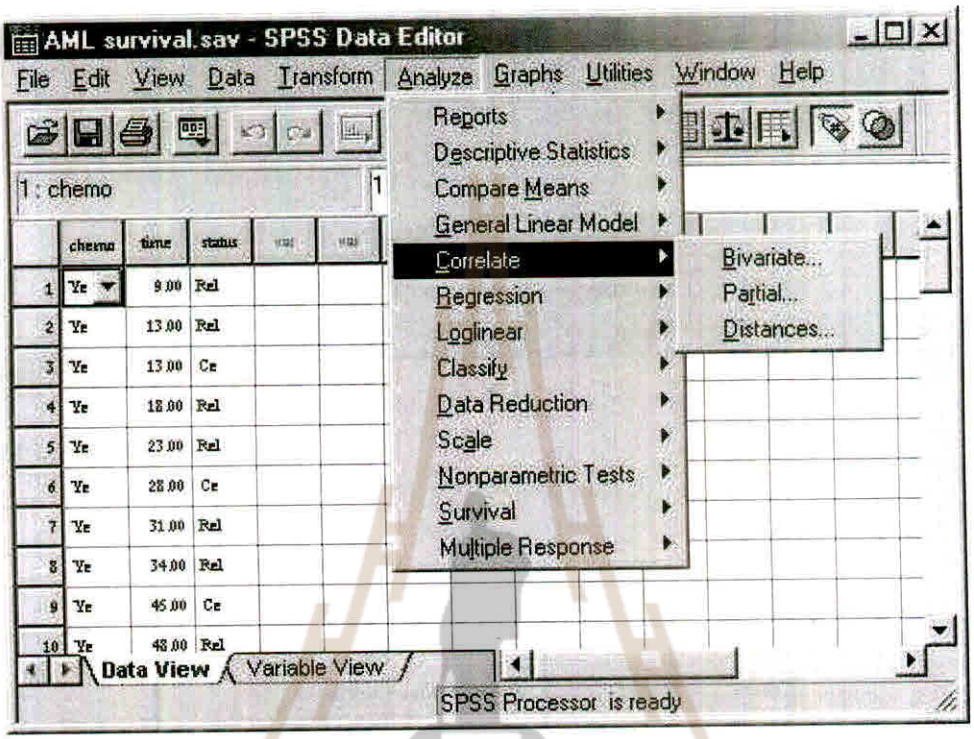
	N	Minimum	Maximum	Mean	Std. Deviation
Time 23 (weeks)	23	5.00	161.00	29.4783	31.7202
Time 23 (weeks)	23	5.00	161.00	29.4783	31.7202
Valid N (listwise)	23				
Valid N (listwise)	23				

เป็นตารางสรุป ค่าสถิติเชิงพรรณาของตัวแปรเวลาเป็นสัปดาห์ โดยค่าที่สังเกตมี 23 ค่า ค่าต่ำสุดเท่ากับ 5 ค่าสูงสุดเท่ากับ 161 ค่าเฉลี่ยเท่ากับ 29.4783 ค่าเบี่ยงเบนมาตรฐานเท่ากับ 31.7202

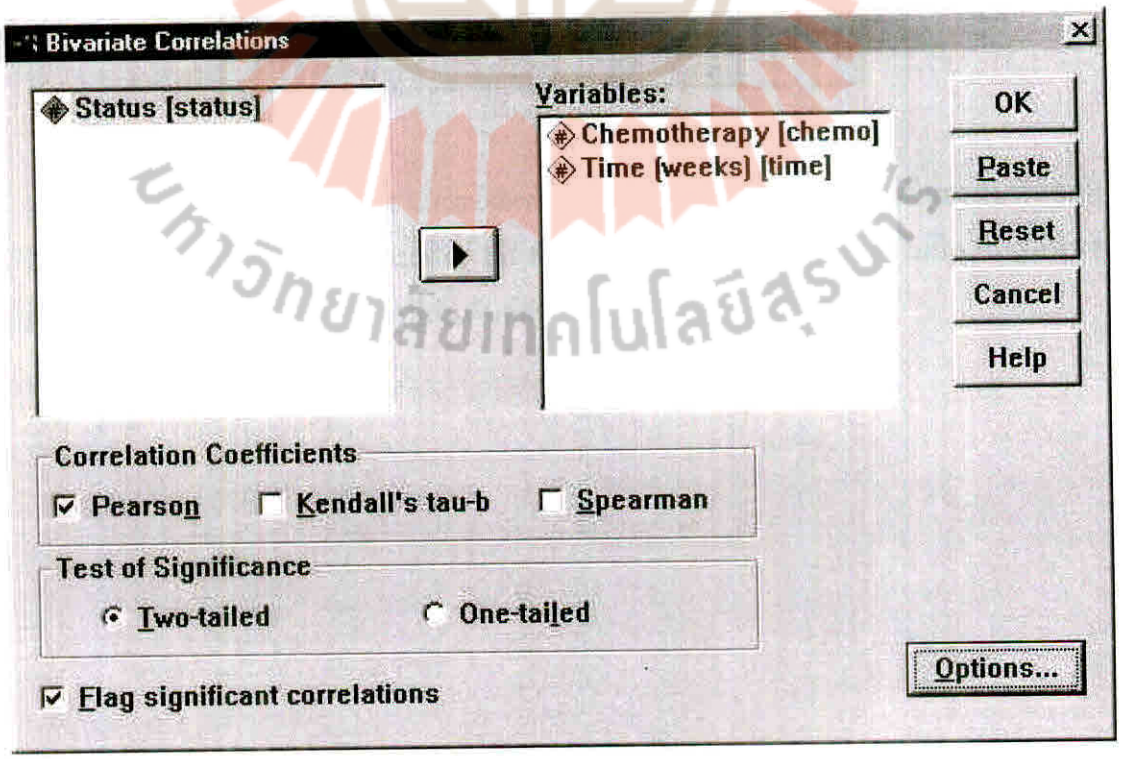
มหาวิทยาลัยเทคโนโลยีสุรนารี

3. คำสั่ง Correlation

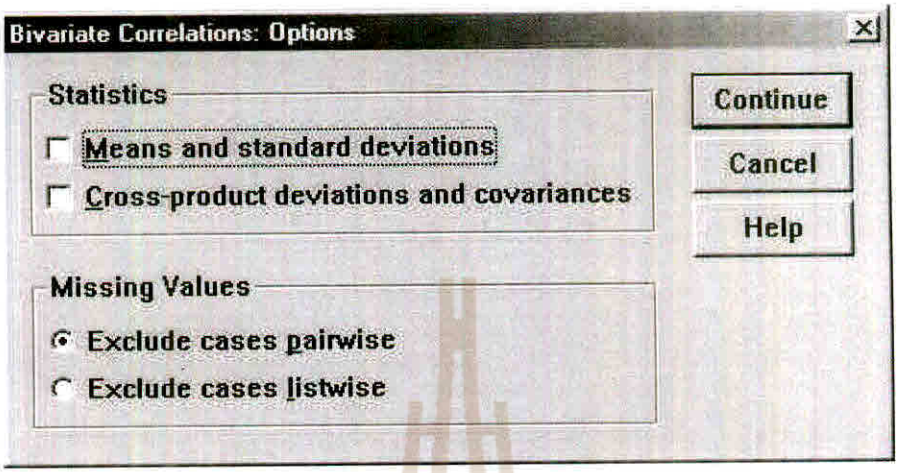
เลือก Analysis ⟶ Correlate



เลือก Bivariate จะได้หน้าจอ



เลือก Options… จะได้



ตัวอย่าง ผลลัพธ์

Correlations

		Chemotherapy	Time (weeks)
Chemotherapy	Pearson Correlation	1.000	.277
	Sig. (2-tailed)		.201
	Sum of Squares and Cross-products	5.739	98.739
	Covariance	.261	4.488
	N	23	23
Time (weeks)	Pearson Correlation	.277	1.000
	Sig. (2-tailed)	.201	.
	Sum of Squares and Cross-products	98.739	22135.739
	Covariance	4.488	1006.170
	N	23	23

จากที่หาค่าสหสัมพันธ์ระหว่าง Chemotherapy และ Time ที่รักษา พบว่ามีค่าสหสัมพันธ์กันเท่ากับ 0.277 เมื่อทดสอบทางสถิติแล้วไม่สามารถปฏิเสธว่าตัวแปรทั้งสองไม่มีความสัมพันธ์ นั่นคือยังบอกไม่ได้ว่าตัวแปรทั้งสองนี้สัมพันธ์กัน

สรุป

กิจกรรมต่อเนื่อง ตอนที่ 8.3

1. ศึกษาผ่านสื่อปฏิสัมพันธ์ผ่านจอภาพ การบรรยาย หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วยโปรแกรมสำเร็จรูป SPSS ตอนที่ 8.3 คำสั่งในการวิเคราะห์สถิติเชิงพรรณา
2. ศึกษาด้วยตนเองผ่านสื่อบทเรียนคอมพิวเตอร์ช่วยสอน หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วย โปรแกรมสำเร็จรูป SPSS ตอนที่ 8.3 คำสั่งในการวิเคราะห์สถิติเชิงพรรณา
3. ทำกิจกรรมประกอบการเรียนในคู่มือการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วย โปรแกรมสำเร็จรูป SPSS ตอนที่ 8.3 คำสั่งในการวิเคราะห์สถิติเชิงพรรณา
4. ทำกิจกรรมเสริมประสบการณ์ในแผนกิจกรรมการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วย โปรแกรมสำเร็จรูป SPSS ตอนที่ 8.3 คำสั่งในการวิเคราะห์สถิติเชิงพรรณา

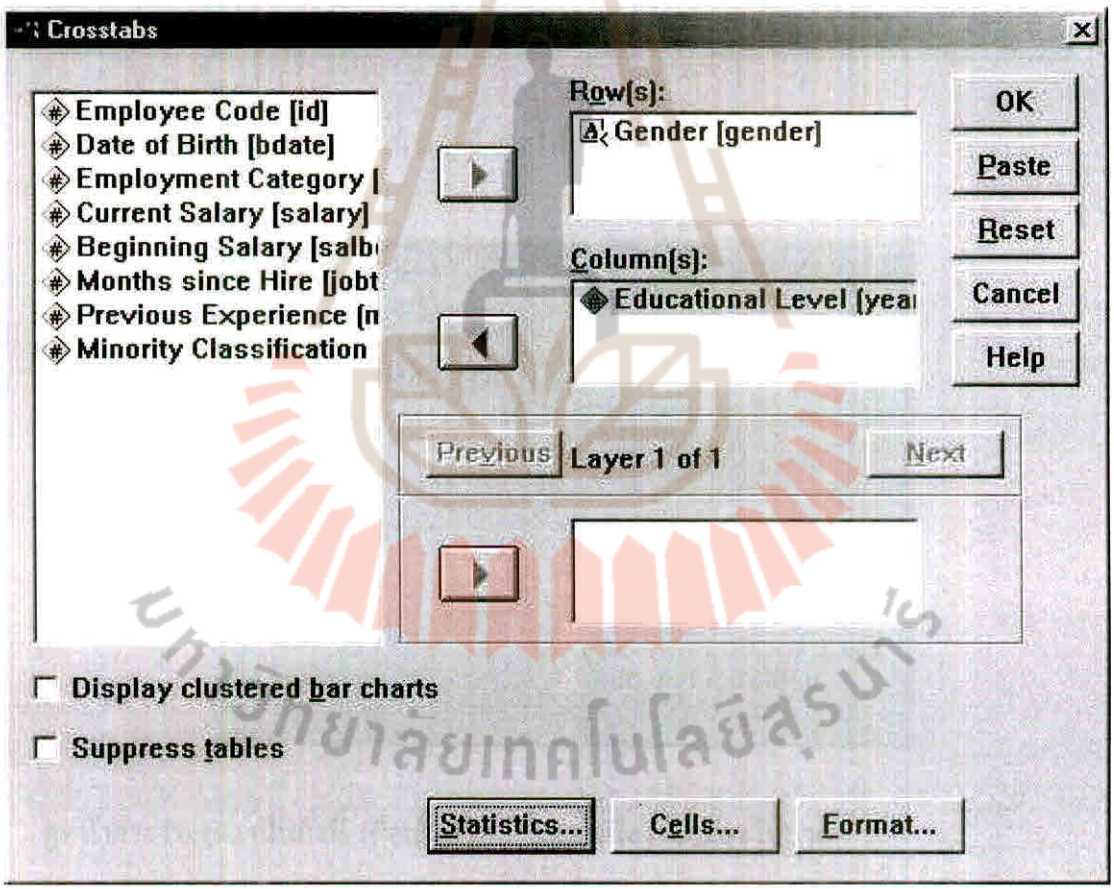
ตอนที่

8.4

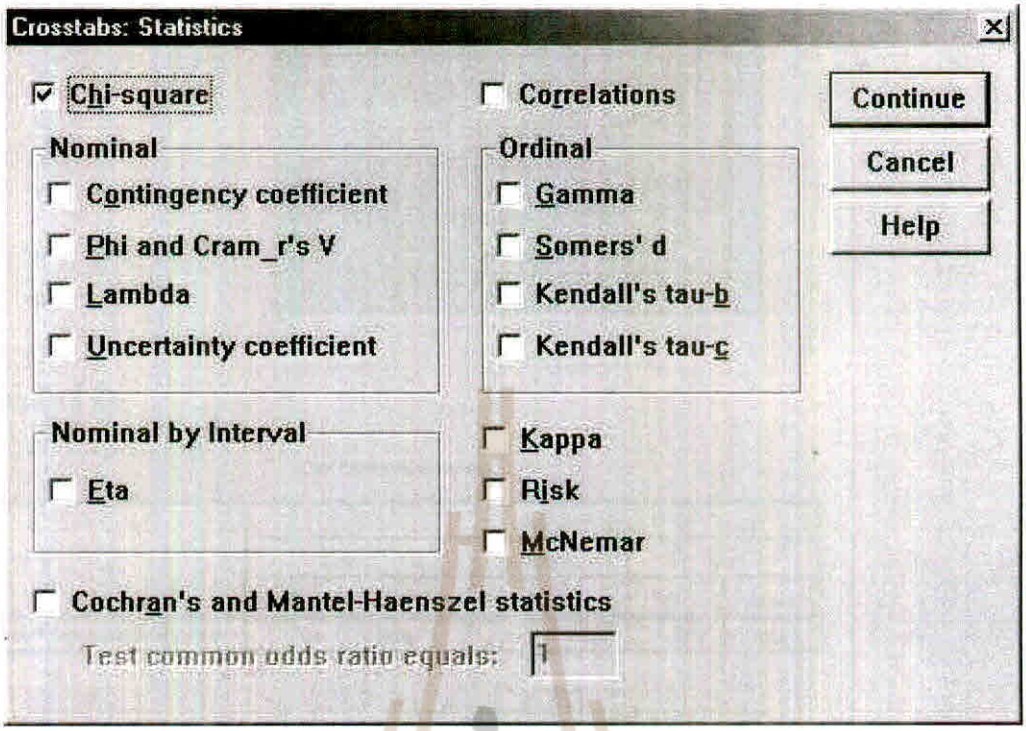
คำสั่งในการวิเคราะห์สถิติเชิงอนุมาน

1. คำสั่งทดสอบ t-test

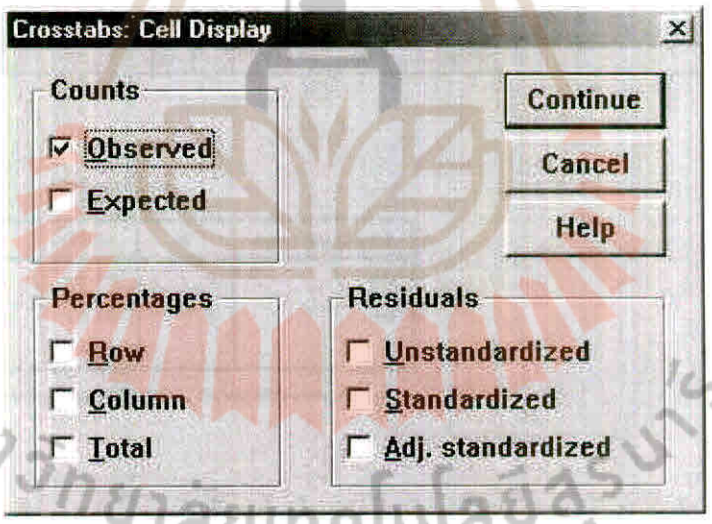
เลือก Descriptive ⟶ Crosstabs จะได้ดังภาพ



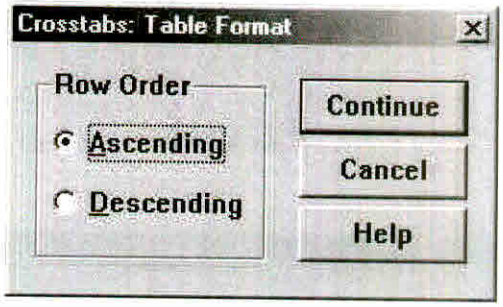
เลือกตัวแปรที่จะมาทดสอบความสัมพันธ์ คือ เพศ(Gender) และ ระดับการศึกษา(Education Level) เลือก Statistics แล้วเลือก Chi-Square ดังภาพ



ไป Continue แล้วดูหน้าต่าง Cell มีดังนี้ เพื่อให้นับค่าที่สังเกตได้(Observed)



ดูหน้าต่าง Format เป็นดังนี้ เพื่อเลือกให้ค่าในแถวเรียงจากน้อยไปมาก



ผลลัพธ์ คือ

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Gender * Educational Level (years)	474	100.0%	0	.0%	474	100.0%

Gender * Educational Level (years) Crosstabulation

Count

		Educational Level (years)										
		8	12	14	15	16	17	18	19	20	21	Total
Gender	Female	30	128		33	24	1					216
	Male	23	62	6	83	35	10	9	27	2	1	258
Total		53	190	6	116	59	11	9	27	2	1	474

Chi-Square Tests

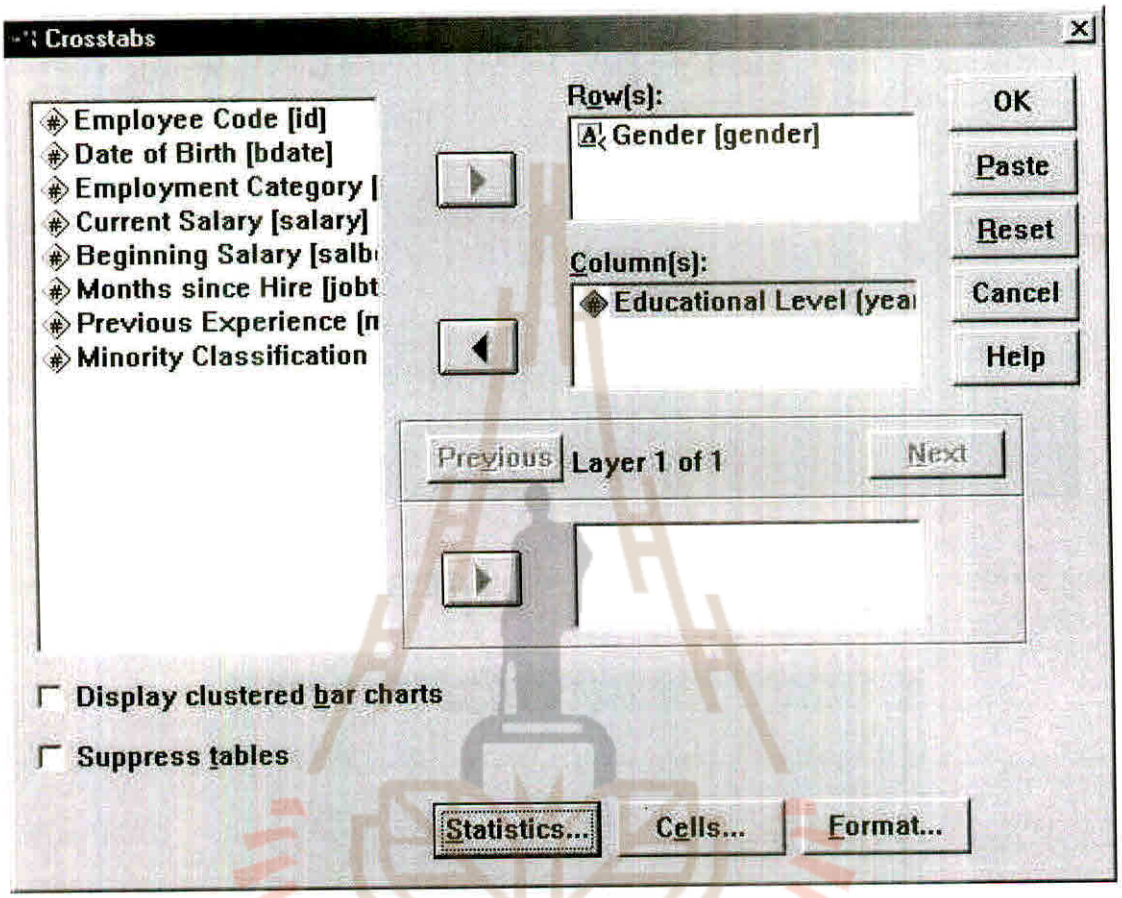
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	96.856[a]	9	.000
Likelihood Ratio	115.880	9	.000
N of Valid Cases	474		

a. 8 cells (40.0%) have expected count less than 5. The minimum expected count is .46.

