12

ชีวสถิติและทักษะทางคอมพิวเตอร์

(Biostatistics and computing skills)

เสนอ

ดร.ปัทมา สุพรรณกุล

โดย

ภก. อุปถัมภ์	พูลหน่าย	รหัส 51077735
นาย พชรดนัย	ทับทิม	รหัส 51077407
นางสาว ชลธิชา	ยรรยงค์	รหัส 51077247
นางสาว ชลธิชา	มาน้อย	รหัส 51077230
นางสาว ทิพวรรณ	อินอ่ำ	รหัส 51077292
นางสาว นภัสวรรณ	เมืองจันทร์	รหัส 51077315
นางสาว ชลธพร	คงจำนงค์	รหัส 51077223

รายงานนี้เป็นส่วนหนึ่งของการศึกษา

ชีวสถิติและทักษะทางคอมพิวเตอร์ (551532)

หลักสูตรสาธารณสุขศาสตรมหาบัณฑิต มหาวิทยาลัยนเรศวร พิษณุโลก

ปีการศึกษา 2551

คำนำ

รายงานฉบับนี้เป็นส่วนหนึ่งของวิชา ชีวสถิติและทักษะการใช้คอมพิวเตอร์ (Biostatistics and computing skills) รหัสวิชา 551532 ซึ่งคณะผู้จัดทำได้ดำเนินการเรียบเรียงและรวบรวมจากเอกสารการเรียน การสอนในชั่วโมงเรียนของอาจารย์ ดร.ปัทมา สุพรรณกุล เป็นการเรียนรู้ในเรื่องความรู้เบื้องต้นเกี่ยวกับสถิติ ความหมายของข้อมูล ตัวแปร ประชากร กลุ่มตัวอย่าง และวิธีการสุ่มตัวอย่าง สถิติเชิงพรรณนาที่ใช้ในงานวิจัย การทดสอบสมมติฐานทางการวิจัย การลงรหัสในแบบสอบถาม การนำเข้าข้อมูล การตรวจสอบคุณภาพเครื่องมือ การทดสอบข้อตกลงเบื้องต้นของสถิติ การทดสอบความสัมพันธ์ การทดสอบค่าเฉลี่ย 2 ประชากร การวิเคราะห์ความแปรปรวน การวิเคราะห์ความถดถอยและสหสัมพันธ์อย่างง่าย การวิเคราะห์ความถดถอยและสหสัมพันธ์เชิงพหุคูณ และทักษะในการใช้โปรแกรมการวิเคราะห์ข้อมูล SPSS

คณะผู้จัดทำ หวังว่า รายงานฉบับนี้ คงจะเป็นประโยชน์สำหรับผู้อ่านบ้างไม่มากก็น้อย หากมีข้อผิดพลาดประการใด คณะผู้จัดทำขออภัยมา ณ ที่นี้

คณะผู้จัดทำ

มกราคม 2552

คณะสาธารณสุขศาสตร์ มหาวิทยาลัยนเรศวร
วันลงทะเบียน 26 มี.ค. 2553
หมายเลข bib b1788160
หมายเลข item 1514033X
เลขเรียกหนังสือ WA20.5
ร581
2552
ฉ.2

สารบัญ

หน้า

คำนำ

สารบัญ

การนำเข้าสู่โปรแกรม SPSS 1

- การเริ่มใช้โปรแกรม SPSS 1
- การใช้คำสั่งโปรแกรม SPSS 7

สถิติ

- Chi-square 23
- T-Test 33
  - O One-Sample T-Test 33
  - O Independent T-Test 36
  - O Paired T-Test 39
- One- way ANOVA 41
- Regression Analysis 45
  - O Simple Regression 48
  - O Multiple Regression Analysis (MRA) 62
  - O Dummy Variables in Multiple Regression Analysis 79

การนำเข้าสู่โปรแกรม SPSS

SPSS เป็นเครื่องมือสำหรับจัดกระทำข้อมูล สามารถนำข้อมูลจากแฟ้มข้อมูลมาประมวลผลด้วยตาราง สร้างกราฟ และวิเคราะห์ทางสถิติที่หลากหลาย ถ้าผู้วิเคราะห์ต้องการใช้โปรแกรม SPSS ดำเนินการจัดการกับข้อมูล วิเคราะห์ข้อมูล และรายงานผลการวิเคราะห์ข้อมูล ผู้วิเคราะห์จะต้องเขียนโปรแกรมคำสั่งเพื่อสั่งให้โปรแกรม SPSS กระทำตามที่ผู้วิเคราะห์ต้องการ