ตารางแรกเป็นตารางสรุปจำนวนตัวอย่างในการวิเคราะห์ว่าทั้งหมดมี 474 ตัวอย่าง ตารางถัดมาเป็นตารางแจกแจงความถี่ 2 ทาง ระหว่างเพศ และ ระดับการศึกษา ตารางสุดท้ายเป็นตารางทดสอบความสัมพันธ์ระหว่างตัวแปรทั้งสอง ว่ามีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติหรือไม่ ปรากฏว่าในกรณีนี้ เพศและระดับการศึกษามีความสัมพันธ์กันอย่างมีนัยสำคัญอย่างยิ่ง

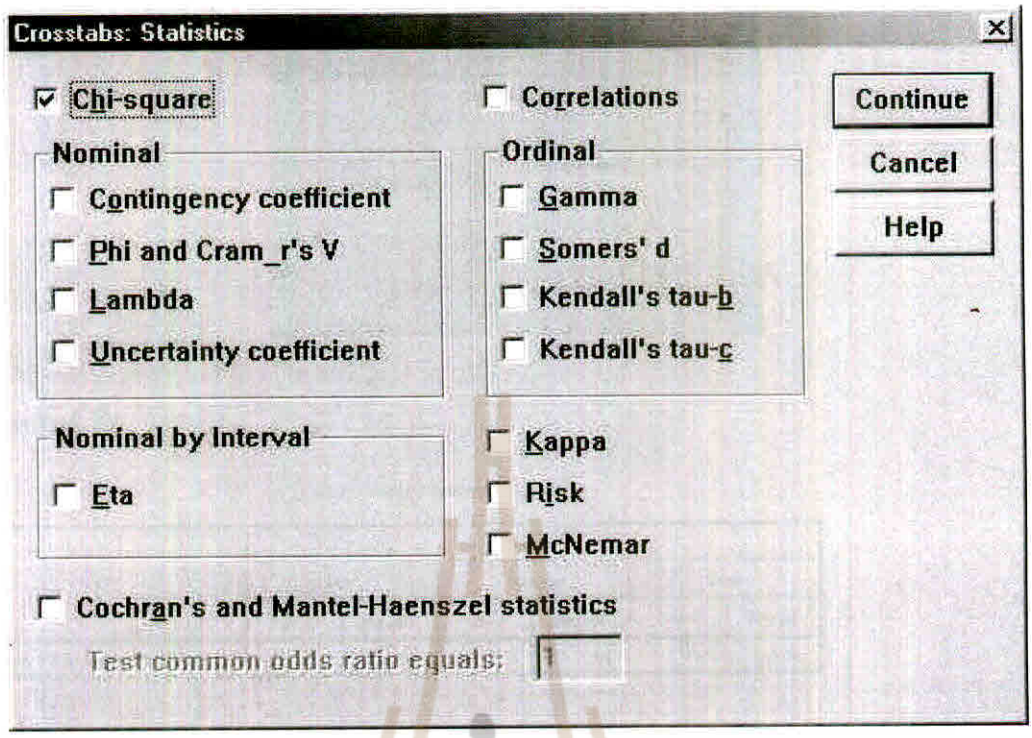
2. คำสั่งทดสอบ χ^2 -test

เลือก Descriptive ⟶ Crosstabs จะได้ดังภาพ

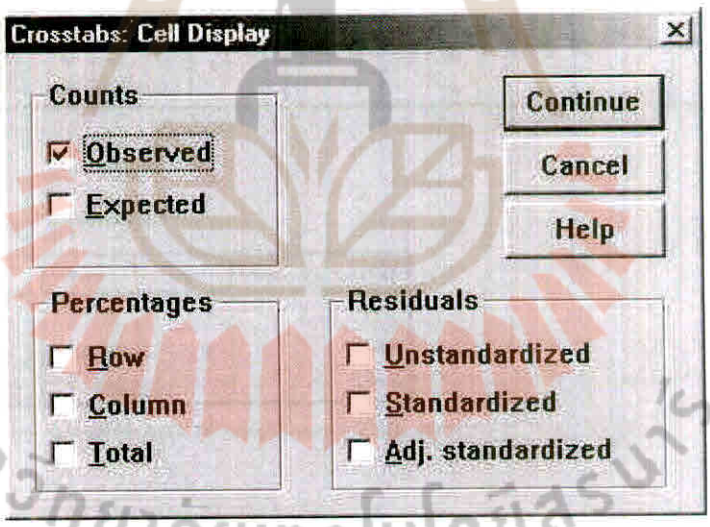


เลือก Statistics แล้วเลือก Chi-Square ดังภาพ

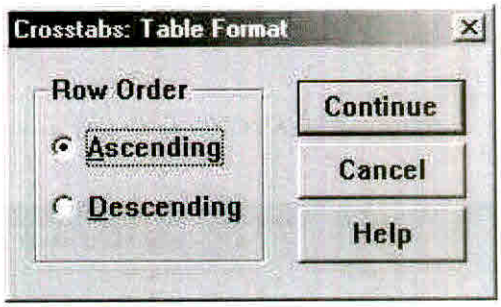
มหาวิทยาลัยเทคโนโลยีสุรนารี



ไป Continue แล้วดูหน้าต่าง Cell มีดังนี้



ดูหน้าต่าง Format เป็นดังนี้



ผลลัพธ์ คือ

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Gender * Educational Level (years)	474	100.0%	0	.0%	474	100.0%

Gender * Educational Level (years) Crosstabulation

Count

		Educational Level (years)										
		8	12	14	15	16	17	18	19	20	21	Total
Gender	Female	30	128		33	24	1					216
	Male	23	62	6	83	35	10	9	27	2	1	258
Total		53	190	6	116	59	11	9	27	2	1	474

Chi-Square Tests

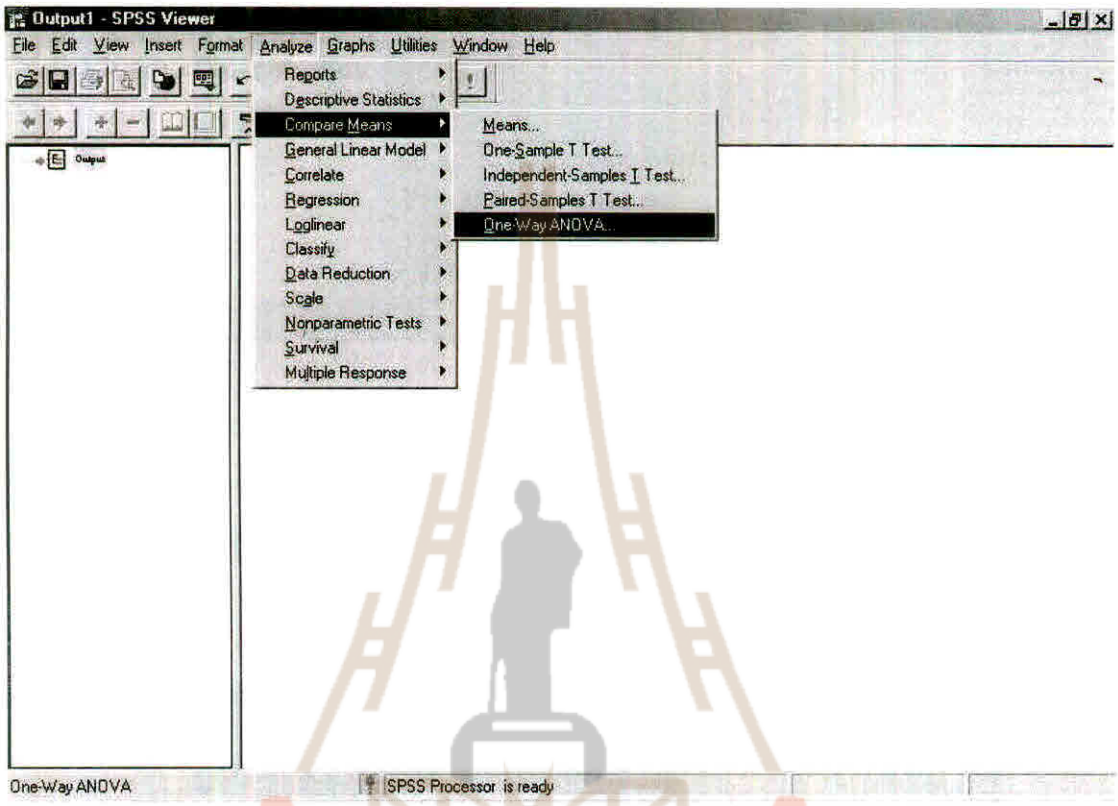
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	96.856[a]	9	.000
Likelihood Ratio	115.880	9	.000
N of Valid Cases	474		

a. 8 cells (40.0%) have expected count less than 5. The minimum expected count is .46.

มหาวิทยาลัยเทคโนโลยีสุรนารี

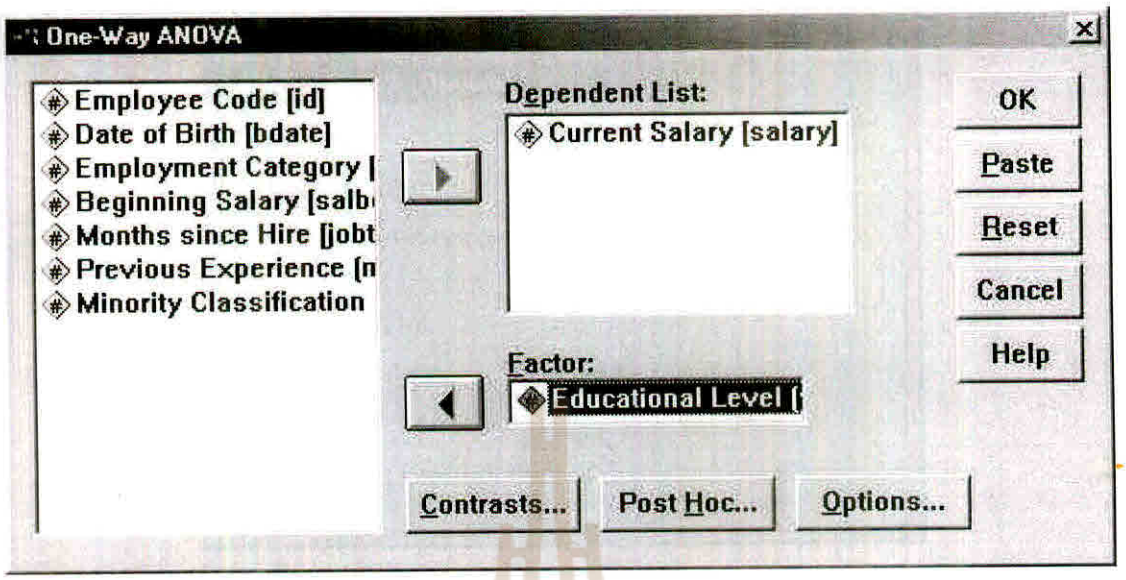
3. คำสั่ง F-test

เลือก Compare Means ⟶ One-Way ANOVA ได้ดังจอภาพ

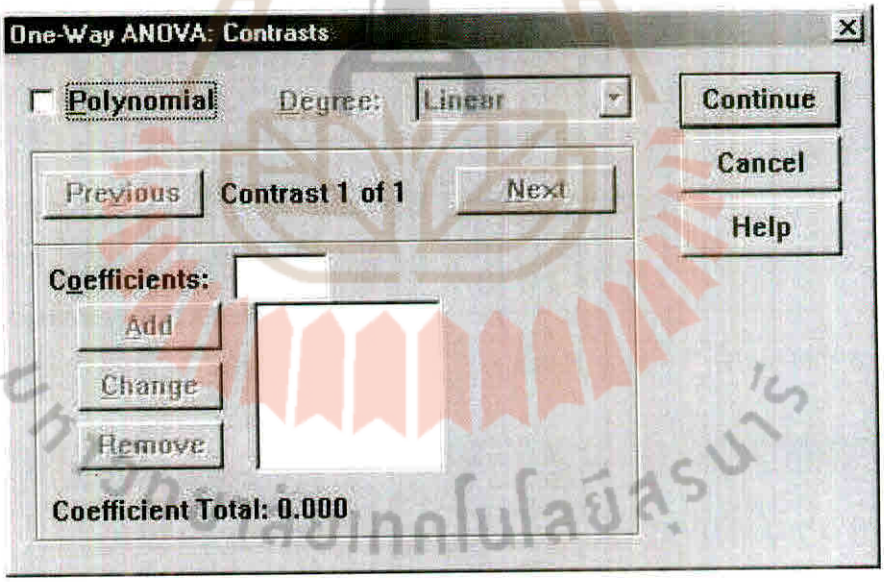


เลือกตัวแปรเพื่อทำการวิเคราะห์ ANOVA หรือ F-test ได้ดังจอภาพ เลือกตัวแปรตาม คือ เงินเดือนปัจจุบัน(Current Salary) ตัวแปรอิสระ คือ ระดับการศึกษา(Educational Level)

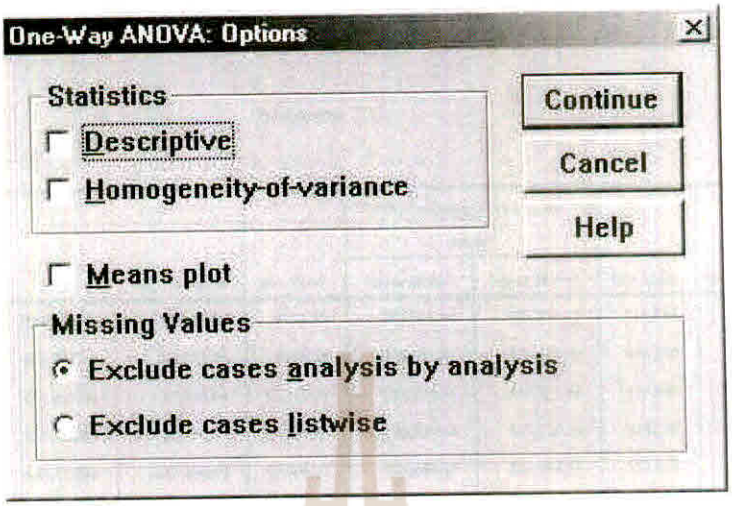
มหาวิทยาลัยเทคโนโลยีสุรนารี



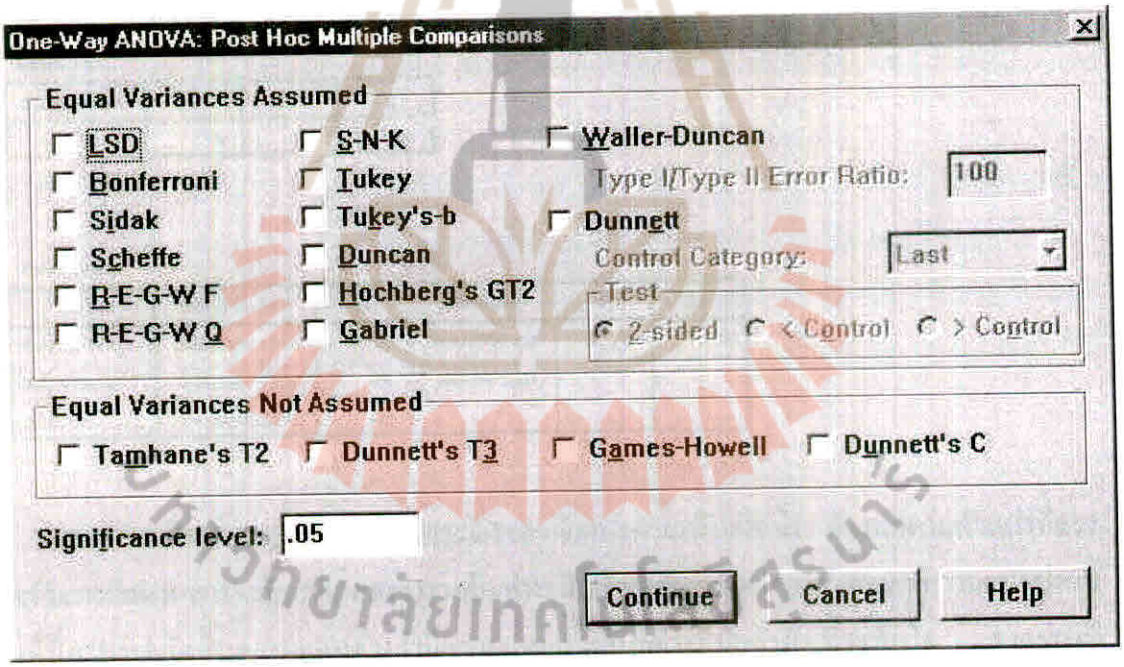
เลือก Contrasts เพื่อเตรียมวิเคราะห์ มีทางเลือกดังจอภาพ เพื่อดูความแตกต่างของค่าเฉลี่ยตามที่เราสนใจ ซึ่งเป็นสถิติขั้นสูงต่อจากการวิเคราะห์ความแปรปรวนแล้วพบว่ามีความแตกต่างอย่างมีนัยสำคัญ เพื่อดูว่ามีกลุ่มตัวแปรอิสระใดแตกต่างกัน หรือ กลุ่มตัวแปรอิสระมีรูปแบบความสัมพันธ์ตามที่กำหนด



เลือก Options เกี่ยวกับสถิติ ที่เกี่ยวข้อง เช่น สถิติเชิงพรรณา หรือ การทดสอบข้อตกลงเบื้องต้นเรื่องความแปรปรวนของแต่ละกลุ่มเท่ากัน (Homogeneity of variance) จะมีทางเลือกดังภาพ



เลือก Post Hoc Multiple Comparisons จะได้ ทางเลือกดังนี้ เมื่อทดสอบแล้วทราบว่า AI มีความแตกต่างกันอย่างมีนัยสำคัญ จึงไปดูว่าค่าเฉลี่ยคู่ใดบ้างที่แตกต่างกัน ด้วยมีวิธีการต่างๆ ที่จะเรียนในวิชาสถิติขั้นสูงต่อไป และด้านล่างของหน้าต่างนี้สามารถระบุระดับนัยสำคัญที่ต้องการได้



ผลลัพธ์ที่ได้รับ

Descriptives

Current Salary

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
8	53	$24,399.06	$5,190.48	$712.97	$22,968.38	$25,829.73	$15,750	$34,500
12	190	$25,887.16	$5,559.76	$403.35	$25,091.52	$26,682.80	$16,200	$59,400
14	6	$31,625.00	$5,790.40	$2,363.92	$25,548.35	$37,701.65	$25,950	$39,900
15	116	$31,685.00	$8,401.57	$780.07	$30,139.84	$33,230.16	$18,750	$80,000
16	59	$48,225.93	$17,430.13	$2,269.21	$43,683.62	$52,768.25	$23,100	$103,750
17	11	$59,527.27	$19,789.56	$5,966.78	$46,232.47	$72,822.08	$27,000	$82,500
18	9	$65,127.78	$13,461.88	$4,487.29	$54,780.06	$75,475.49	$43,950	$86,250
19	27	$72,520.37	$20,285.48	$3,903.94	$64,495.70	$80,545.04	$36,000	$135,000
20	2	$64,312.50	$6,982.68	$4,937.50	$1,575.61	$127,049.39	$59,375	$69,250
21	1	$65,000.00	.	.	.	.	$65,000	$65,000
Total	474	$34,419.57	$17,075.66	$784.31	$32,878.40	$35,960.73	$15,750	$135,000

Test of Homogeneity of Variances

Current Salary

Levene Statistic	df1	df2	Sig.
14.473	9	464	.000

ANOVA

Current Salary

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	88653535061.984	9	9850392784.665	92.779	.000
Within Groups	49262960374.356	464	106170173.221		
Total	137916495436.340	473			

ตารางแรกเป็นการสรุปค่าสถิติเชิงพรรณาของตัวแปรตาม(เงินเดือน) จำแนกตามตัวแปรอิสระ (ระดับการศึกษา) ตารางถัดมาอีกสองตารางเป็นค่าระดับนัยสำคัญหลังจากการวิเคราะห์ความแปรปรวน ในที่นี้สามารถตัดสินความแตกต่างว่ามีความแตกต่างกันอย่างมีนัยสำคัญที่ระดับใด จากช่องที่ตรงกับ Sig ว่ามีค่าน้อยกว่าระดับนัยสำคัญที่ตั้งไว้หรือไม่ ในที่นี้ค่า Sig =.000 ซึ่งแสดงว่าระดับเงินเดือนแตกต่างกันตามระดับการศึกษาอย่างมีนัยสำคัญทางสถิติ ที่ระดับที่ตั้งไว้ ในที่นี้ ตั้งไว้ที่ 0.05 ค่าที่ได้จากการคำนวณมีค่าน้อยกว่า 0.05 ก็แสดงว่าผู้ที่มีการศึกษาระดับต่างกันจะมีเงินเดือนแตกต่างกันอย่างมีนัยสำคัญทางสถิติ

สรุป

กิจกรรมต่อเนื่อง ตอนที่ 8.4

1. ศึกษาผ่านสื่อปฏิสัมพันธ์ผ่านจอภาพ การบรรยาย หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วยโปรแกรมสำเร็จรูป SPSS ตอนที่ 8.4 คำสั่งในการวิเคราะห์สถิติเชิงอนุมาน
2. ศึกษาด้วยตนเองผ่านสื่อบทเรียนคอมพิวเตอร์ช่วยสอน หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วย โปรแกรมสำเร็จรูป SPSS ตอนที่ 8.4 คำสั่งในการวิเคราะห์สถิติเชิงอนุมาน
3. ทำกิจกรรมประกอบการเรียนในคู่มือการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วย โปรแกรมสำเร็จรูป SPSS ตอนที่ 8.4 คำสั่งในการวิเคราะห์สถิติเชิงอนุมาน
4. ทำกิจกรรมเสริมประสบการณ์ในแผนกิจกรรมการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วย โปรแกรมสำเร็จรูป SPSS ตอนที่ 8.4 คำสั่งในการวิเคราะห์สถิติเชิงอนุมาน

ตอนที่

8.5

การตีความจากผลการวิเคราะห์ข้อมูลด้วยสถิติ

การแปลผลการวิเคราะห์ข้อมูลทางสถิติ จะต้องคำนึงถึงระดับนัยสำคัญทางสถิติ และสมมติฐานทางสถิติที่ได้กำหนดไว้ใน

1. นัยสำคัญทางสถิติ

ระดับนัยสำคัญทางสถิติ (Statistical Significance Level) ในการทดสอบสมมติฐานทางสถิติ เราจะยอมรับสมมติฐานที่ตั้งขึ้นก็ต่อเมื่อเราคำนวณค่าความน่าจะเป็นได้มากกว่าค่าระดับนัยสำคัญทางสถิติ ซึ่งใช้แทนด้วยสัญลักษณ์ α และถ้าเราคำนวณค่าความน่าจะเป็นออกมาได้น้อยกว่า α ก็แสดงว่าเราจะปฏิเสธสมมติฐานที่ตั้ง

ถ้า $\alpha = 0.05$ และปฏิเสธสมมติฐานที่ตั้ง เรียกว่าปฏิเสธสมมติฐานอย่างมีนัยสำคัญ

ถ้า $\alpha = 0.01$ และปฏิเสธสมมติฐานที่ตั้ง เรียกว่าปฏิเสธสมมติฐานอย่างมีนัยสำคัญอย่างยิ่ง

ตัวอย่าง ถ้าเราต้องการทดสอบสมมติฐานว่าเหรียญบาทมีความสมดุล คือโอกาสที่จะเกิดหน้าหัวหรือหน้าก้อยเท่ากัน เท่ากับ $\frac{1}{2}$ หรือไม่ โดยทำการทดลองโยนเหรียญบาท 100 ครั้ง ปรากฏว่าออกหน้าหัว 60 ครั้ง