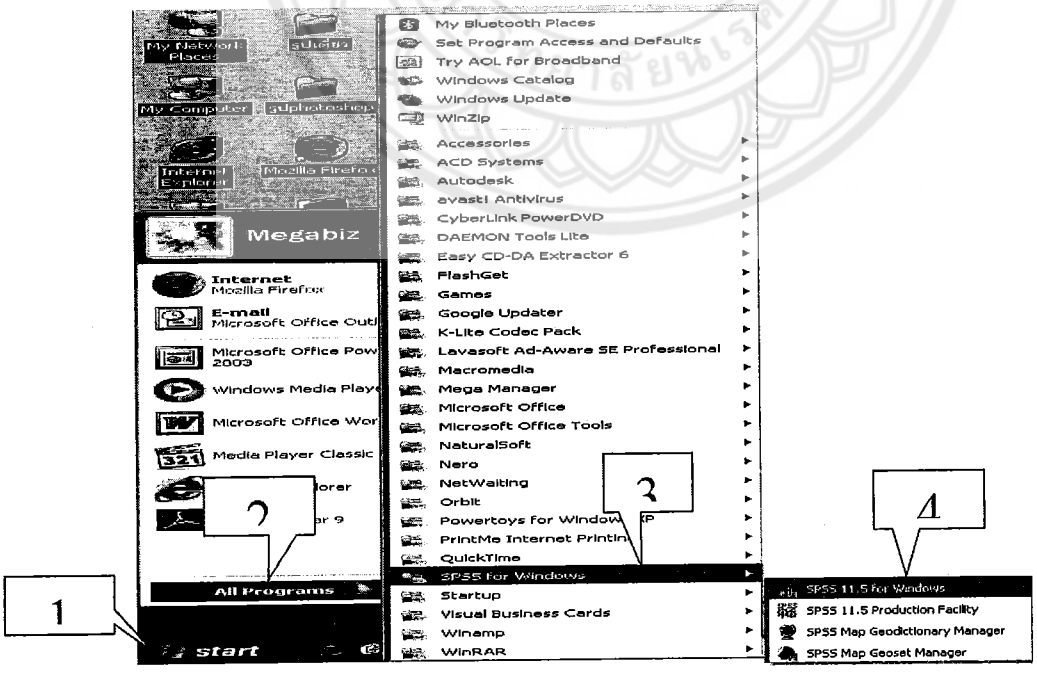
โดยต้องมีส่วนประกอบสำคัญ 2 อย่างคือ

1. โปรแกรม
2. ข้อมูลจากแบบประเมินหรือแบบสอบถาม

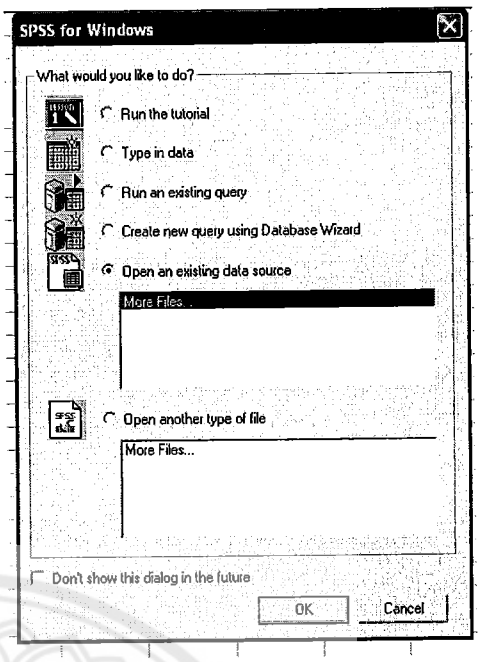
1. การเริ่มใช้โปรแกรม SPSS

เมื่อทำการติดตั้งโปรแกรม SPSS เรียบร้อยแล้ว โปรแกรม SPSS จะทำการเพิ่มรายการการทำงานของ SPSS เข้าไปในเมนู Programs ของ Windows ดังนั้น ในการเรียกใช้งานโปรแกรม SPSS ผู้ใช้สามารถเลือกการใช้งานได้ ตามขั้นตอนดังนี้

1. เลือก Start
2. เลือก All Program
3. เลือก SPSS for Windows
4. เลือก SPSS 11.5 for Windows



เมื่อเริ่มเข้าสู่การทำงาน โปรแกรม SPSS จะแสดง Dialog การทำงานให้ผู้ใช้เลือกการใช้งาน SPSS ดังรูป



**** Dialog เมื่อเริ่มเข้าสู่การใช้งานนั้น ผู้ใช้สามารถเลือกตัวเลือก* ☐ Don't show this dialog in the future *เพื่อกำหนดให้ SPSS ไม่ต้องแสดง Dialog นี้เมื่อทำการเปิดโปรแกรมครั้งต่อไปได้*

สำหรับ Dialog ที่ปรากฏนี้ จะมีรายการให้ผู้ใช้ทำการเลือกการทำงานดังนี้

☐ Run the tutorial เป็นการเลือกให้ SPSS ทำการเปิด tutorial เพื่อแนะนำการใช้งานในเบื้องต้นของโปรแกรม SPSS (หากผู้ใช้ต้องการเลือกตัวเลือกนี้ จะต้องใส่แผ่น CD ที่มีไฟล์ tutorial ของ SPSS ไว้ที่เครื่องด้วย)

☐ Type in data เลือกเมื่อต้องการใส่ข้อมูลใหม่เอง

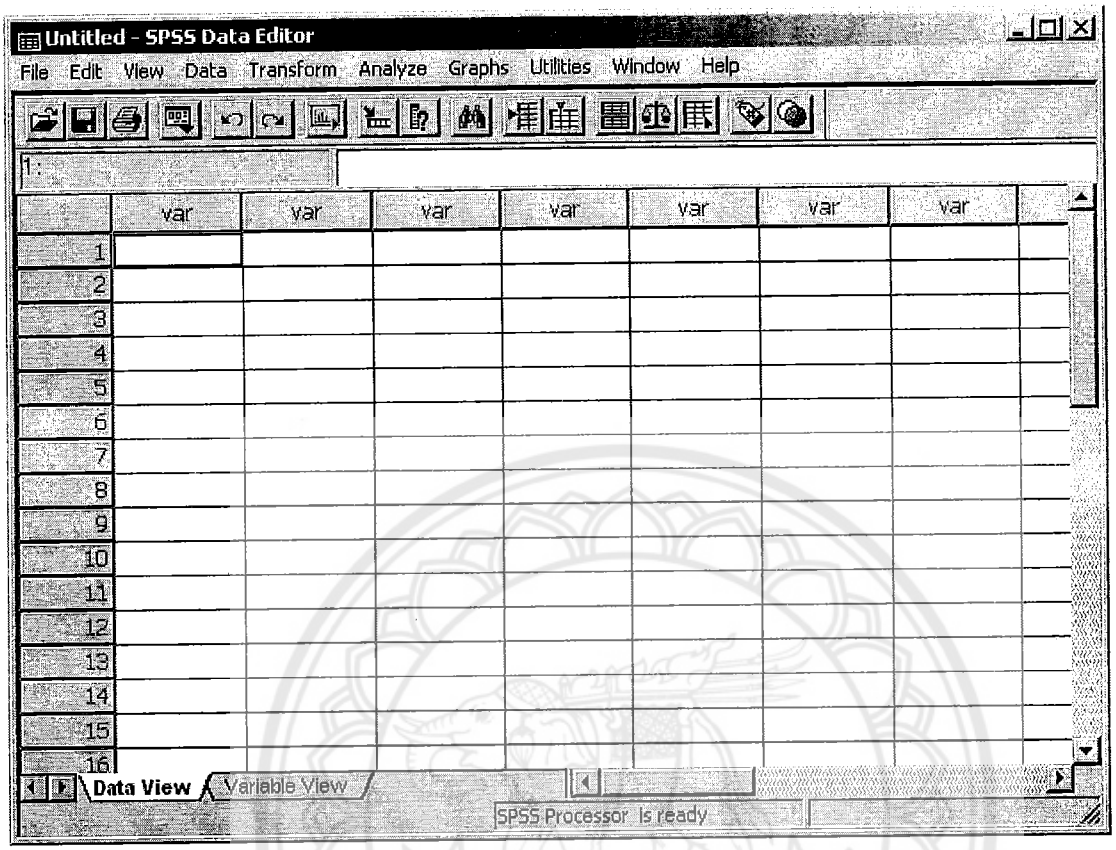
☐ Run an existing query เลือกเพื่อเปิดฐานข้อมูลที่ผู้ใช้ทำการสร้างจากโปรแกรมจัดการฐานข้อมูลอื่นๆ ที่ SPSS สามารถอ่านได้ และทำการกำหนดให้ SPSS เชื่อมโยงข้อมูลไว้แล้ว

☐ Create new query using Database Wizard เลือกเพื่อให้โปรแกรม SPSS ทำการเปิดไฟล์ฐานข้อมูลที่ถูกสร้างไว้โดยโปรแกรมจัดการฐานข้อมูลอื่น และนำมาวิเคราะห์ในโปรแกรม SPSS เช่น ฐานข้อมูลที่ถูกสร้างโดย MS Access

☐ Open an existing data source เลือกเพื่อเปิด File data ที่ได้ทำการสร้างไว้ใน SPSS ที่มีอยู่แล้ว