วิธีทำ เพื่อหาคำตอบสมมติฐานที่ตั้ง เราก็ต้องดูว่าความน่าจะเป็นที่เหรียญบาทที่สมดุลจะออกหน้าหัวมากกว่า 60 ครั้ง เป็นเท่าใด แล้วไปเปรียบเทียบกับระดับนัยสำคัญทางสถิติ

$$P\{X \geq 60\} = \sum_{X=60}^{100} \binom{100}{x} \left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{100-x}$$

ประมาณด้วยโค้งปกติมาตรฐาน

$$\mu = np = 100.\left(\frac{1}{2}\right) = 50$$

$$\sigma = \sqrt{pqn} = \sqrt{\frac{1}{2}.\frac{1}{2}.100} = \sqrt{25} = 5$$

$$\therefore Z = \frac{59.5-50}{5} = 1.90$$

นำค่า Z ไปเปิดตารางหาพื้นที่ที่ $Z \geq 1.9$ จะได้เท่ากับ 0.0287

$\therefore P\{X \geq 60\} = 0.0287$

นั้นก็แสดงว่าโอกาสที่จะเกิดเหตุการณ์ที่เหรียญจะออกหน้าหัวมากกว่า 60 ครั้ง ถ้าเหรียญนั้นเป็นเหรียญสมดุล เท่ากับ 0.0287 ซึ่งมีค่าน้อยกว่า $\alpha = 0.05$ ก็แสดงว่าปฏิเสธสมมติฐานที่ตั้งว่าเหรียญสมดุลอย่างมีนัยสำคัญ แต่มีมากกว่า $\alpha = 0.01$ ก็แสดงว่ายอมรับสมมติฐานที่ตั้งว่าเหรียญสมดุลที่ระดับสำคัญอย่างยิ่ง

ตอบ

2. สมมติฐานทางสถิติ

สมมติฐานว่างและสมมติฐานทางเลือกอื่น (Null Hypothesis and Alternative Hypothesis)

การตั้งสมมติฐานทางสถิติ มีลักษณะเดียวกับการตัดสินคดีของศาล คือเมื่อมีผู้ต้องหา เช่น อัยการยื่นฟ้องว่า นาย ก.เป็นฆาตกร ศาลก็จะตั้งสมมติฐานว่างว่า นาย ก.เป็นผู้บริสุทธิ์ แล้วจากพยานหลักฐานที่มี สามารถจะปฏิเสธสมมติฐานว่างของ นาย ก. ได้หรือไม่ ถ้ามีพยานและหลักฐานเพียงพอที่จะปฏิเสธสมมติฐานว่างนี้ได้ นั่นก็แสดงว่ายอมรับว่า นาย ก.เป็นฆาตกร ถ้าพิจารณาให้ดีก็จะเห็นว่า สมมติฐานทางเลือกอื่นก็คือ นาย ก. เป็นฆาตกร ในทางสถิติก็เช่นเดียวกัน เพียงแต่สมมติฐานที่ตั้งนั้นจะต้องเป็นสมมติฐานทางสถิติคือเป็นสมมติฐานที่เกี่ยวกับค่าพารามิเตอร์ ปกติสมมุติฐานว่างแทนด้วยสัญลักษณ์ H_o และสมมติฐานทางเลือกอื่นแทนด้วย H_A ดังตัวอย่าง เช่น

H_o : สัดส่วนความนิยมนายไข่เท่ากับ 0.8

H_A : สัดส่วนความนิยมนายไข่น้อยกว่า 0.8

อาจจะแทนด้วยสัญลักษณ์ได้ดังนี้

H_o : $P = 0.8$

H_A : $P < 0.8$

ตามปกติทั่วไปการตั้งสมมติฐานว่าง จะตั้งเป็นแบบแน่นอนตายตัว ซึ่งเรียกว่าสมมติฐานเชิงเดียว (Simple Hypothesis) โดยตั้งอยู่ในรูปของพารามิเตอร์เท่ากับค่าคงตัว ส่วนการตั้งสมมติฐานทางเลือกอื่น จะต้องเป็นแบบประกอบ (Composite Hypothesis) โดยตั้งอยู่ในรูปเป็นช่วงของพารามิเตอร์ ได้แก่ พารามิเตอร์น้อยกว่า หรือมากกว่า หรือไม่เท่ากับ ค่าคงตัวในสมมติฐานว่าง นั่นคือนอกจากจะตั้งสมมติฐานข้างต้นแล้ว อาจจะตั้งเป็น

H_o : $P = 0.8$

H_A : $P > 0.8$

หรือ H_o : $P = 0.8$

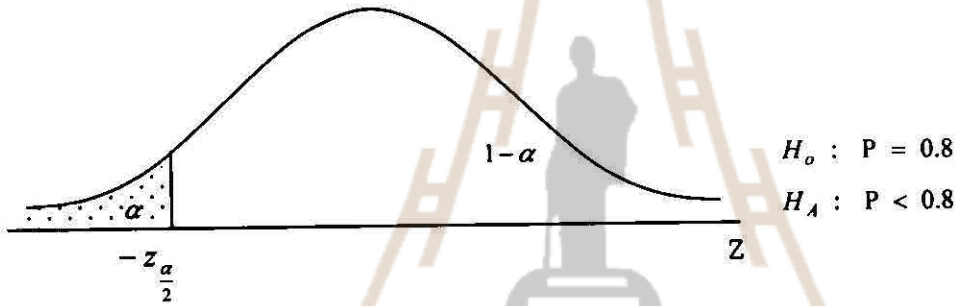
H_A : $P \neq 0.8$

การตั้งสมมติฐานแบบทางเดียวหรือสองทาง (One-tailed and Two-tailed Testing Hypothesis)

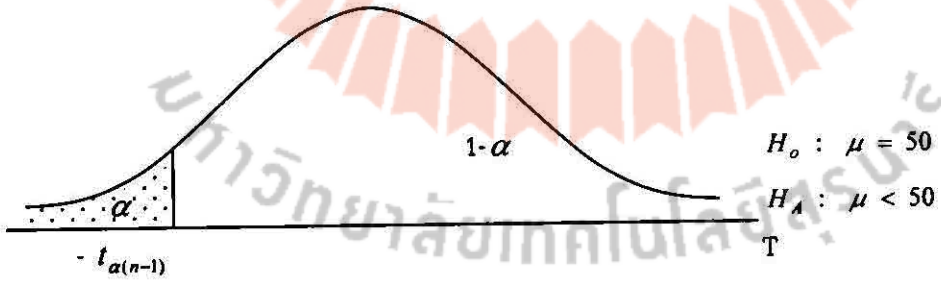
การตั้งสมมติฐานแบบทางเดียวหรือสองทาง ต่างกันตรงที่พื้นที่ของค่านัยสำคัญ (α) ในการพิจารณาว่าจะยอมรับหรือปฏิเสธสมมติฐานว่าง การตั้งสมมติฐานว่างเหมือนกันในทุกกรณี เพราะสมมติฐานว่างเป็นสมมติฐานที่เราตั้งไว้เพื่อปฏิเสธ นั่นคือถ้ามีข้อมูลเพียงพอก็จะปฏิเสธสมมติฐานว่างเพื่อไปยอมรับสมมติฐานทางเลือกอื่นที่เราคาดว่าจะเป็น บางทีเราเรียกสมมติฐานทางเลือกอื่นว่าเป็นสมมติฐานของการวิจัย เพราะมีทิศทางที่เราคาดการว่าจะมากกว่า น้อยกว่า หรือไม่เท่ากับ ค่าพารามิเตอร์ที่เราสนใจ ซึ่งสามารถพิจารณาได้เป็นกรณี ๆ ดังต่อไปนี้

การทดสอบทางเดียวข้างน้อย

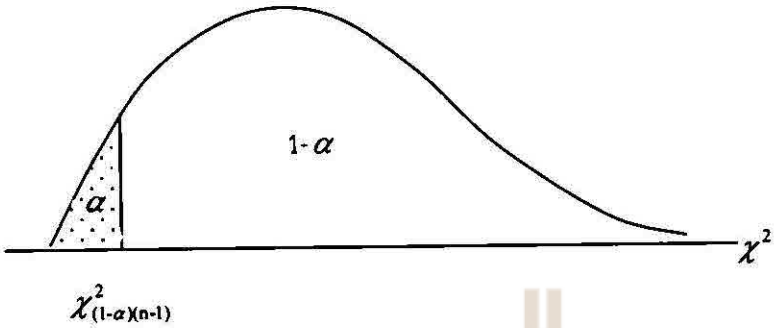
เป็นการตั้งสมมติฐานในกรณีที่เราคาดว่า ถ้ามีการปฏิเสธสมมติฐานว่างแล้ว สมมติฐานทางเลือกอื่นก็คือค่าพารามิเตอร์จะต้องน้อยกว่าค่าคงตัวในสมมติฐานว่าง พิจารณาได้จากภาพ



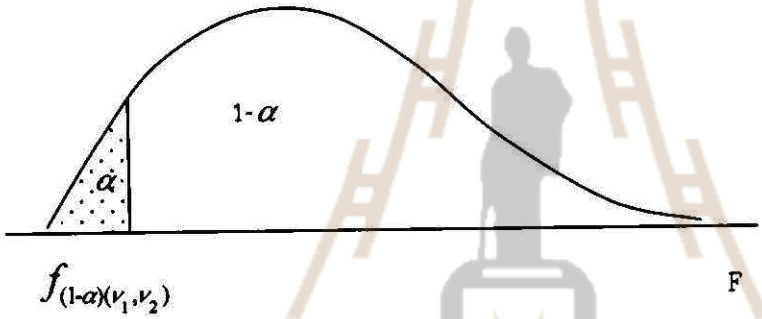
ภาพที่ 8.1 แสดงพื้นที่ของการปฏิเสธสมมติฐานว่างในกรณีเป็นการทดสอบเดียวข้างน้อย และการแจกแจงของตัวประมาณค่าแบบปกติมาตรฐาน



ภาพที่ 8.2 แสดงพื้นที่ของการปฏิเสธสมมติฐานว่างในกรณีเป็นการทดสอบทางเดียวข้างน้อย และการแจกแจงของตัวประมาณค่าเป็นแบบ t



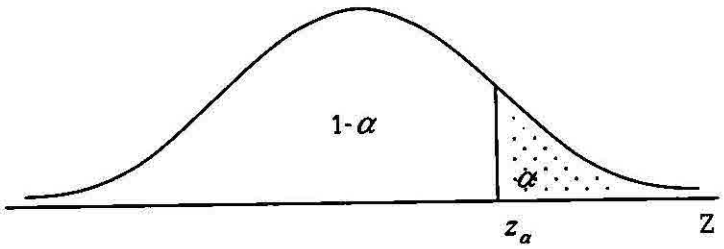
ภาพที่ 8.3 แสดงพื้นที่ของการปฏิเสธสมมติฐานว่างในกรณีเป็นการทดสอบทางเดียวข้างน้อย และการแจกแจงของตัวประมาณค่าเป็นแบบ χ^2



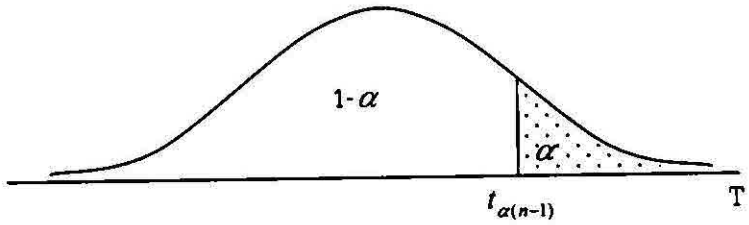
ภาพที่ 8.4 แสดงพื้นที่ของการปฏิเสธสมมติฐานว่างในกรณีเป็นการทดสอบทางเดียวข้างน้อย และการแจกแจงของตัวประมาณค่าเป็นแบบ F

การทดสอบทางเดียวข้างมาก

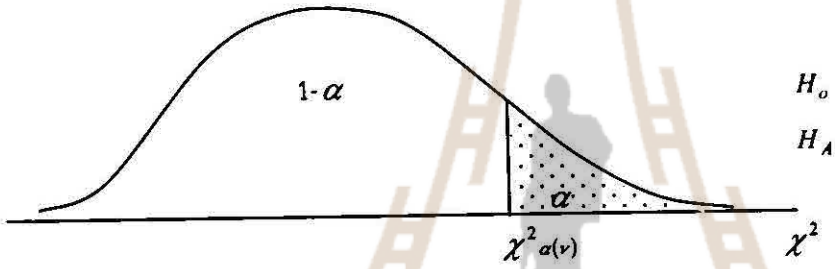
เป็นการตั้งสมมติฐานในกรณีที่เราคาดว่า ถ้ามีการปฏิเสธสมมติฐานว่างแล้วสมมติฐานทางเลือกอื่นก็คือค่าพารามิเตอร์ จะต้องมากกว่าค่าคงตัว ในสมมติฐานว่าง พิจารณาได้จากภาพ



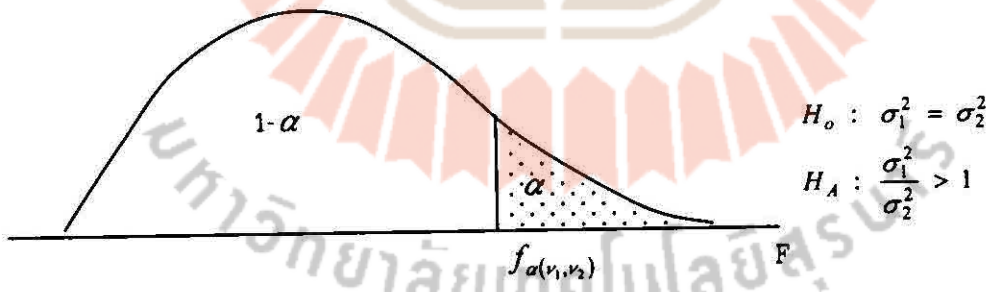
ภาพที่ 8.5 แสดงพื้นที่ของการปฏิเสธสมมติฐานว่างในกรณีที่เป็นการทดสอบทางเดียวข้างมาก และการแจกแจงของตัวประมาณค่าเป็นแบบปกติมาตรฐาน



ภาพที่ 8.6 แสดงพื้นที่ของการปฏิเสธสมมติฐานว่างในกรณีที่เป็นการทดสอบทางเดียวข้างมาก และการแจกแจงของตัวประมาณค่า เป็นแบบ t



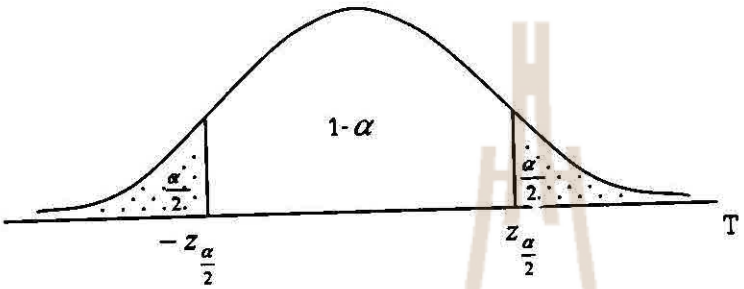
ภาพที่ 8.7 แสดงพื้นที่ของการปฏิเสธสมมติฐานในกรณีที่เป็นการทดสอบทางเดียวข้างมาก และการแจกแจงของตัวประมาณค่าเป็นแบบ χ^2



ภาพที่ 8.8 แสดงพื้นที่ของการปฏิเสธสมมติฐานว่างในกรณีที่เป็นการทดสอบทางเดียวข้างมาก และการแจกแจงของตัวประมาณค่าเป็นแบบ F

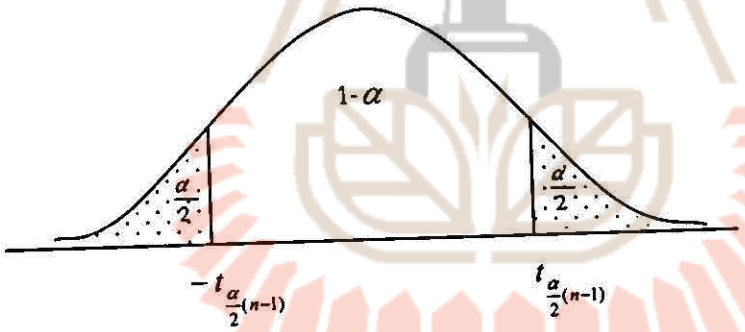
การทดสอบสองทาง

เป็นการตั้งสมมติฐานในกรณีที่เราคาดว่าถ้ามีการปฏิเสธสมมติฐานว่างแล้วไม่ทราบว่าสมมติฐานทางเลือกอื่นมีค่าพารามิเตอร์น้อยกว่าหรือมากกว่าค่าคงตัวในสมมติฐานว่าง ดังนั้นเราจึงแบ่ง α ออกไปสองข้าง ๆ ละ $\frac{\alpha}{2}$ พิจารณาได้จากภาพที่ 8.9, 8.10, 8.11 และ 8.12



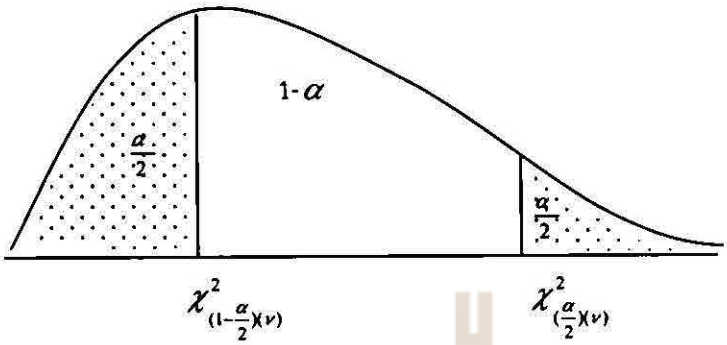
H_o : P = 0.8
H_A : P ≠ 0.8

ภาพที่ 8.9 แสดงพื้นที่ของการปฏิเสธสมมติฐานว่างในกรณีที่เป็นการทดสอบสองทางและการแจกแจงของตัวประมาณค่าเป็นแบบปกติมาตรฐาน

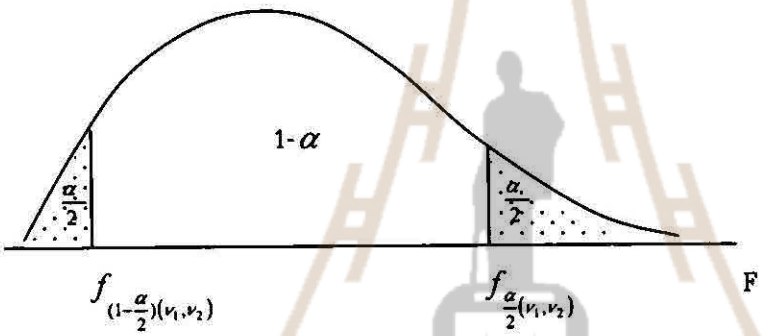


H_o : μ = 50
H_A : $\mu \neq$ 50

ภาพที่ 8.10 แสดงพื้นที่ของการปฏิเสธสมมติฐานว่างในกรณีที่เป็นการทดสอบสองทางและการแจกแจงของตัวประมาณค่าเป็นแบบ t



ภาพที่ 8.11 แสดงพื้นที่ของการปฏิเสธสมมติฐานว่างในกรณีที่เป็นการทดสอบสองทางและการแจกแจงของตัวประมาณค่าเป็นแบบ χ^2



ภาพที่ 8.12 แสดงพื้นที่ของการปฏิเสธสมมติฐานว่างในกรณีที่เป็นการทดสอบสองทางและแจกแจงของตัวประมาณค่าเป็นแบบ F

เกณฑ์ในการยอมรับหรือปฏิเสธสมมติฐานว่าง (Criterion for Accept or Reject Null Hypothesis)

เกณฑ์ในการยอมรับหรือปฏิเสธสมมติฐาน ขึ้นอยู่กับระดับนัยสำคัญ ดังที่อธิบายไว้ในหัวข้อที่ผ่านมา ซึ่งเราสามารถนำมาเปรียบเทียบกับค่าตัวแปรของการแจกแจงของความน่าจะเป็น (Probability Distribution) อาทิเช่น ค่า Z, T, χ^2 และ F โดยเรียกขอบเขตในการปฏิเสธสมมติฐานว่างว่า เขตวิกฤต (Critical Region) ซึ่งสัมพันธ์กับระดับนัยสำคัญอันขึ้นอยู่กับการตั้งสมมติฐานว่าเป็นแบบทางเดียวข้างน้อย ทางเดียวข้างมาก หรือสองทาง โดยพอจะสรุป ดังตารางที่ 8.1 ต่อไปนี้

ตารางที่ 8.1 แสดงหาเขตวิกฤตในกรณีต่าง ๆ

จากรูปที่	สมมติฐานว่าง	สมมติฐานทางเลือกอื่น	เขตวิกฤต
1	P = 0.8	P < 0.8	$Z < -z_{\alpha}$
2	$\mu = 50$	$\mu < 50$	$T < -t_{\alpha(n-1)}$
3	$\sigma^2 = 2$	$\sigma^2 < 2$	$\chi^2 < \chi^2_{(1-\alpha)(\nu)}$
4	$\sigma_1^2 = T_2^2$	$\frac{\sigma_2^2}{\sigma_2^2} < 1$	$F < f_{(1-\alpha)(\nu_1,\nu_2)}$
5	P=8	P>0.8	$Z > z_{\alpha}$
6	$\mu = 50$	$\mu < 50$	$T > t_{\alpha(n-1)}$
7	$\sigma^2 = 2$	$\sigma_2^2 > 2$	$x^2 > x^2_{\alpha(\nu)}$
8	$\sigma_1^2 = \sigma_2^2$	$\frac{\sigma_2^2}{\sigma_2^2} > 1$	$F > f_{\alpha(\nu_1,\nu_2)}$
9	P=0.8	$P \neq 0.8$	$Z < -z_{\frac{\alpha}{2}}$ และ $Z > z_{\frac{\alpha}{2}}$
10	$\mu = 50$	$\mu \neq 50$	$T < -t_{\frac{\alpha}{2}(n-1)}$ และ $T > t_{\frac{\alpha}{2}(n-1)}$
11	$\sigma^2 = 2$	$\sigma^2 \neq 2$	$\chi^2 < x^2_{(1-\frac{\alpha}{2})(\nu)}$ และ $\chi^2 > x^2_{\frac{\alpha}{2}(\nu)}$
12	$\sigma_1^2 = \sigma_2^2$	$\frac{\sigma_1^2}{\sigma_2^2} \neq 1$	$F < f_{(1-\frac{\alpha}{2})(\nu_1,\nu_2)}$ และ $F > f_{\frac{\alpha}{2}(\nu_1,\nu_2)}$

ความคลาดเคลื่อนชนิดที่ 1 และความคลาดเคลื่อนชนิดที่ 2 (Type I Error and Type II Error)

ถ้าสมมติฐานว่างเป็นจริง แต่ข้อมูลที่ได้จากตัวอย่างคำนวณค่าสถิติตกอยู่ในช่วงเขตวิกฤต ทำให้เราปฏิเสธสมมติฐานที่เป็นจริง เรียกว่าเกิดความผิดชนิดที่ 1 ความน่าจะเป็นที่จะเกิดความผิดชนิดที่ 1 แทนด้วย α เรียกว่า ระดับนัยสำคัญ

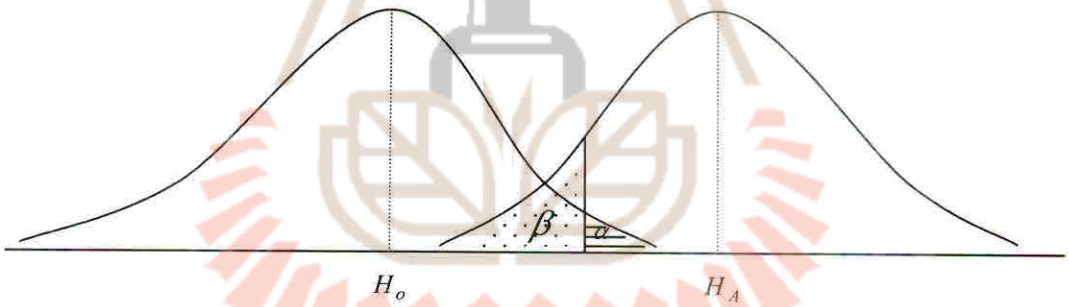
ถ้าสมมติฐานว่างไม่จริง แต่ข้อมูลที่ได้จากตัวอย่างคำนวณค่าสถิติตกอยู่ในช่วงยอมรับทำให้เรายอมรับสมมติฐานว่าง เรียกว่าเกิดความผิดชนิดที่ 2 ความน่าจะเป็นจะเกิดความผิดชนิดที่ 2 แทนด้วย β

ส่วน $1-\beta$ เรียกว่าเป็นกำลังของการทดสอบ (Power of The Test) ซึ่งพอจะสรุปความสัมพันธ์ดังตาราง

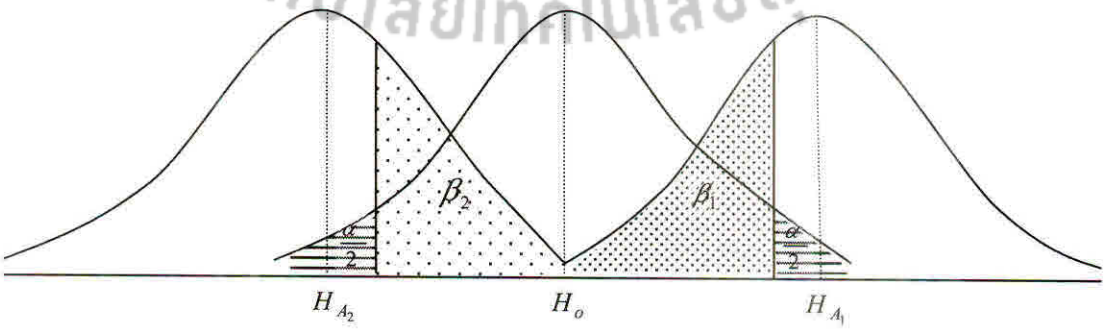
ข้อสรุปจากตัวอย่าง	สมมติฐานว่างเป็นจริง	สมมติฐานว่างไม่จริง
ปฏิเสธสมมติฐานว่าง	ความคลาดเคลื่อนชนิดที่ 1	ถูกต้อง
ยอมรับสมมติฐานว่าง	ถูกต้อง	ความคลาดเคลื่อนชนิดที่ 2

พิจารณา α และ β ได้ในแต่ละกรณี ดังนี้

ภาพที่ 8.13 **แสดง α และ β ในกรณีการทดสอบทางเดียวข้างน้อย**



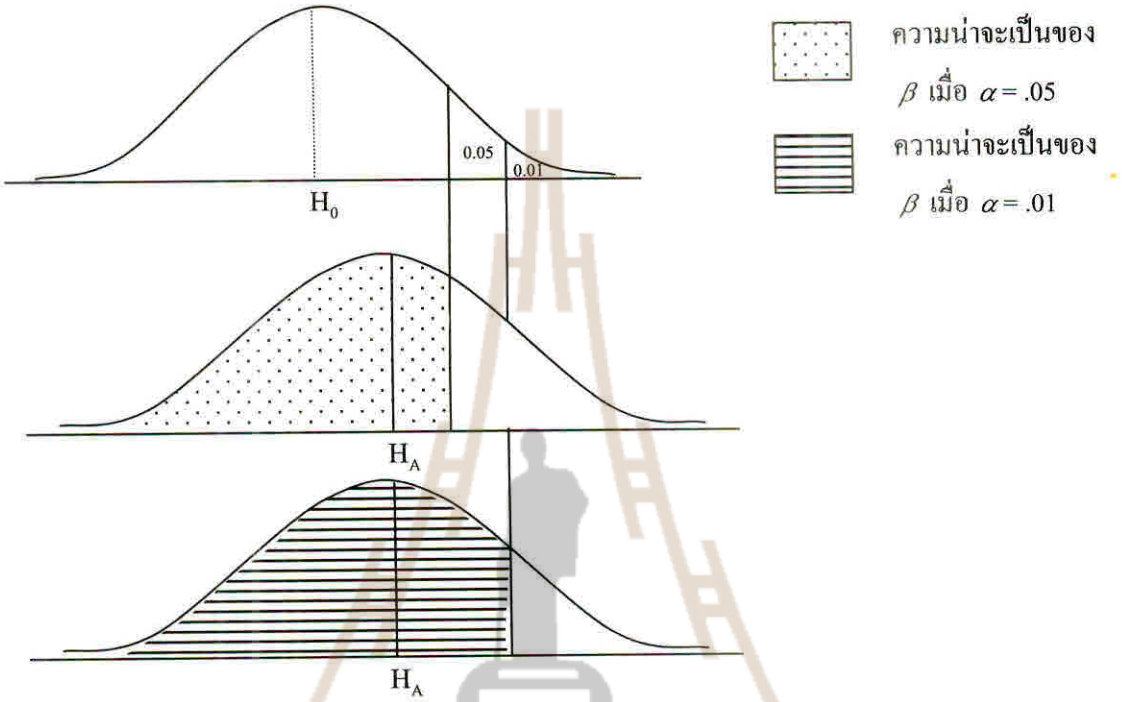
ภาพที่ 8.14 **แสดง α และ β ในกรณีการทดสอบทางเดียวข้างมาก**



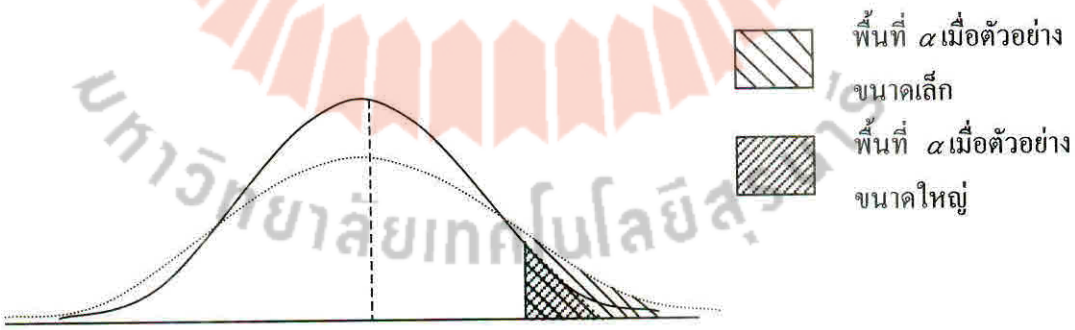
ภาพที่ 8.15 **แสดง α และ β ในกรณีการทดสอบสองทาง**

ความสัมพันธ์ของ α และ β

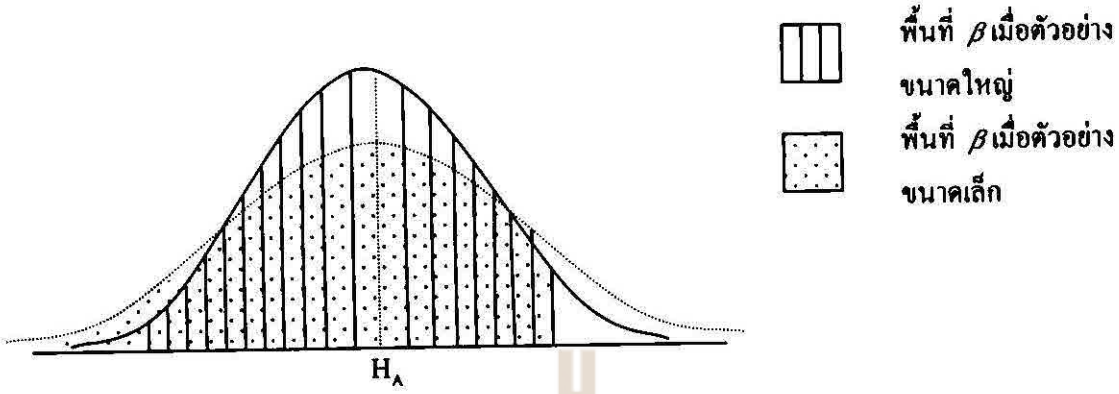
ทั้ง α และ β เป็นความน่าจะเป็นของความผิดพลาดที่เกิดขึ้น ดังนั้นในทางปฏิบัติเราต้องพยายามทำให้ทั้ง α และ β มีค่าน้อยที่สุด แต่โดยปกติ α และ β จะเป็นปฏิภาคผกผันกัน นั่นคือ ถ้าลด α ลง ค่า β ก็จะเพิ่มขึ้น หรือถ้าลดค่า β ค่า α ก็จะเพิ่มขึ้น ซึ่งพิจารณาได้จากภาพ



แต่เราสามารถที่จะลด α และ β พร้อม ๆ กันได้ โดยเพิ่มอนาคตของตัวอย่าง เมื่อตัวอย่างมีขนาดใหญ่ขึ้น ก็สามารถที่จะลด α และ β ลงได้ ซึ่งจะพิจารณาได้จากภาพ



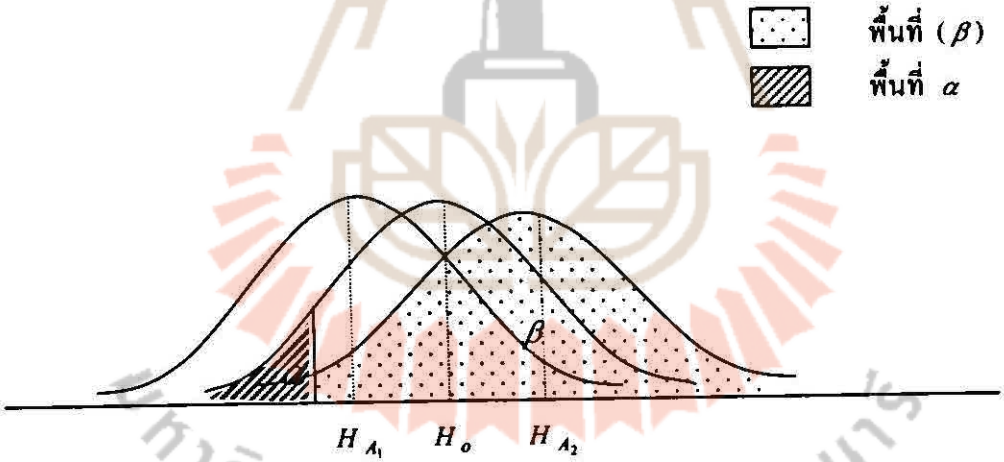
ภาพที่ 8.16 **แสดงขนาด ของ α เมื่อตัวอย่างขนาดเล็กและตัวอย่างขนาดใหญ่**



ภาพที่ 8.17 แสดงขนาดของ β เมื่อตัวอย่างขนาดเล็กและตัวอย่างขนาดใหญ่

ผลกระทบต่อกำลังของการทดสอบ (1- β)

กำลังของการทดสอบจะมีค่าน้อยมาก แสดงว่าความน่าจะเป็นของความผิดชนิดที่ 2 มีค่ามาก ในกรณีที่เราตั้งสมมติฐานการทดสอบทางเดียวได้ H_A ที่ถูกต้องจะมีค่าน้อยกว่า H_o พิจารณาได้จากภาพที่ผ่านมา สมมติฐานทางเลือกที่ถูกต้อง ก็คือ H_{A_2} แต่ เราไปตั้งเป็น H_{A_1}

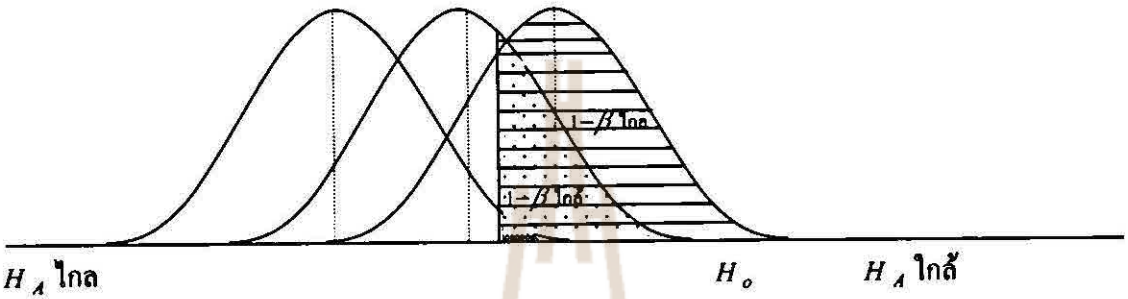


ภาพที่ 8.18 แสดง $1-\beta$ ในกรณีที่ตั้งสมมติฐานการทดสอบทางเดียวผิดข้าง จะเห็นว่าพื้นที่ $(1-\beta)$ ใน H_{A_2} ที่ถูกต้องนั้นมีค่าน้อยมาก และอาจจะเป็นศูนย์ ถ้า H_{A_2} มีค่าห่างจาก H_o พอสมควร

ในกรณีที่ตั้งสมมติฐานทางเลือกถูกต้องตามที่ควรจะเป็นค่าของกำลังการทดสอบ ก็จะมีค่ามากขึ้น ถ้า H_A อยู่ห่างจาก H_o มาก และจะน้อยลง ถ้า H_A มีค่าเข้าใกล้ H_o มาก พิจารณาได้จากภาพ

พื้นที่ $(1-\beta)$ เมื่อ H_A ใกล้ H_o

พื้นที่ $(1-\beta)$ เมื่อ H_A ไกล H_o



ภาพที่ 8.19 แสดงเปรียบเทียบ $(1-\beta)$ กรณีที่ H_A ใกล้ H_o กับ H_A ไกล H_o

ตัวอย่าง ในการสอบแบบเลือกตอบ 20 ข้อ แต่ละข้อมีคำตอบให้เลือก 4 หนทาง และมีคำตอบถูกเพียงคำตอบเดียว และถ้าผู้สอบทำคะแนนได้ตั้งแต่ 9 ข้อ แสดงว่ามีความรู้เกี่ยวกับข้อสอบ จงหา α และ β เมื่อสมมติฐานทางเลือก มีโอกาสที่จะตอบถูกร้อยละ 50

วิธีทำ (1) ตั้งสมมติฐานว่าง : ผู้สอบตอบโดยไม่มีความรู้ (เดาตอบ)

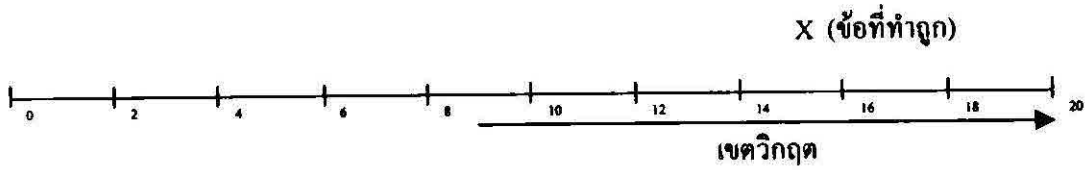
สมมติฐานทางเลือก : ผู้สอบมีความรู้ในเรื่องที่สอบ

ตั้งเป็นสัญลักษณ์

H_o : $P = \frac{1}{4}$

H_A : $P > \frac{1}{4}$

(2) เขตวิกฤตกำหนดโดย ถ้าตอบถูกตั้งแต่ 9 ข้อขึ้นไป แสดงว่ามีความรู้ในเรื่องที่สอบ



(3) α คือความน่าจะเป็นของความผิดปกติชนิดที่ 1

$$= P\{X > 9 \mid P = \frac{1}{4}\}$$

$$= \sum_{x=9}^{20} b(X,20,\frac{1}{4})$$

$$= \sum_{X=9}^{20}\binom{20}{x}\left(\frac{1}{4}\right)^{x}\left(\frac{3}{4}\right)^{20-x}$$

$$= 1 - \sum_{x=0}^{8}\binom{20}{X}\left(\frac{1}{4}\right)^{x}\left(\frac{3}{4}\right)^{20-x} = 0.0409$$

(4) β คือความน่าจะเป็นของความผิดชนิดที่ 2 เมื่อสมมติฐานทางเลือกคือ

$$P = \frac{1}{2}$$

$$\beta = P\{X < 9 \mid P = \frac{1}{2}\}$$

$$= \sum_{X=0}^{8} b\left(X;20,\frac{1}{2}\right)$$

$$= \sum_{X=0}^{8}\binom{20}{X}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{20-X}$$

$$= 0.2517$$ ตอบ

ในกรณีที่ตัวอย่างมีขนาดใหญ่ขึ้น เราสามารถพิจารณาขนาดของ α และ β ได้จาก ตัวอย่างต่อไปนี้

ตัวอย่าง จากโจทย์ตัวอย่างที่ผ่านมา แต่ทำการทดสอบ 100 ข้อ และถ้าทำข้อสอบถูกมากกว่า 37 ข้อ จะถือว่าปฏิเสธสมมติฐานว่าง

$$\alpha = \sum_{X=37}^{100}\binom{100}{X}\left(\frac{1}{4}\right)^{X}\left(\frac{3}{4}\right)^{100-X}$$

ประมาณโดยโค้งปกติมาตรฐานโดย $\mu = np = 100\left(\frac{1}{4}\right) = 25$

$$\sigma = \sqrt{npq} = \sqrt{100.\frac{1}{4}.\frac{3}{4}} = 4.33$$

$$\alpha = p\{X > 36.5 \mid N(X;25,4.33)\}$$

$$= p\{Z > 2.6559\}$$

$$= 0.0039$$ ตอบ

$$\beta = \sum_{X=0}^{36}\binom{100}{X}\left(\frac{1}{2}\right)^{X}\left(\frac{1}{2}\right)^{100-X}$$

ประมาณโดยโค้งปกติมาตรฐานโดย $\mu = 100\left(\frac{1}{2}\right) = 50$

$$\sigma = \sqrt{100.\frac{1}{2}.\frac{1}{2}} = 5$$

$$\beta = p\{X < 36.5 \mid N(X;50,5)\}$$

$$\beta = p\{Z < -2.7\}$$

$$= 0.0035$$ ตอบ

ตัวอย่าง มีผู้ตั้งข้อสังเกตว่าความสูงของนักศึกษาปีที่ 1 ในปัจจุบัน มีความสูงเฉลี่ยไม่เท่ากับ 68 นิ้ว ดังที่เคยบันทึกเอาไว้ เมื่อ 5 ปีที่แล้ว ถ้ากำหนดให้ขอบเขตวิกฤตที่จะปฏิเสธความสูงเฉลี่ยเดิม ถ้าค่าเฉลี่ยของความสูงในปัจจุบันน้อยกว่า 67 นิ้ว หรือมากกว่า 69 นิ้ว และทำการศึกษาโดยสุ่มตัวอย่างนักศึกษามา 36 คน คำนวณหาค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3.6 นิ้ว จงคำนวณหาค่า α และ β เมื่อสมมติฐานทางเลือก = 70 นิ้ว

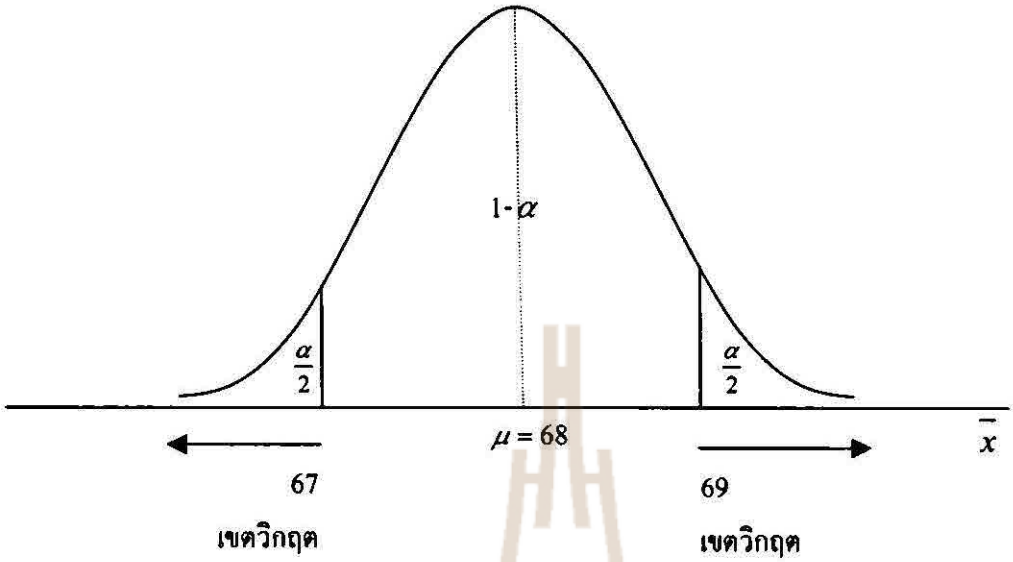
วิธีทำ (1) ตั้งสมมติฐาน

H_o : $\mu = 68$

H_A : $\mu \neq 68$

(2) กำหนดขอบเขตวิกฤต $\overline{X} < 67$ และ $\overline{X} > 69$ เนื่องจากเป็นตัวแปรเชิงสุ่มแบบต่อเนื่อง ที่มีการแจกแจงแบบปกติ ดังนั้นสามารถพิจารณาได้จากภาพต่อไปนี้

มหาวิทยาลัยเทคโนโลยีสุรนารี



(3) $= P\{\overline{X} < 67 \mid \mu = 68\} + P\{\overline{X} > 69 \mid \mu = 68\}$

แปลงเป็นค่าปกติมาตรฐาน

$$Z_{67} = \frac{67-68}{\frac{3.6}{\sqrt{36}}} = -1.67$$

$$Z_{69} = \frac{69-68}{\frac{3.6}{\sqrt{36}}} = 1.67$$

$$\alpha = P\{Z < -1.67\} + P\{Z > 1.67\}$$

$$= 0.0950$$ ตอบ

$$\beta = P\{67 \le \overline{X} \le 69 \mid \mu = 70\}$$

แปลงเป็นค่าปกติมาตรฐาน

$$Z_{67} = \frac{67-68}{\frac{3.6}{\sqrt{36}}} = -1.67$$

$$Z_{69} = \frac{69-68}{\frac{3.6}{\sqrt{36}}} = -1.67$$

$$\alpha = P\{Z < -1.67\} + P\{Z > 1.67\}$$

$$= 0.0950$$ ตอบ

มหาวิทยาลัยเทคโนโลยีสุรนารี

$$Z_{67} = \frac{67-70}{\frac{3.6}{\sqrt{36}}} = -5$$

$$Z_{69} = \frac{69-70}{\frac{3.6}{\sqrt{36}}} = 1.6667$$

$$\beta = P\{-5 \le Z \le -1.6667\}$$

$$= 0.0475$$ ตอบ

ตัวอย่าง จากตัวอย่างที่ผ่านมา ถ้าสุ่มตัวอย่างนักศึกษามา 64 คน จงหาค่าของ α และ β เมื่อสมมติฐานทางเลือก $\mu = 70$

วิธีทำ $\alpha = P\{\overline{X} < 67 \mid \mu = 68\} + P\{\overline{X} > 69 \mid \mu = 68\}$

แปลงเป็นค่าปกติมาตรฐาน

$$Z_{67} = \frac{67-68}{\frac{3.6}{\sqrt{64}}} = -2.22$$

$$Z_{69} = \frac{69-68}{\frac{3.6}{\sqrt{64}}} = 2.22$$

$$\alpha = P\{Z < -2.22\} + P\{Z > 2.22\}$$

$$= 0.0264$$ ตอบ

$$\beta = P\{67 \le \overline{X} \le 69 \mid \mu = 70\}$$

แปลงเป็นค่าปกติมาตรฐาน

$$Z_{67} = \frac{67-70}{\frac{3.6}{\sqrt{64}}} = -6.67$$

$$Z_{69} = \frac{69-70}{\frac{3.6}{\sqrt{64}}} = -2.22$$

$$\beta = P\{06.67 \le Z \le -2.22\}$$

$$= 0.0132$$ ตอบ

มหาวิทยาลัยเทคโนโลยีสุรนารี

ดังนั้นพอจะสรุปความองค์ประกอบ ที่มีผลกระทบต่อ α และ β ดังต่อไปนี้

1. ถ้า α มีค่ามากขึ้น ค่า β จะลดลง

   ถ้า β มีค่ามากขึ้น ค่า α จะลดลง

2. ถ้า α ลดลงจะทำให้ขอบเขตวิกฤตลดลงด้วย แต่ขอบเขตการยอมรับจะเพิ่มขึ้น

3. ถ้าขนาดของตัวอย่าง (n) มากขึ้น ค่า α และ β จะลดลง

4. ถ้าสมมติฐานว่างไม่เป็นจริง ค่า β จะมีค่ามากขึ้น เมื่อสมมติฐานทางเลือกมีค่าใกล้สมมติฐานว่าง ถ้าสมมติฐานว่างทางเลือกไกลสมมติฐานว่างเท่าใด ค่า β ก็ยิ่งจะลดลงเท่านั้น

สรุป

มหาวิทยาลัยเทคโนโลยีสุรนารี

กิจกรรมต่อเนื่อง ตอนที่ 8.5

1. ศึกษาผ่านสื่อปฏิสัมพันธ์ผ่านจอภาพ การบรรยาย หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วยโปรแกรมสำเร็จรูป SPSS ตอนที่ 8.5 การตีความจากผลการวิเคราะห์ข้อมูลด้วยสถิติ
2. ศึกษาด้วยตนเองผ่านสื่อบทเรียนคอมพิวเตอร์ช่วยสอน หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วย โปรแกรมสำเร็จรูป SPSS ตอนที่ 8.5 การตีความจากผลการวิเคราะห์ข้อมูลด้วยสถิติ
3. ทำกิจกรรมประกอบการเรียนในคู่มือการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วย โปรแกรมสำเร็จรูป SPSS ตอนที่ 8.5 การตีความจากผลการวิเคราะห์ข้อมูลด้วยสถิติ
4. ทำกิจกรรมเสริมประสบการณ์ในแผนกิจกรรมการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 8 การวิเคราะห์ข้อมูลทางสถิติด้วย โปรแกรมสำเร็จรูป SPSS ตอนที่ 8.5 การตีความจากผลการวิเคราะห์ข้อมูลด้วยสถิติ

มหาวิทยาลัยเทคโนโลยีสุรนารี

หน่วยที่

9

การนำเข้าข้อมูลจากรูปแบบอื่นๆ

มหาวิทยาลัยเทคโนโลยีสุรนารี

โดย รองศาสตราจารย์ ดร.คณิต ไข่มุกด์

ตอนที่

9.1

การอ่านข้อมูลจากแฟ้มข้อมูลแบบอื่นๆ

ในการวิเคราะห์ข้อมูลด้วยโปรแกรม SPSS เราสามารถนำเข้าข้อมูลจากแฟ้มข้อมูลรูปแบบอื่นๆ ได้แก่

(1) ข้อมูลที่อยู่แล้วในรูปแบบ text หรือ ASCII File

(2) ข้อมูลจากฐานข้อมูลสำเร็จรูป เช่น dBASE, FoxPro และ Access และ

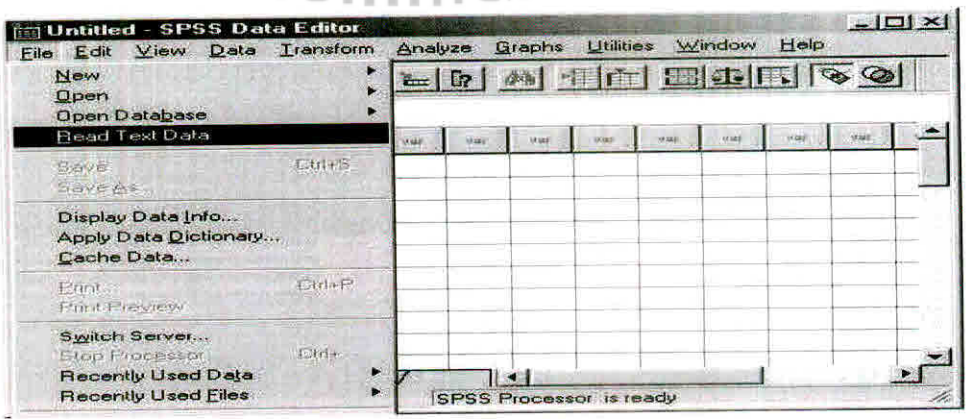
(3) ข้อมูลจากตารางคำนวณ เช่น Microsoft Excel

1. อ่านข้อมูลจาก ASCII flile

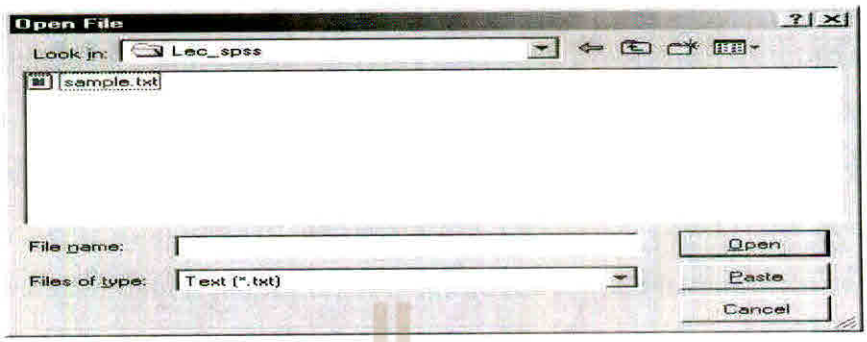
การเก็บข้อมูลที่เป็นรูปแบบมาตรฐานอย่างหนึ่งในการเก็บข้อมูลก็คือเก็บแบบข้อความรหัสแอสกี้ ที่นิยมเรียกทับศัพท์กันว่า แบบเท็กซ์ เพราะว่าประหยัดเนื้อที่ในการเก็บข้อมูล ยิ่งข้อมูลปริมาณมากก็จะยิ่งประหยัด และยังเป็นรหัสมาตรฐานที่ใช้ในการกำหนดอักขรไม่ว่าเป็นตัวเลขหรือตัวอักษรที่ใช้กันทั่วโลก ดังนั้นซอฟท์แวร์ทั่วๆ ไป ไม่ว่าจะเป็นตระกูลพิมพ์เอกสาร ตระกูลตารางคำนวณ ตระกูลฐานข้อมูล หรือตระกูลวิเคราะห์ข้อมูลทางสถิติ ก็มักจะมีทางเลือกในการเก็บข้อมูลแบบเท็กซ์อยู่ด้วยเสมอ ดังนั้นการเก็บข้อมูลแบบเท็กซ์จึงเป็นเสมือนสื่อกลางระหว่างซอฟท์แวร์ต่างตระกูลกัน เพื่อประโยชน์ในการใช้ข้อมูลร่วมกันโดยไม่ต้องป้อนข้อมูลเข้าไปใหม่

ซอฟท์แวร์ SPSS ก็เช่นกัน นอกจากจะเก็บข้อมูลในลักษณะแฟ้มข้อมูลแบบเท็กซ์แล้ว ก็สามารถรับข้อมูลแบบเท็กซ์ได้ด้วย

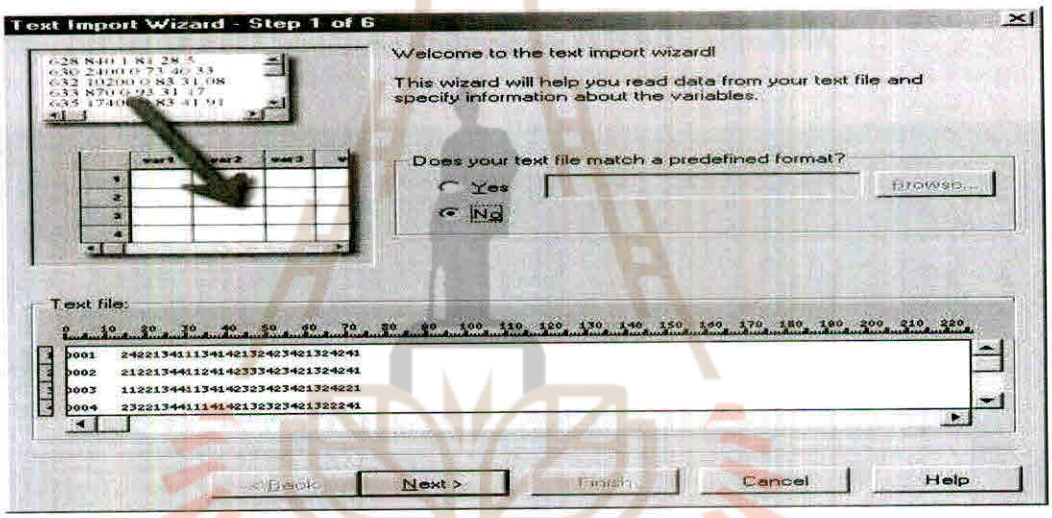
เริ่มจากหน้าต่างแรกของ SPSS ไปที่ File ⟶ Read Text Data จะได้หน้าต่างดังนี้



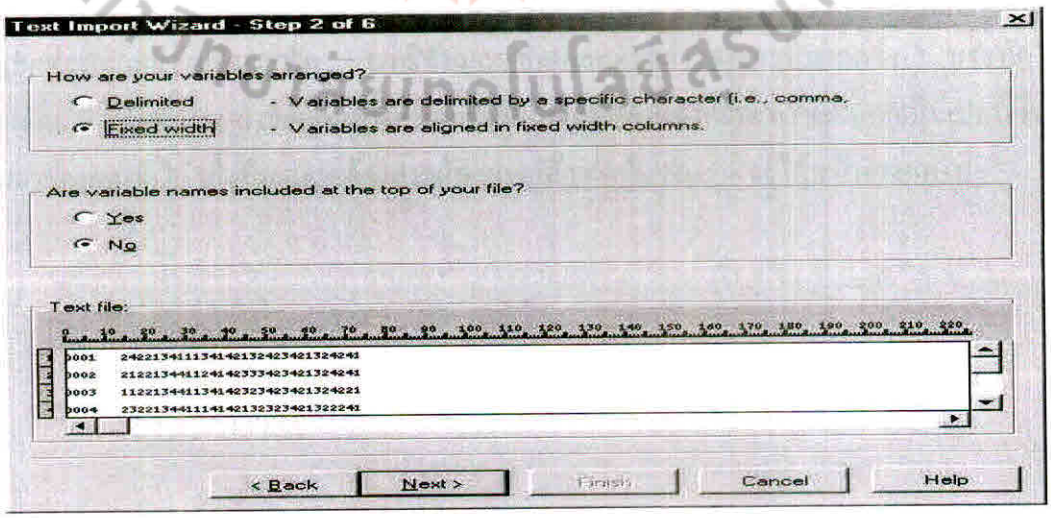
กดคลิ๊กที่ Read Text Data จะได้หน้าต่างถัดไปดังต่อไปนี้



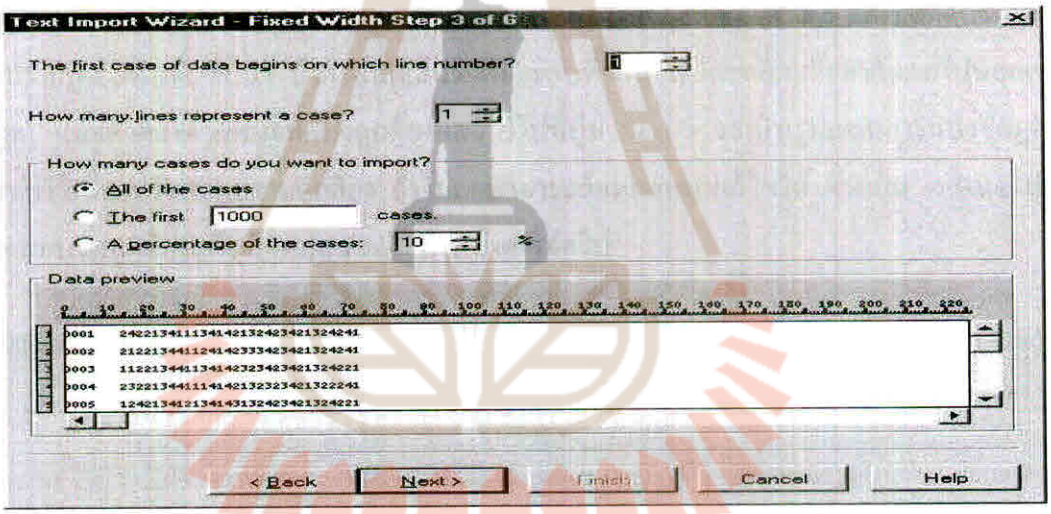
ระบุแฟ้มข้อมูลที่เป็นแฟ้มข้อมูลแบบเท็กซ์ ในที่นี้ก็คือ sample.txt แล้วกดปุ่ม Open จะได้หน้าต่างถัดไปคือ



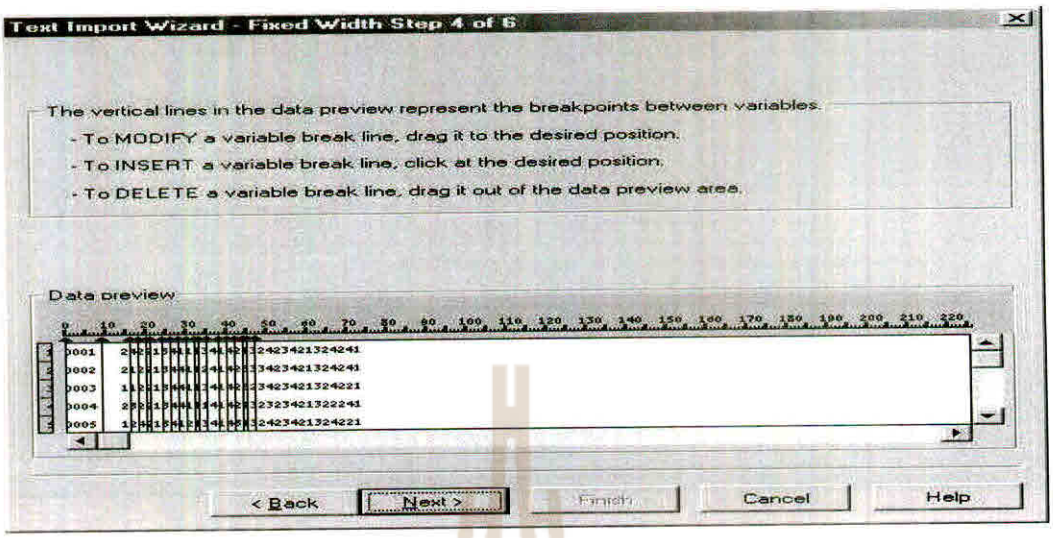
หน้าต่างนี้จะเห็นข้อมูลแบบเท็กซ์เพื่อเป็นการยืนยันว่าเป็นข้อมูลที่เราต้องการจะนำไปใช้ในโปรแกรม SPSS ถูกต้องใช่หรือไม่ ถ้าถูกต้อง ให้กดปุ่ม Next > จะได้หน้าต่างถัดไปคือ



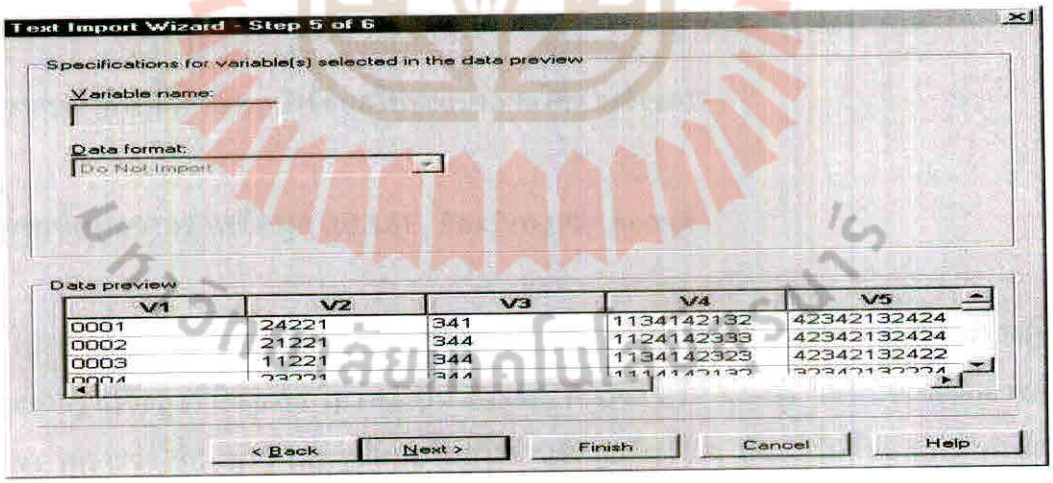
คำถามแรกในหน้าต่างนี้ก็คือ ข้อมูลที่ต้องการอ่านเข้ามาจัดการตัวแปรแบบใด ทางเลือกแรกก็คือ Delimited(เครื่องหมายคั่น) หมายความว่าข้อมูลของแต่ละตัวแปรมีเครื่องหมายคั่นเพื่อแยกตัวแปรอยู่ด้วยหรือไม่ ปกติถ้ามีเครื่องหมายคั่นข้อมูลอยู่แล้วก็จะง่ายในการจำแนกข้อมูลเพื่อนำเข้าสู่ตัวแปร เพราะตัวคั่นจะเป็นตัวบอกว่าข้อมูลใดเป็นค่าของตัวแปรใด ตัวคั่นอาจจะเป็น, หรือที่ว่างก็ได้ ปกติข้อมูลประเภทนี้มาจากโปรแกรมตารางคำนวณที่มีทางเลือกในการเก็บข้อมูลแต่ละสดมภ์ให้มีเครื่องหมายแยกกันชัดเจน แต่ถ้าเตรียมด้วย Text Editor โดยตรง เช่น NotePad หรือ WordPad ปกติจะเก็บข้อมูลติดต่อกันเพื่อประหยัดเนื้อที่ในการเก็บแต่จะแยกค่าตัวแปรด้วยสดมภ์ที่บันทึก จะต้องให้ข้อมูลของตัวแปรเดียวกันอยู่ตรงสดมภ์กัน จึงจะระบุข้อมูลได้ถูกต้อง กรณีนี้คือทางเลือกถัดมา แบบ Fixed Width จะใช้สดมภ์เป็นตัวระบุตัวแปรของข้อมูล ในตัวอย่าง sample.txt จะใช้ทางเลือกหลัง ส่วนกล่องถัดไปเป็นการระบุ ว่าข้อมูลที่เตรียมมาบรรทัดแรกเป็นชื่อของตัวแปรใช่หรือไม่ ทางเลือกนี้ก็เช่นกันปกติเป็นทางเลือกของข้อมูลที่เอามาจากโปรแกรมตระกูลตารางคำนวณ ในที่นี้เป็นข้อมูลล้วนเลือก No กดปุ่ม Next > จะได้หน้าต่างต่อไปคือ



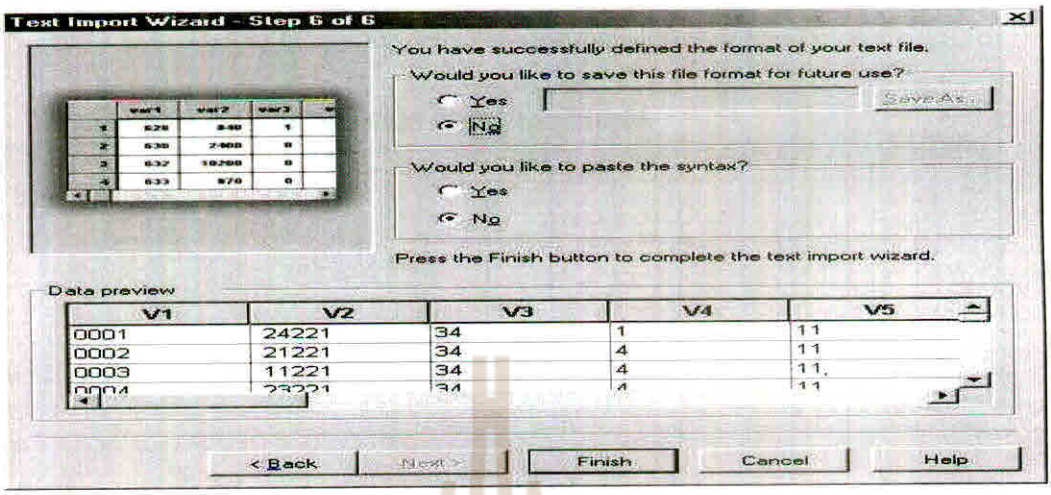
จากหน้าต่างแรกเป็นการยืนยันข้อมูลว่าตัวอย่างแรกข้อมูลเริ่มจากบรรทัดใด ช่องถัดมาถามว่าข้อมูลแต่ละตัวอย่างมีกี่บรรทัด กรณีข้อมูลมากแต่ละตัวอย่างอาจจะมีมากกว่า 1 บรรทัดได้ กล่องถัดมาให้เลือกว่าจะเอาข้อมูลมาทั้งหมดหรือไม่ หรือจะเอากี่ตัวอย่าง หรือจะบอกเป็นเปอร์เซ็นต์ที่จะเอาข้อมูลมาก็ได้ เมื่อข้อมูลถูกต้องตามต้องการแล้ว กดปุ่ม Next > จะได้หน้าต่างต่อไป



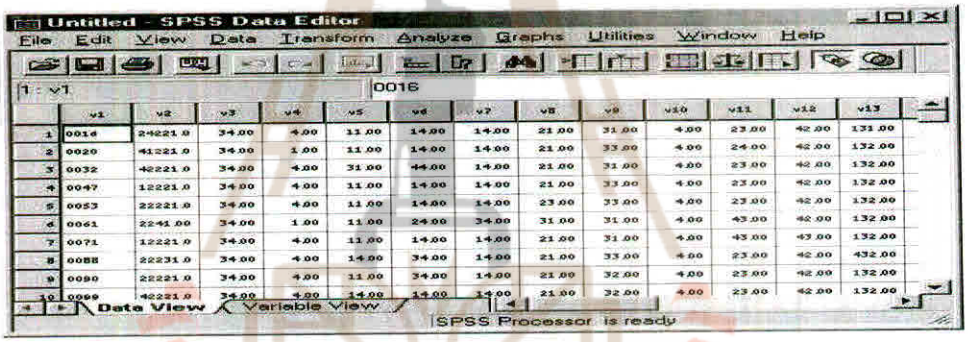
เราสามารถลากเส้นแนวดิ่งเพื่อแบ่งข้อมูลออกไปตามสดมภ์ของแต่ละตัวแปรตามที่ต้องการ โดยสามารถปรับเปลี่ยนตำแหน่งเส้นดิ่งได้จากการคลิ๊กเมาส์แล้วลากเส้นดิ่งไปในตำแหน่งที่ต้องการ และถ้าต้องการแทรกเส้นดิ่ง ณ ตำแหน่งใดๆ ก็สามารถคลิ๊กเมาส์ ณ ตำแหน่งที่ต้องการได้ ก็จะปรากฏเส้นดิ่งเพิ่มขึ้นมา ถ้ากำหนดเส้นดิ่งผิดจะลบทิ้งก็ทำได้โดยคลิ๊กแล้วลากออกไปนอกเขตข้อมูล เมื่อคาดว่าตำแหน่งที่แบ่งถูกต้องแล้ว ก็ไปที่ปุ่ม Next > จะเห็นรูปแบบการแบ่งข้อมูลที่กำหนดไว้เพื่อตรวจสอบความถูกต้อง ถ้าไม่ถูกสามารถย้อนกลับมาแก้ไขได้ ด้วยปุ่ม < Back เมื่อตรวจสอบจนถูกต้องแล้ว ก็กดปุ่ม Finish ดังหน้าต่างต่อไป



ในหน้าต่างนี้สามารถเลือกเฉพาะบางตัวแปรที่จะนำเข้ามาวิเคราะห์ได้อีก ถ้าไม่ระบุแสดงว่าเลือกมาทุกตัว เมื่อกดปุ่ม Next ก็จะได้หน้าต่างถัดไป



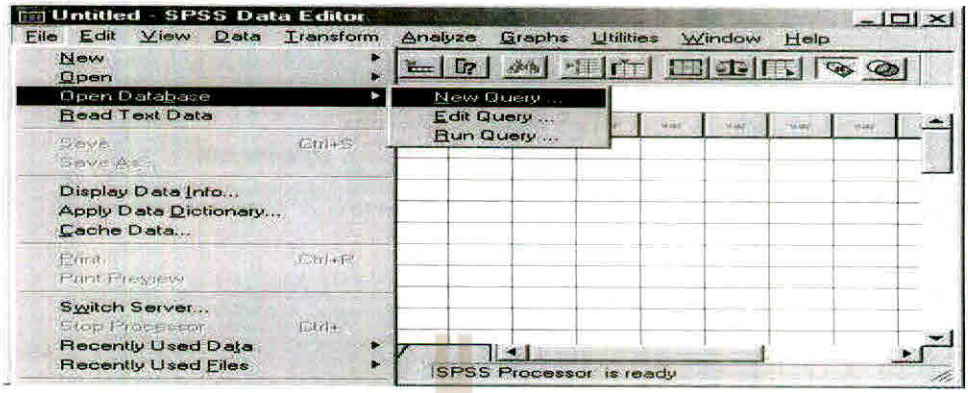
หน้าต่างนี้เป็นขั้นตอนสุดท้าย ถ้าต้องการจะบันทึกข้อมูลไว้ก็ให้ระบุชื่อ ถ้าไม่ก็กดปุ่ม Finish เป็นอันจบกระบวนการนำเข้าข้อมูล ได้ผลดังหน้าต่างต่อไปนี้



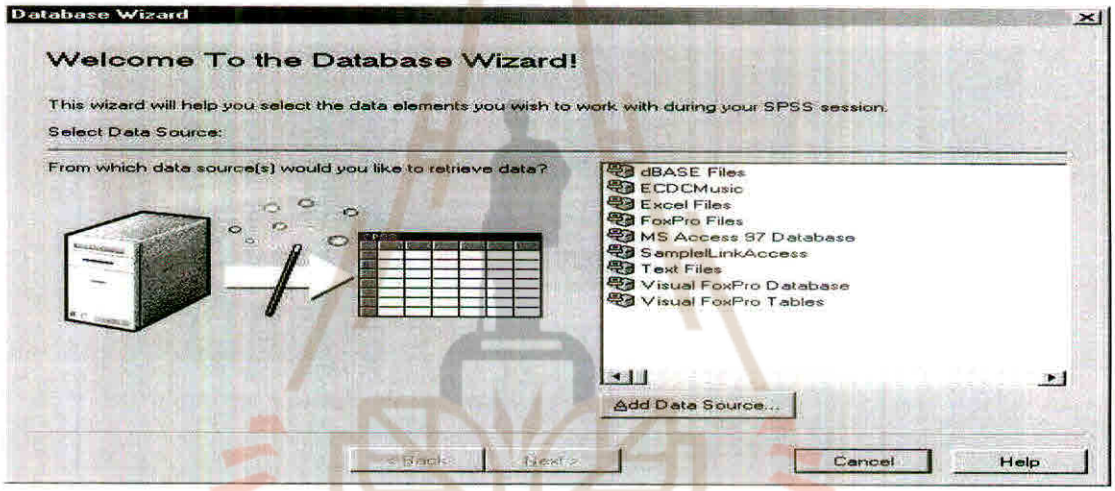
จากหน้าต่างนี้ก็สามารถนำข้อมูลไปประมวลผลได้ตามต้องการ

2. อ่านข้อมูลจากฐานข้อมูล dBASE, FoxPro และ Access

โปรแกรม SPSS ได้เตรียมการรับข้อมูลจากฐานข้อมูลโดยตรงได้ด้วยเพื่อความสะดวกของผู้ใช้ ข้อมูลจากฐานข้อมูลที่มีอยู่แล้ว ไม่ว่าจะเป็น dBASE, FoxPro หรือ Access ในที่นี้จะขอยกตัวอย่าง Access เพราะว่าใช้กันมาก และรูปแบบในการอ่านข้อมูลก็คล้ายๆ กัน เริ่มต้นจากหน้าต่างแรกต่อไปนี้

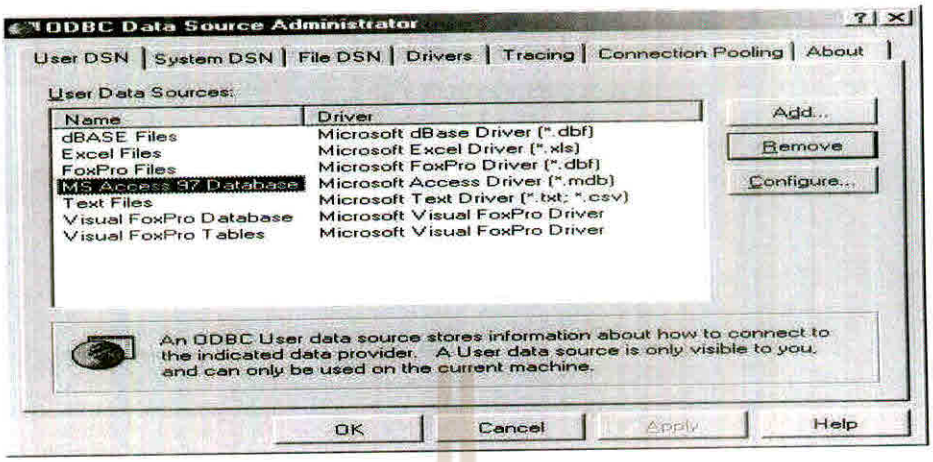


เลือก Open Database ⟶ New Query จะได้หน้าต่างต่อไป

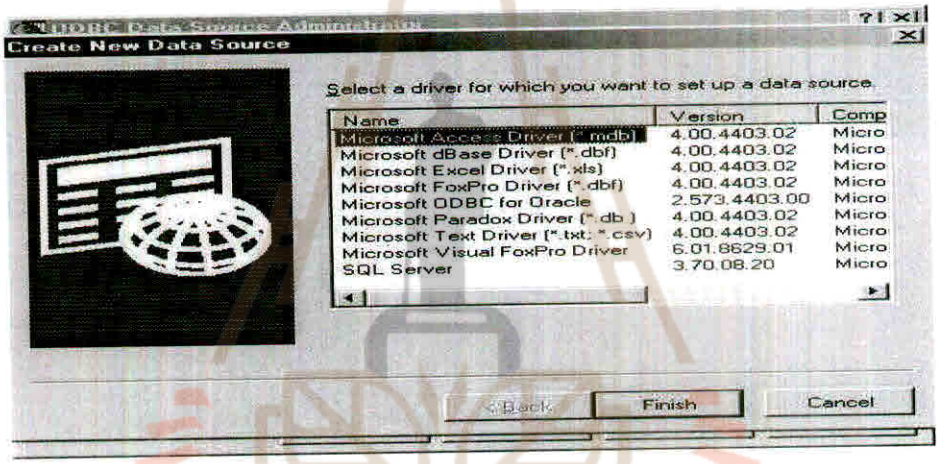


หน้าต่างนี้จะมีฐานข้อมูลให้เลือกเพื่อเชื่อมกับโปรแกรม SPSS ผ่าน ODBC เสียก่อน จึงจะเรียกใช้ข้อมูลได้ ในที่นี้เลือก MS Access 97 Database แล้วไปที่ปุ่ม Add Data Source เพื่อบอกว่าจะเชื่อมกับฐานข้อมูลชนิดนี้เข้ากับ SPSS เมื่อเลือกแล้วจะได้หน้าต่างถัดไป ดังนี้

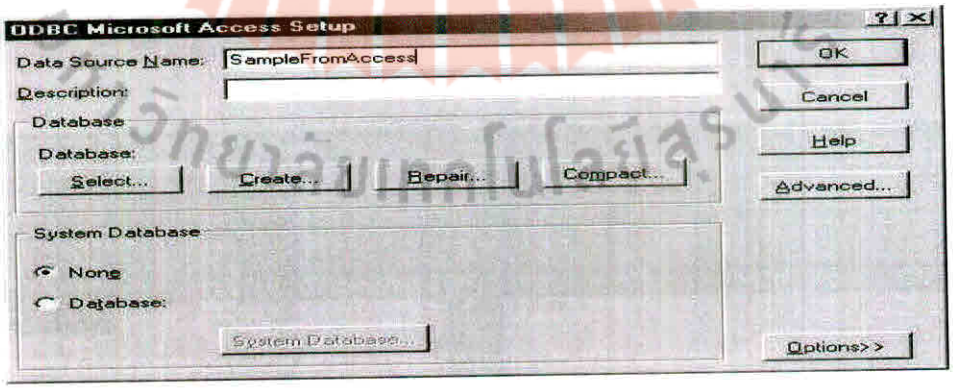
มหาวิทยาลัยเทคโนโลยีสุรนารี



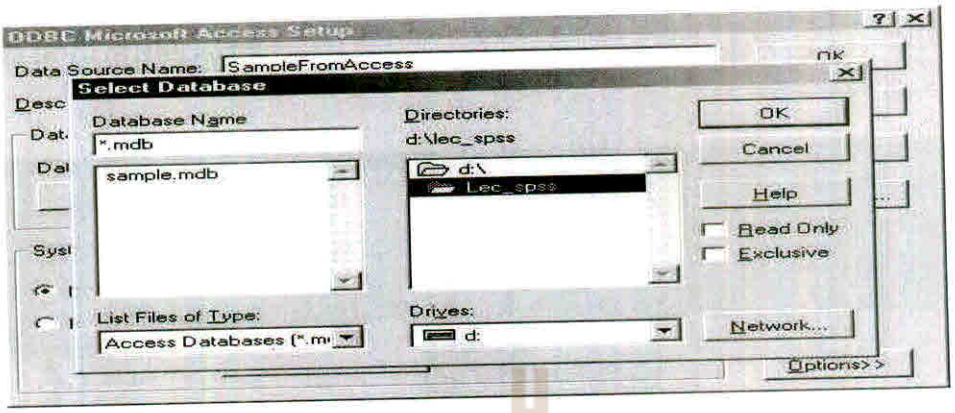
เลือกปุ่ม OK จะได้ หน้าต่างถัดไป



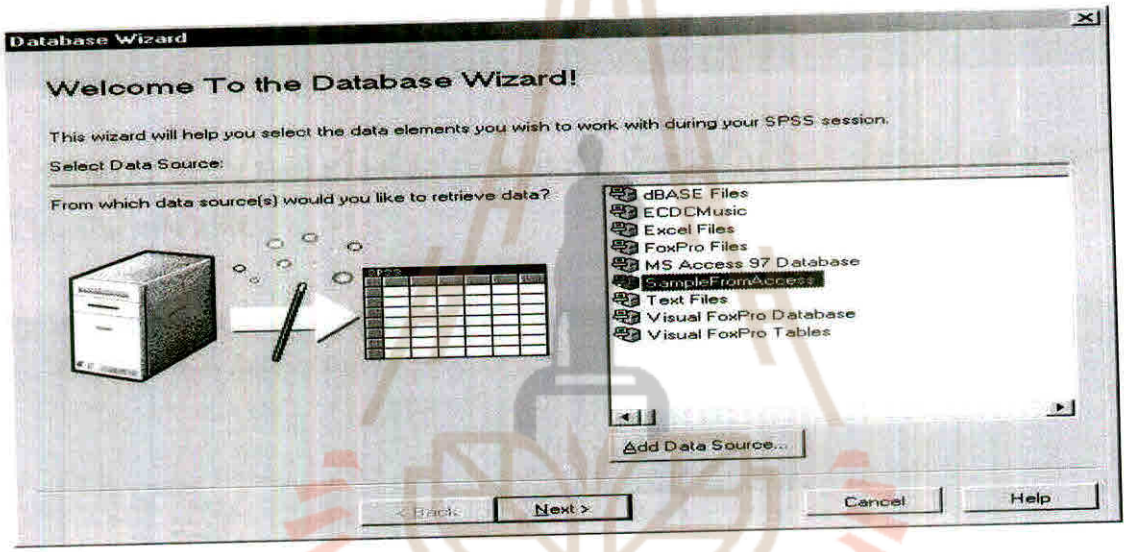
เลือก Microsoft Access Driver(*.mdb) แล้วกดปุ่ม Finish จะได้หน้าต่างต่อไปนี้



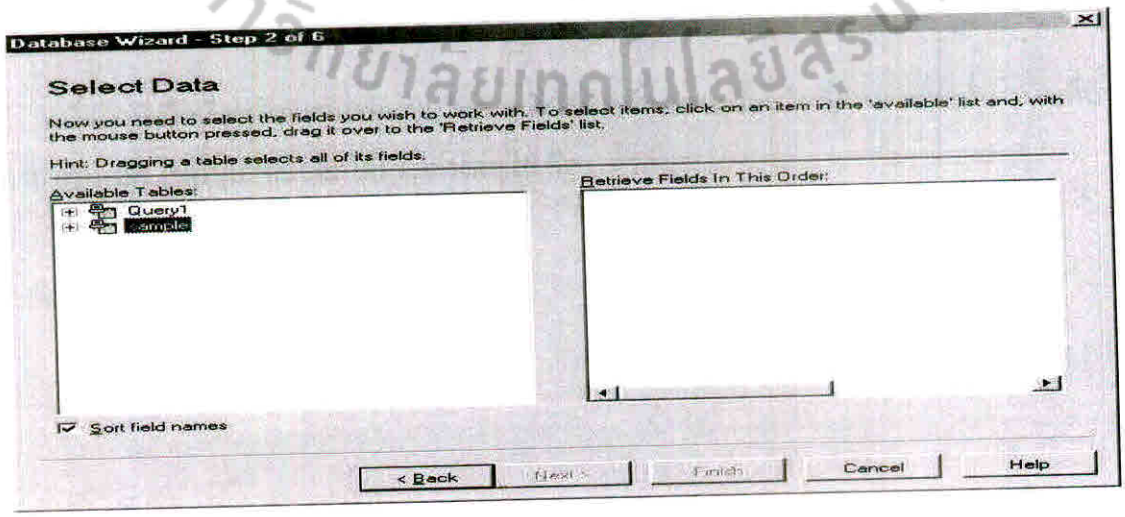
ระบุชื่อที่เป็นตัวเชื่อมนำข้อมูลจากฐานข้อมูลที่กำหนดให้ SPSS รู้จัก แล้วกดปุ่ม OK จะได้หน้าต่างถัดไปดังนี้



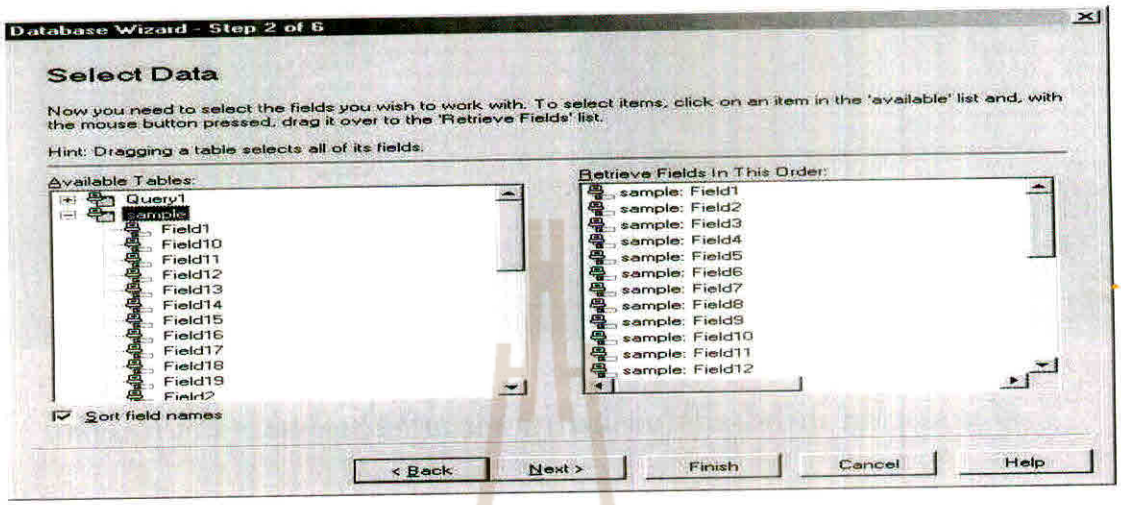
ระบุฐานข้อมูลที่จะนำมาเข้า SPSS ในที่นี้ชื่อ sample.mdb กดปุ่ม OK จะได้หน้าต่างต่อไป



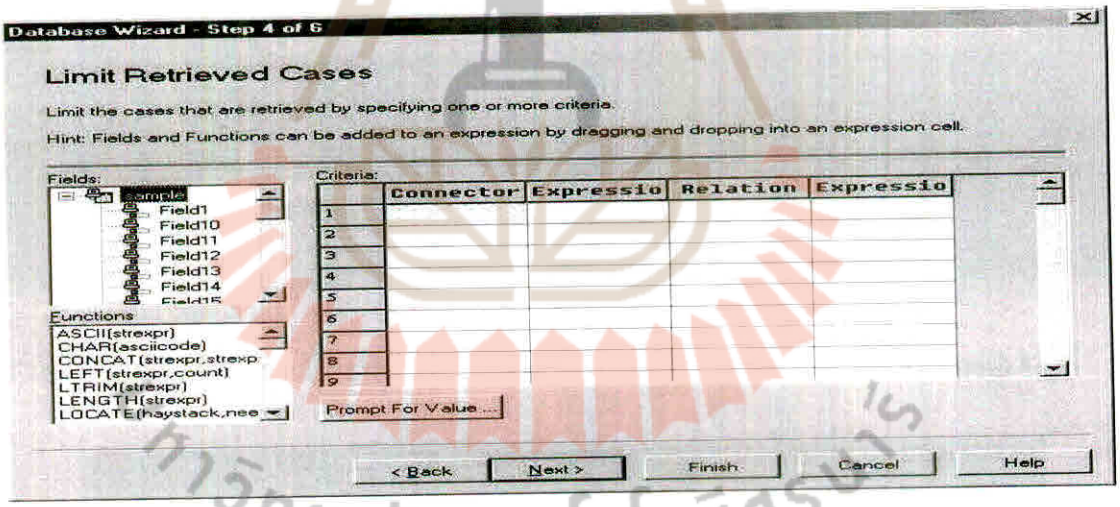
จะเห็นชื่อตัวเชื่อมที่เราตั้งไว้ในที่นี้ ชื่อว่า SampleFromAccess ก็แสดงว่าเราได้เตรียมการนำเข้าข้อมูลใน SPSS เรียบร้อยแล้ว กดปุ่ม Next จะได้



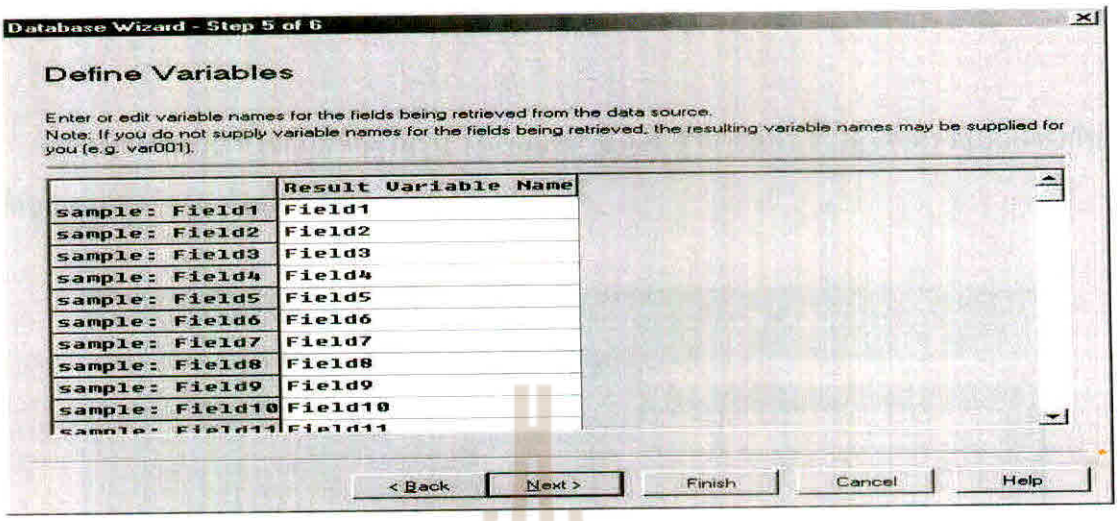
หน้าต่างนี้ ให้ระบุ Table หรือ Query ที่จะนำข้อมูลเข้ามาใน SPSS ในที่นี้ระบุ Table ชื่อ Sample จะได้หน้าต่างต่อไป คือ



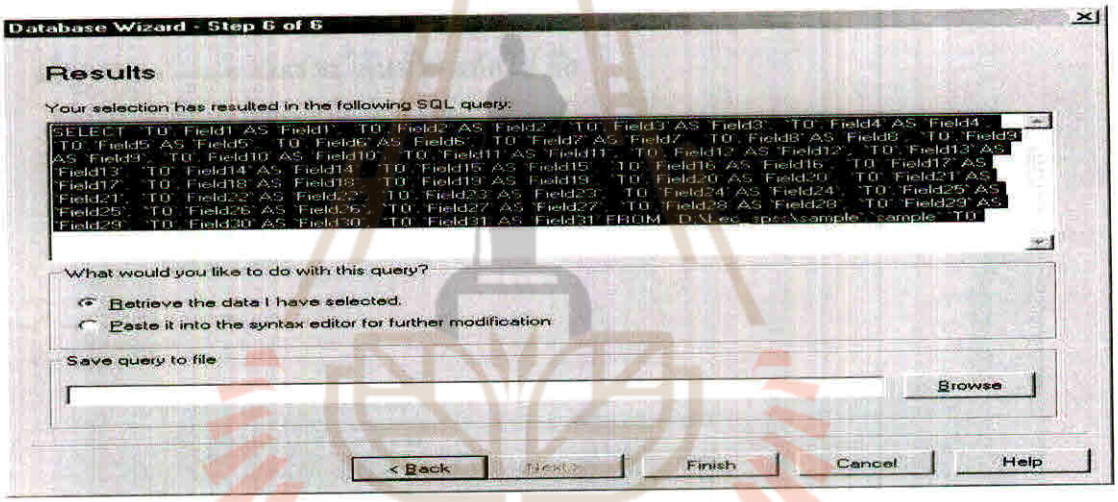
หน้าต่างนี้ ให้ระบุ Field หรือ ตัวแปรที่จะโอนเข้ามาวิเคราะห์ใน SPSS เลือกแล้ว กดปุ่ม Next จะได้หน้าต่างถัดไปคือ



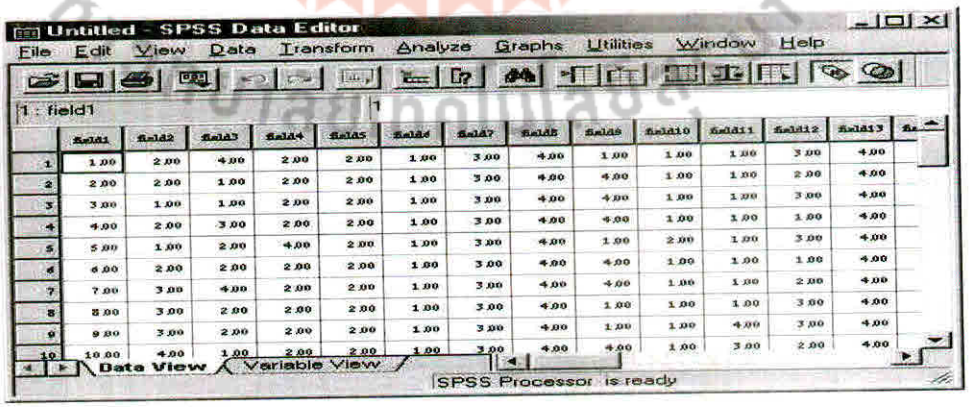
หน้าต่างนี้เป็นเงื่อนไขที่สามารถระบุในการนำข้อมูลเข้ามาตามเงื่อนไขที่กำหนดเท่านั้น ในที่นี้เอามาทั้งหมด ไปที่ปุ่ม Next จะได้หน้าต่างต่อไป คือ



ถ้าต้องการนิยามหรือเปลี่ยนชื่อตัวแปรก็สามารถเปลี่ยนได้ที่หน้าต่างนี้ ไปที่ Next จะได้



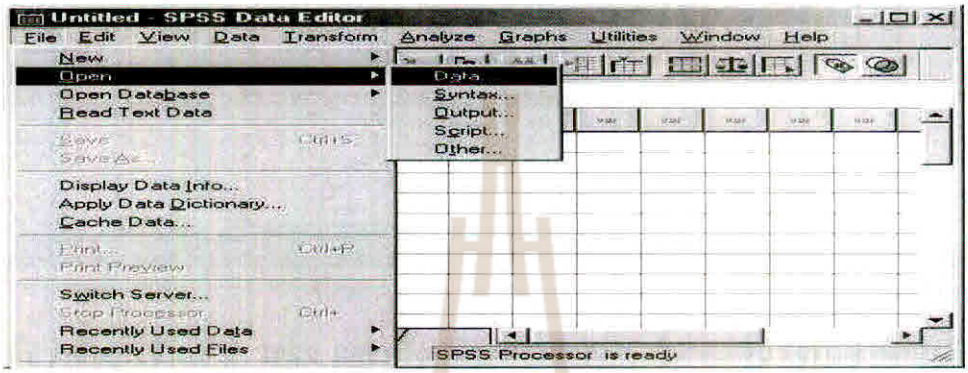
สรุปเราเลือกทั้งหมด แล้วจะบันทึกไว้หรือไม่ ถ้าบันทึกก็ให้ระบุชื่อ ถ้าไม่ก็ไปที่ Finish จะได้



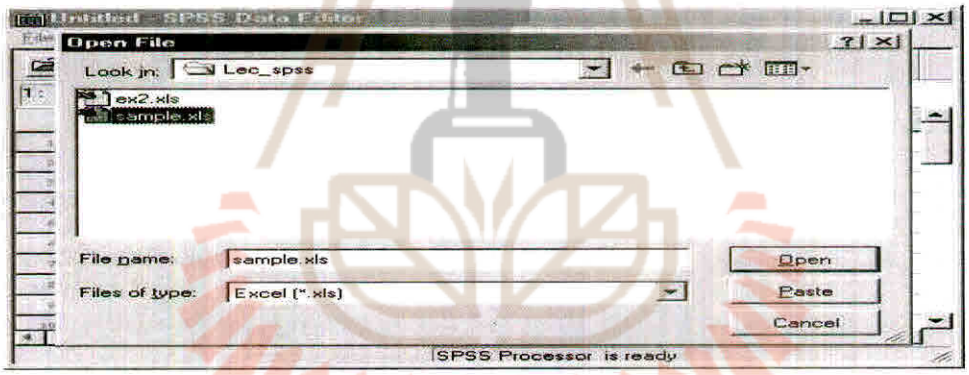
เป็นอันสิ้นสุดการทำงานนำข้อมูลที่อยู่ในฐานข้อมูลเข้า โปรแกรม SPSS

3. อ่านข้อมูลจากตารางคำนวณ Microsoft Excel

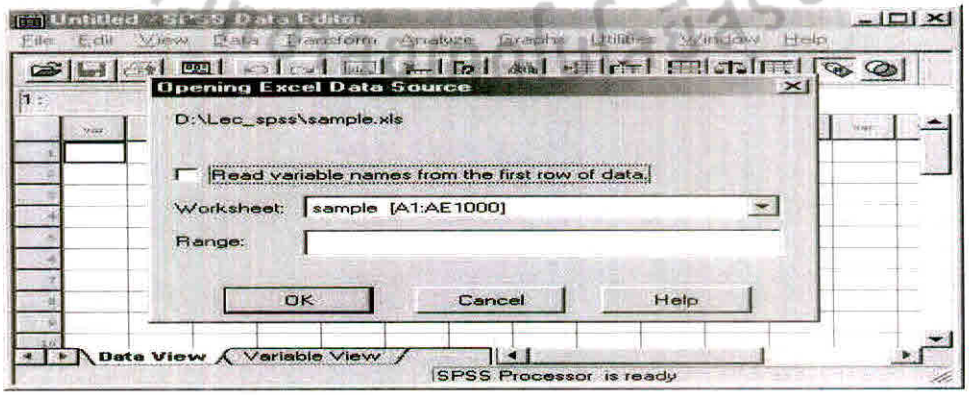
ข้อแนะนำในการนำข้อมูลที่อยู่ในรูป Microsoft Excel สามารถนำเข้าได้โดยตรงแบบเดียวกับข้อมูลแบบข้อความ ดังต่อไปนี้



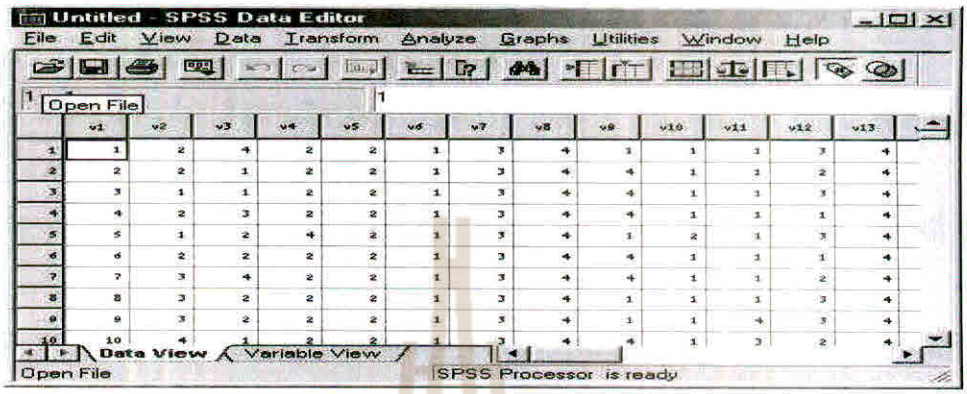
เลือก Open ⟶ Data จะได้หน้าต่างถัดไป คือ



ระบุชื่อแฟ้มข้อมูล Excel ในที่นี้ ชื่อ sample.xls แล้วไปที่ Open จะได้หน้าต่างถัดไป คือ



โปรแกรม SPSS ให้ระบุพิสัยของข้อมูลที่ต้องการนำมาใช้ในการวิเคราะห์กับ SPSS ว่าอยู่ในช่วงใด ในที่นี้ไม่ระบุ หมายความว่าเอาทั้งหมด ไปที่ OK จะได้



จะเห็นว่าข้อมูลเข้ามาใน SPSS และพร้อมที่จะประมวลผลแล้ว จะเห็นว่าการนำข้อมูลที่อยู่ในรูปแบบของตารางคำนวณ Excel ทำได้ง่ายที่สุด

สรุป

มหาวิทยาลัยเทคโนโลยีสุรนารี

กิจกรรมต่อเนื่อง ตอนที่ 9.1

1. ศึกษาผ่านสื่อปฏิสัมพันธ์ผ่านจอภาพ การบรรยาย หน่วยที่ 9 การนำเข้าข้อมูลจากรูปแบบอื่นๆ ตอนที่ 9.1 การอ่านข้อมูลจากแฟ้มข้อมูลแบบอื่นๆ
2. ศึกษาด้วยตนเองผ่านสื่อบทเรียนคอมพิวเตอร์ช่วยสอน หน่วยที่ 9 การนำเข้าข้อมูลจากรูปแบบอื่นๆ ตอนที่ 9.1 การอ่านข้อมูลจากแฟ้มข้อมูลแบบอื่นๆ
3. ทำกิจกรรมประกอบการเรียนในคู่มือการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 9 การนำเข้าข้อมูลจากรูปแบบอื่นๆ ตอนที่ 9.1 การอ่านข้อมูลจากแฟ้มข้อมูลแบบอื่นๆ
4. ทำกิจกรรมเสริมประสบการณ์ในแผนกิจกรรมการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 9 การนำเข้าข้อมูลจากรูปแบบอื่นๆ ตอนที่ 9.1 การอ่านข้อมูลจากแฟ้มข้อมูลแบบอื่นๆ

มหาวิทยาลัยเทคโนโลยีสุรนารี

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตอนที่

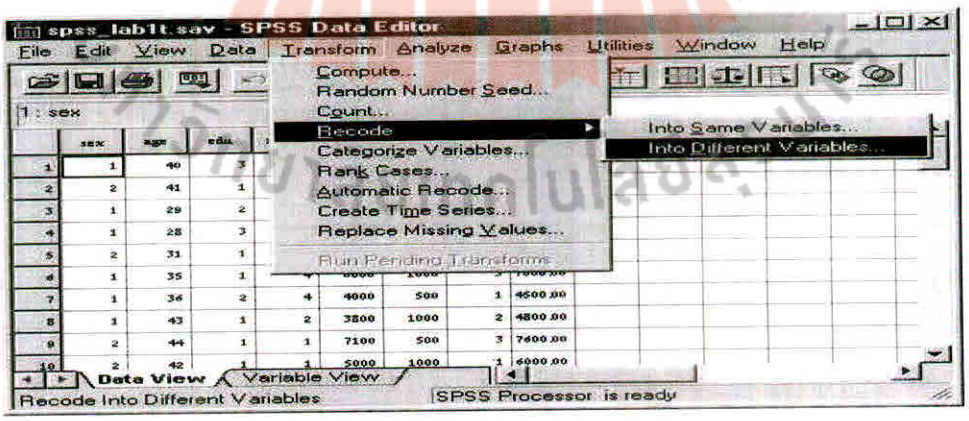
9.2

การปรับปรุงค่าข้อมูล

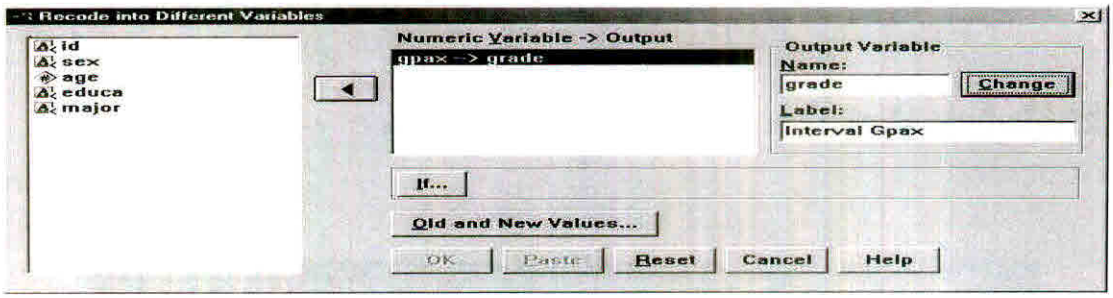
ข้อมูลที่มีการนำเข้าเพื่อวิเคราะห์ทางสถิติ ในบางครั้งเราจำเป็นต้องแปลงข้อมูลก่อนการวิเคราะห์ โดยจัดกลุ่มข้อมูลใหม่ด้วยคำสั่ง RECODE หรือคำนวณตัวแปรใหม่ ด้วยคำสั่ง COMPUTE

คำสั่ง RECODE และ คำสั่ง COMPUTE ใช้ในการปรับปรุงค่าของข้อมูลมีที่ใช้ต่างกันคือ คำสั่ง RECODE ใช้ในการจัดกลุ่มของข้อมูล เช่น จัดกลุ่มอายุของผู้ตอบแบบสอบถาม จัดกลุ่มของรายได้ของกลุ่มตัวอย่าง เพื่อใช้ในการวิเคราะห์ข้อมูล หรือใช้ในการเปลี่ยนค่าของข้อมูล เช่น ต้องการให้ค่าคะแนนของผู้ตอบข้อถามที่เป็นข้อถามทางลบมีค่าสวนทางกับคะแนนของผู้ตอบข้อถามที่มีข้อความมีความหมายทางบวก ส่วนคำสั่ง COMPUTE เป็นการคำนวณค่าตัวแปรใหม่จากค่าตัวแปรที่มี เช่น ต้องการทราบผลรวมของรายได้ทั้งหมด โดยมีรายได้จากเงินเดือนและรายได้จากการทำงานล่วงเวลา หรือต้องการหาผลบวกของคะแนนความคิดเห็นที่มีหลายข้อถาม เพื่อจะสรุปว่าคะแนนรวมเป็นเท่าใด หรือหาผลบวกของคะแนนสอบทั้งหมดจากที่มีคะแนนสอบย่อยหลายๆ ครั้ง

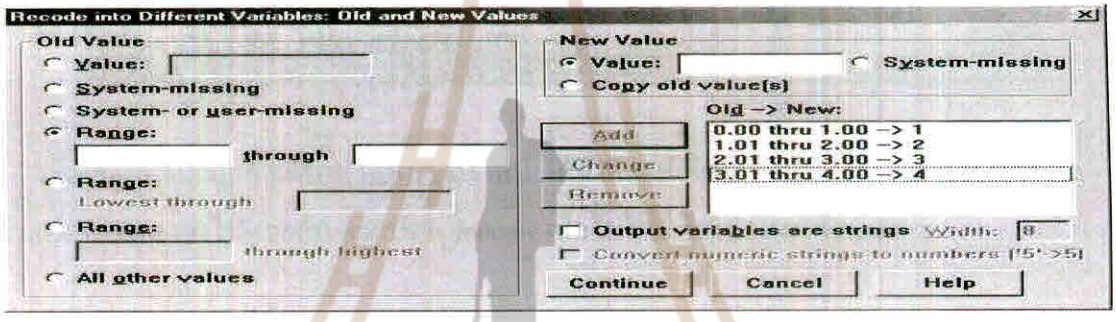
1. การแปลงข้อมูล (RECODE)



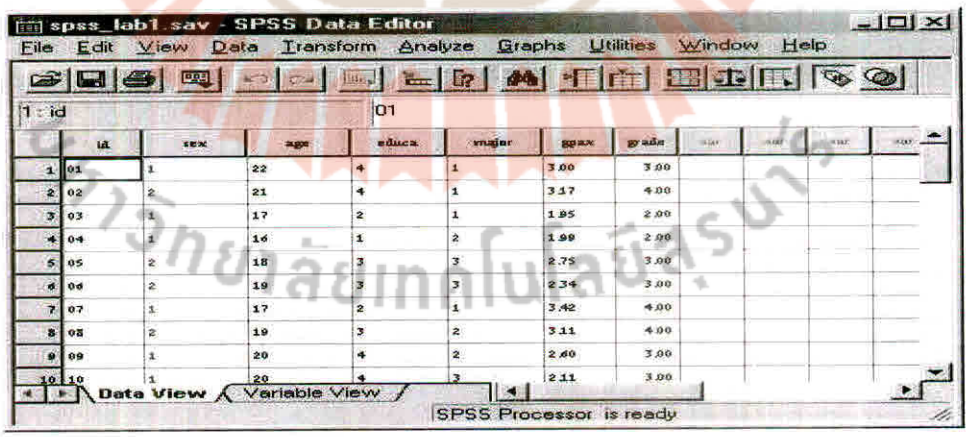
เลือกทางเลือกจากเมนูรายการดังนี้ Transform ⟶ Recode ⟶ Into Different Variables ดังแสดงในหน้าต่าง จะได้หน้าต่างต่อไปดังนี้



เลือก gpax มาทำการ RECODE ตั้งชื่อตัวแปรใหม่ว่า grade ชื่อตัวแปรนี้ว่า interval gpax กดปุ่ม Change เพื่อบอกให้เครื่องทราบชื่อตัวแปรใหม่ที่จะเอาค่าที่เปลี่ยนแปลงแล้วไปเก็บ แล้วเลือกปุ่ม Old and New Value เพื่อกำหนดค่าเดิมให้เป็นค่าใหม่ตามที่เราต้องการ จะได้หน้าต่างใหม่ดังนี้



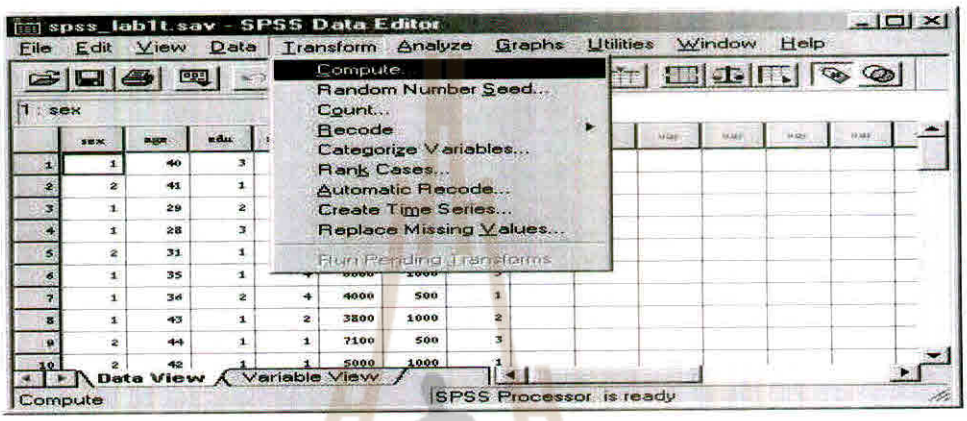
เมื่อกำหนดค่าเก่าเป็นช่วงแล้วบอกค่าใหม่ที่ต้องการให้เป็นกดปุ่ม Add จนได้ดังจอภาพแล้ว กดปุ่ม Continue จะได้หน้าต่างในขั้นต่อไป



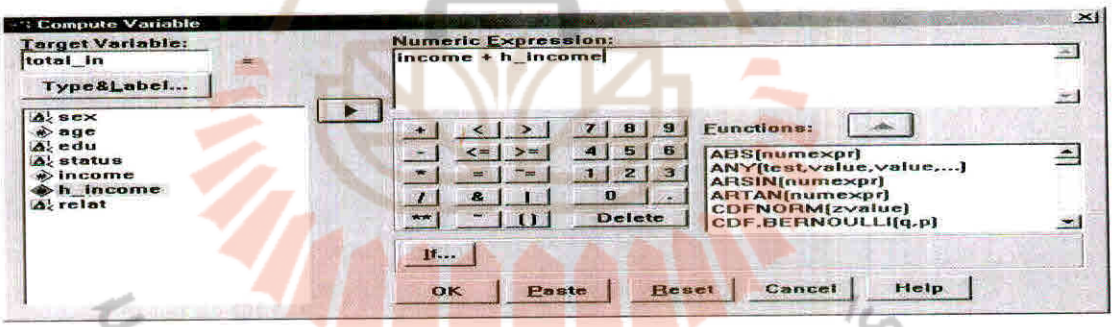
จะเห็นว่ามีตัวแปรสดมภ์ใหม่ ชื่อ grade เกิดขึ้น และค่าของข้อมูลก็เป็นค่าที่เปลี่ยนแปลงเกรดตามช่วงที่กำหนดให้เป็นค่าที่เรากำหนด เป็นอันเสร็จสิ้นการใช้ คำสั่ง RECODE ส่วนทางเลือกอื่นๆ สามารถจะทำได้ในทำนองเดียวกัน

2. การสร้างตัวแปรใหม่ (COMPUTE)

เลือก Transform ⟶ Compute… เพื่อคำนวณค่าผลรวมระหว่างเงินเดือนกับเงินที่ส่งให้ทางบ้าน

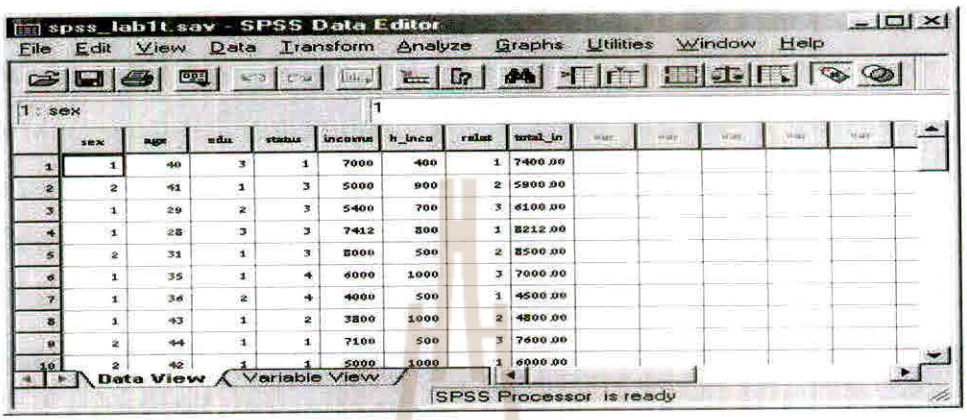


เมื่อเลือกทางเลือกข้างต้นจะได้หน้าต่างใหม่ ให้ตั้งตัวแปรใหม่ ในที่นี้ใช้ชื่อ total_in แล้วเลือกตัวแปร income เอาไปบวกกับ ตัวแปร h_income เพื่อไปใส่ที่ตัวแปรใหม่ คือ total_in



เมื่อตรวจสอบความถูกต้องแล้ว จากนั้นกดปุ่ม OK เพื่อสั่งให้โปรแกรมทำงาน ก็จะได้หน้าต่างสุดท้ายดังนี้

จะเห็นว่ามี สดมภ์เพิ่มขึ้นอีกสดมภ์และมีชื่อ total_in ค่าของข้อมูลเป็นผลบวกของสดมภ์ที่ชื่อ income และ h_income เป็นอันเสร็จสิ้นกระบวนการในการคำนวณข้อมูลใหม่ด้วยคำสั่ง COMPUTE



สรุป

มหาวิทยาลัยเทคโนโลยีสุรนารี

กิจกรรมต่อเนื่อง ตอนที่ 9.2

1. ศึกษาผ่านสื่อปฏิสัมพันธ์ผ่านจอภาพ การบรรยาย หน่วยที่ 9 การนำเข้าข้อมูลจากรูปแบบอื่นๆ ตอนที่ 9.2 การปรับปรุงค่าข้อมูล
2. ศึกษาด้วยตนเองผ่านสื่อบทเรียนคอมพิวเตอร์ช่วยสอน หน่วยที่ 9 การนำเข้าข้อมูลจากรูปแบบอื่นๆ ตอนที่ 9.2 การปรับปรุงค่าข้อมูล
3. ทำกิจกรรมประกอบการเรียนในคู่มือการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 9 การนำเข้าข้อมูลจากรูปแบบอื่นๆ ตอนที่ 9.2 การปรับปรุงค่าข้อมูล
4. ทำกิจกรรมเสริมประสบการณ์ในแผนกิจกรรมการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 9 การนำเข้าข้อมูลจากรูปแบบอื่นๆ 9.2 การปรับปรุงค่าข้อมูล

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตอนที่

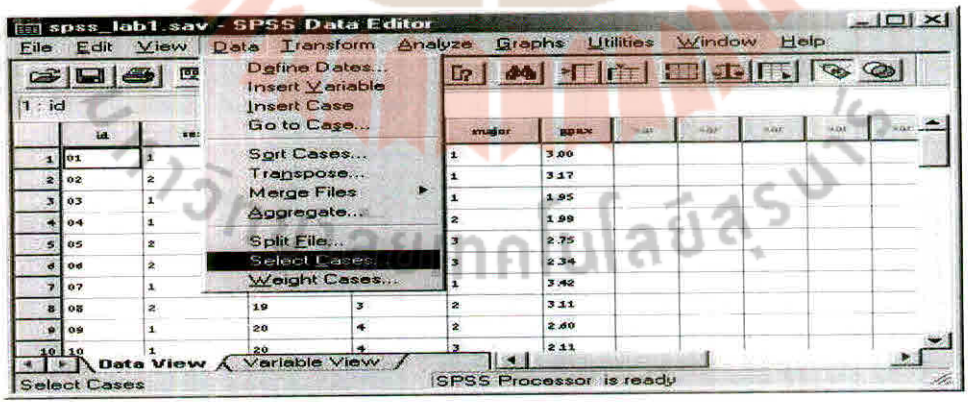
9.3

การเลือกข้อมูลมาวิเคราะห์

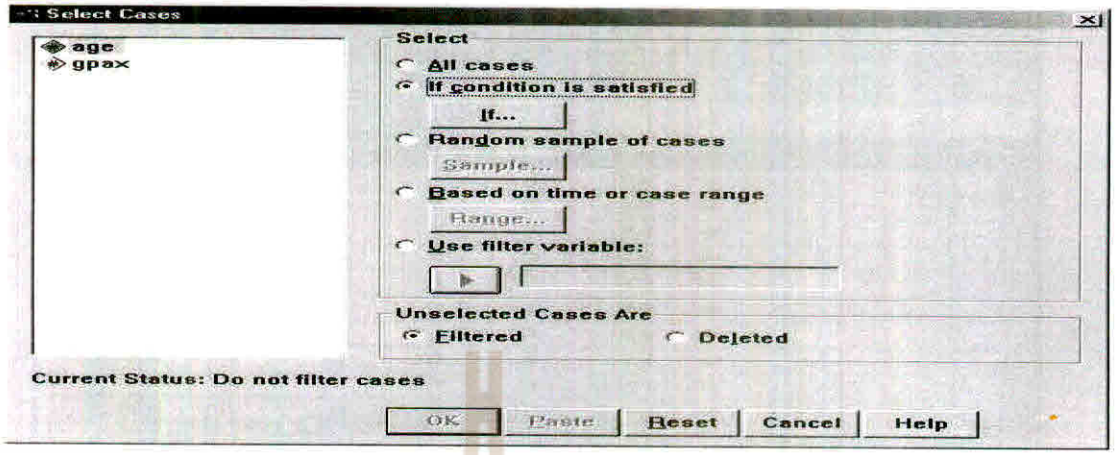
สามารถกำหนดเงื่อนไขที่ต้องการ ในการเลือกข้อมูลบางตัวมาวิเคราะห์ทางสถิติ โดยใช้คำสั่ง SELECT IF หรือคำสั่ง PROCESS IF

คำสั่ง SELECT IF และ คำสั่ง PROCESS IF มีที่ใช้ในกรณีต้องการเลือกบ้างข้อมูลเข้ามาทำการวิเคราะห์ เช่น ต้องการวิเคราะห์เฉพาะข้อมูลของเพศหญิง หรือข้อมูลของผู้ตอบที่มีอายุครบเกณฑ์ในการเลือกตั้ง หรือผู้ที่มี GPA ต่ำกว่า 2.00 เท่านั้น คำสั่ง SELECT IF ใช้ในกรณีต้องการเลือกข้อมูลเข้ามาทำการวิเคราะห์โดยกำหนดเงื่อนไขแบบถาวรตลอดโปรแกรม ส่วนคำสั่ง PROCESS IF ใช้ในกรณีต้องการเลือกข้อมูลมาทำการวิเคราะห์กับคำสั่งนั้นคำสั่งเดียว ดังนั้นสามารถที่จะเลือกข้อมูลด้วยเงื่อนไขที่แตกต่างกันในแต่ละคำสั่ง แม้ว่าเป็นโปรแกรมชุดเดียวกัน คำสั่ง PROCESS IF ได้นำมารวมกับคำสั่งทางสถิติที่กำหนดให้สามารถใช้เลือกเงื่อนไขของข้อมูลในการวิเคราะห์ได้เป็น ทางเลือก If ...

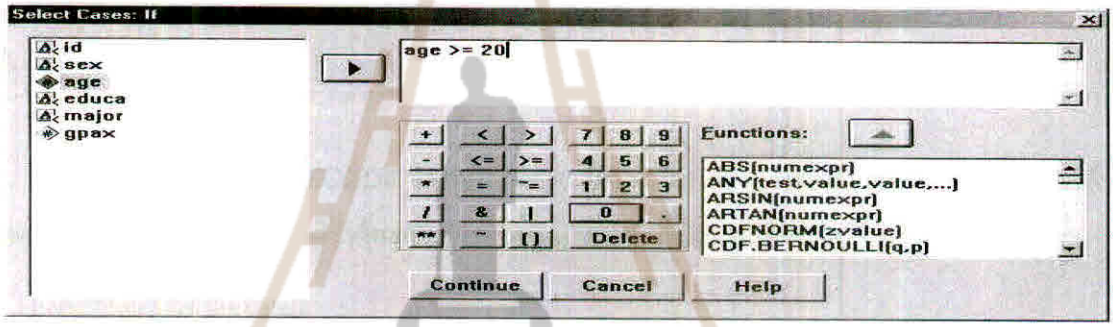
1. คำสั่ง SELECT IF



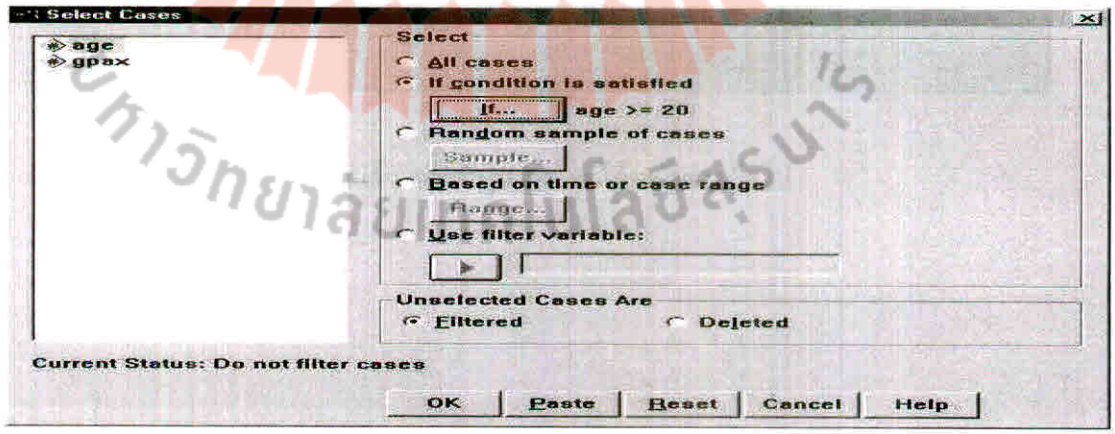
เลือกหน้าต่าง Data ⟶ Select Cases...



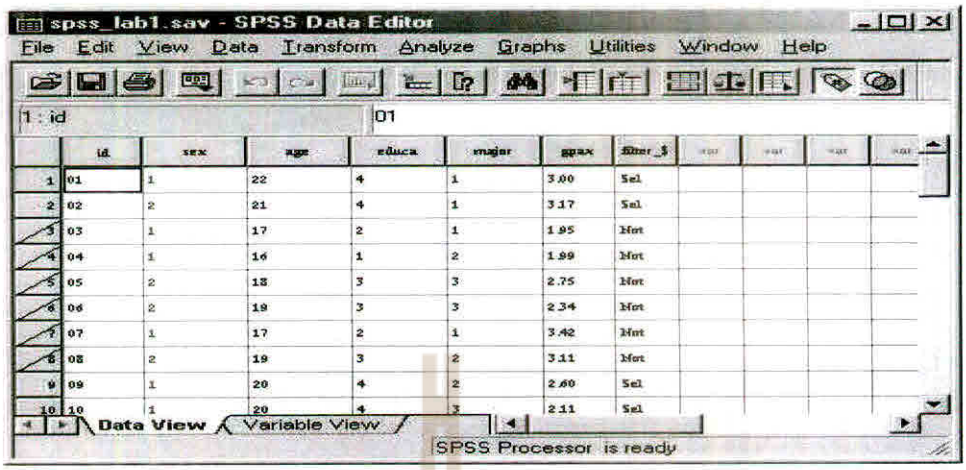
เลือกทางเลือก If condition is satisfied ⟶ If จะได้หน้าต่างถัดไปคือ



ลองเลือก age แล้วกำหนดว่า เลือกเฉพาะกรณี อายุมากกว่าหรือเท่ากับ 20 ปี เมื่อได้ข้อมูลดังหน้าต่างแล้ว กดปุ่ม Continue จะได้หน้าต่างถัดไป



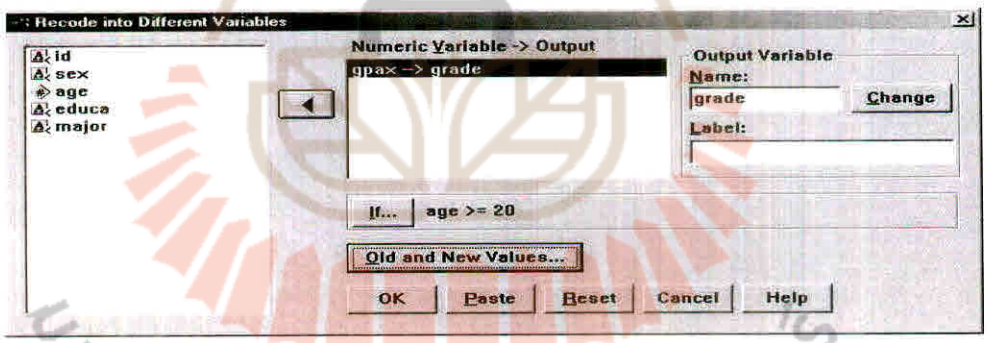
กดปุ่ม OK ก็จะได้หน้าต่างสุดท้ายดังต่อไปนี้



สังเกตเห็นว่าที่สดมภ์สุดท้าย จะมีข้อมูลระบุว่า case ใดบ้างถูกเลือกมาตามเงื่อนไขที่เรากำหนด

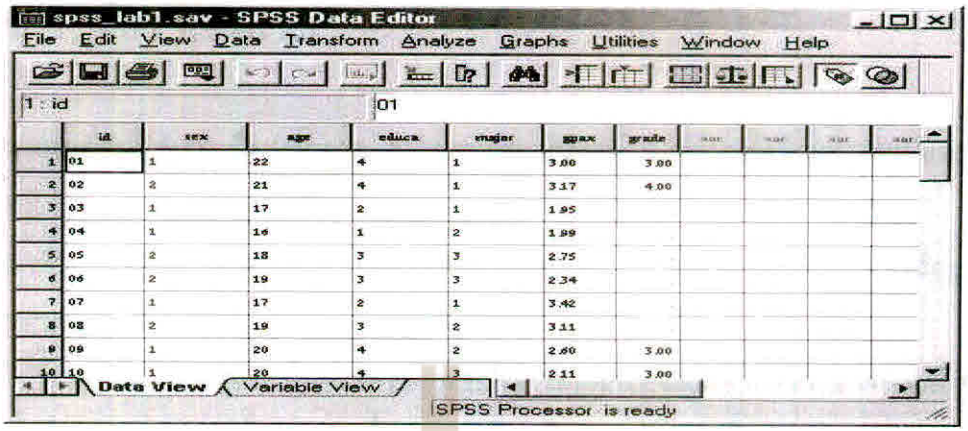
2. คำสั่ง PROCESS IF

อ้างถึงข้อมูลในคำสั่ง RECODE จัดกลุ่ม gpa จะมีทางเลือก If... ให้สามารถกำหนดเงื่อนไขได้ในการจัดกลุ่มเกรดนี้ได้ สมมติว่าต้องการจัดกลุ่ม gpa เฉพาะผู้ที่มีอายุมากกว่าหรือเท่ากับ 20 ปีเท่านั้น



เมื่อกำหนดเงื่อนไขในคำสั่งแล้ว คลิ๊กที่ปุ่ม OK จะได้หน้าต่างถัดไป

มหาวิทยาลัยเทคโนโลยีสุรนารี



สังเกตว่าตรงสดมภ์ grade จะมีข้อมูลหลังจาก RECODE แล้วเฉพาะผู้ที่มีอายุตั้งแต่ 20 ปีขึ้นไปเท่านั้น ในคำสั่งอื่นๆ ก็ทำนองเดียวกัน ถ้ามีทางเลือก If... ก็สามารถกำหนดเงื่อนไขแบบ PROCESS IF เช่นเดียวกัน ตั้งแต่เวอร์ชั่นวินโดว์ใช้คำสั่ง If... แทน PROCESS IF

สรุป

มหาวิทยาลัยเทคโนโลยีสุรนารี

กิจกรรมต่อเนื่อง ตอนที่ 9.3

1. ศึกษาผ่านสื่อปฏิสัมพันธ์ผ่านจอภาพ การบรรยาย หน่วยที่ 9 การนำเข้าข้อมูลจากรูปแบบอื่นๆ ตอนที่ 9.3 การเลือกข้อมูลมาวิเคราะห์
2. ศึกษาด้วยตนเองผ่านสื่อบทเรียนคอมพิวเตอร์ช่วยสอน หน่วยที่ 9 การนำเข้าข้อมูลจากรูปแบบอื่นๆ ตอนที่ 9.3 การเลือกข้อมูลมาวิเคราะห์
3. ทำกิจกรรมประกอบการเรียนในคู่มือการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 9 การนำเข้าข้อมูลจากรูปแบบอื่นๆ ตอนที่ 9.3 การเลือกข้อมูลมาวิเคราะห์
4. ทำกิจกรรมเสริมประสบการณ์ในแผนกิจกรรมการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 9 การนำเข้าข้อมูลจากรูปแบบอื่นๆ ตอนที่ 9.3 การเลือกข้อมูลมาวิเคราะห์

มหาวิทยาลัยเทคโนโลยีสุรนารี

3. ปัจจัยด้านบวกที่สนับสนุน และปัจจัยด้านลบที่คอยขัดขวางการบรรลุวัตถุประสงค์ของกองทุนหมู่บ้านและชุมชนเมือง

3.1 ปัจจัยด้านบวกที่สนับสนุน

จากการสรุปโดยภาพรวมทั้ง 5 อำเภอ 667 กองทุน พบว่าปัจจัยด้านบวกที่สนับสนุนการบรรลุวัตถุประสงค์ของกองทุน ประกอบด้วยปัจจัยหลัก ๆ ดังนี้ คือ

3.1.1 การมีส่วนร่วมในการจัดตั้งกองทุน มีสัดส่วนร้อยละ 86.36 (576 กองทุน)

3.1.2 มีการรวมกลุ่มองค์กรต่าง ๆ ในชุมชน มีสัดส่วนร้อยละ 59.82 (399 กองทุน)

3.1.3 มีความเสียสละ / มีความสามัคคี มีสัดส่วนร้อยละ 45.58 (304 กองทุน)

3.1.4 มีการแลกเปลี่ยนข้อมูล มีสัดส่วนร้อยละ 43.33 (289 กองทุน)

เมื่อพิจารณาจากปัจจัยด้านบวกที่สนับสนุนการบรรลุวัตถุประสงค์ของกองทุนในแต่ละอำเภอ สามารถสรุปได้ดังนี้ คือ

1) ปัจจัยด้านบวกที่พบในทุกอำเภอ คือ

- การมีส่วนร่วมในการจัดตั้งกองทุน
- การรวมกลุ่มองค์กรต่าง ๆ ในชุมชน
- ความเสียสละ / ความสามัคคี
- การแลกเปลี่ยนข้อมูล
- สมาชิกมีทัศนคติที่ดีต่อกองทุน
- ประสบการณ์ของคณะกรรมการ
- ความสามารถของคณะกรรมการ
- จำนวนกองทุนเดิมในชุมชน
- การเกิดเครือข่ายการเรียนรู้ ค้นพบเทคนิควิธีใหม่ในการเพิ่มผลผลิต

2) ปัจจัยด้านบวกที่พบเฉพาะอำเภอสีคิ้ว คือ

- มีเงินสะสมเพิ่มขึ้น
- รู้จักการอดออม
- หน่วยงานราชการให้ความช่วยเหลือ
- การตรวจสอบจากทางราชการ
- ผู้นำมีความเข้มแข็ง