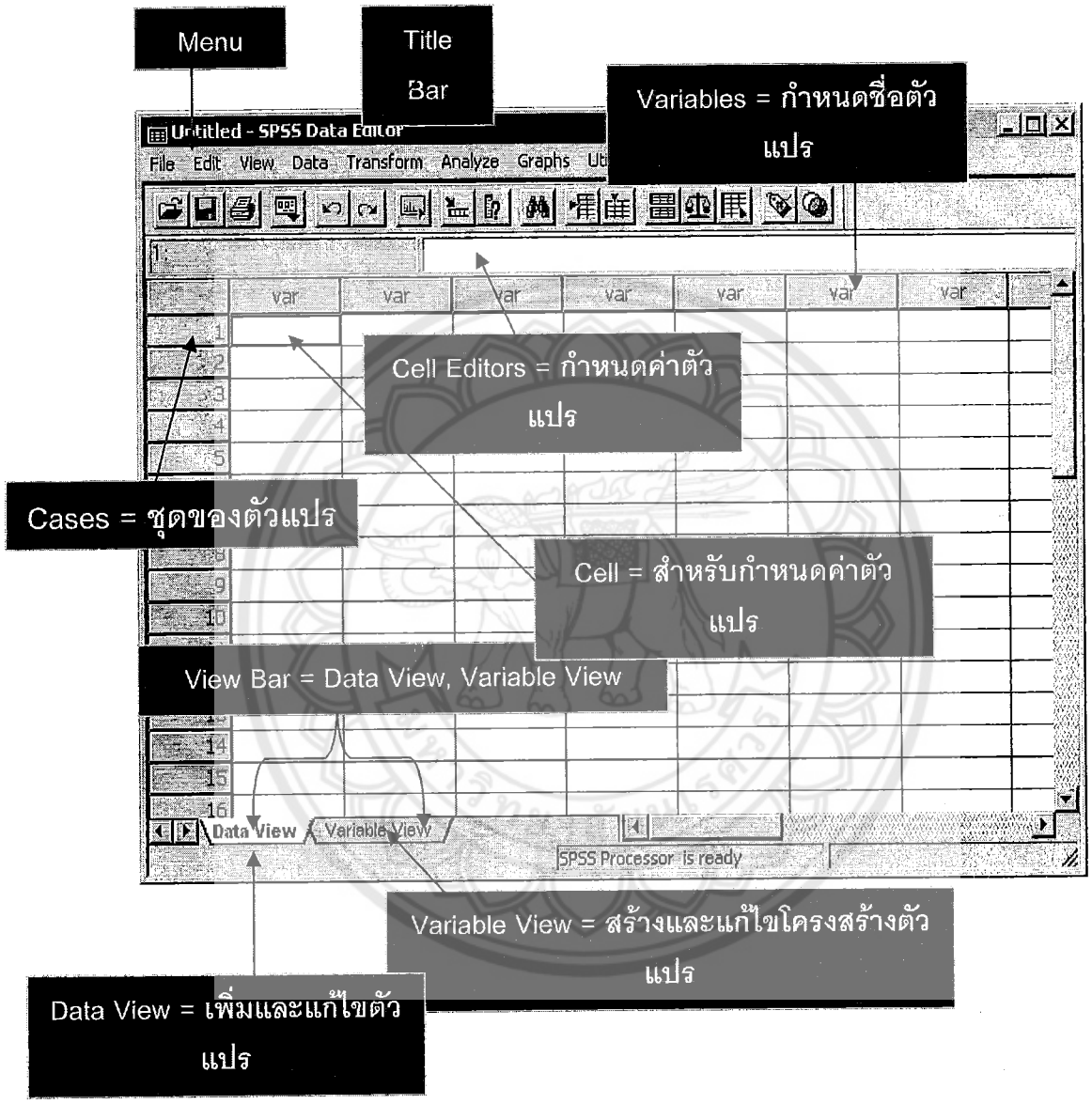
☐ Open another type of file เลือกเพื่อเปิด File ชนิดอื่นๆ ของ SPSS ที่มีอยู่แล้ว

เริ่มต้นใช้งาน ให้เลือกตัวเลือก Type in data แล้วเลือก OK (หรือหากไม่เลือกรายการใดๆ ให้กด Cancel โปรแกรมจะแสดงหน้าต่าง SPSS Data Editor ดังรูป



ส่วนประกอบต่างๆ บนหน้าต่าง SPSS Data Editor

หน้าต่าง SPSS Data Editor จะเป็นหน้าต่างหลัก และเป็นหน้าต่างแรกสุดเมื่อผู้ใช้ทำการเปิดโปรแกรม SPSS ใช้งาน เมื่อเริ่มเข้าสู่หน้าต่างนี้ ยังไม่มีการเปิดไฟล์ข้อมูลใดๆ หน้าต่างจะแสดงชื่อไฟล์ว่า "Untitled" ดังรูป



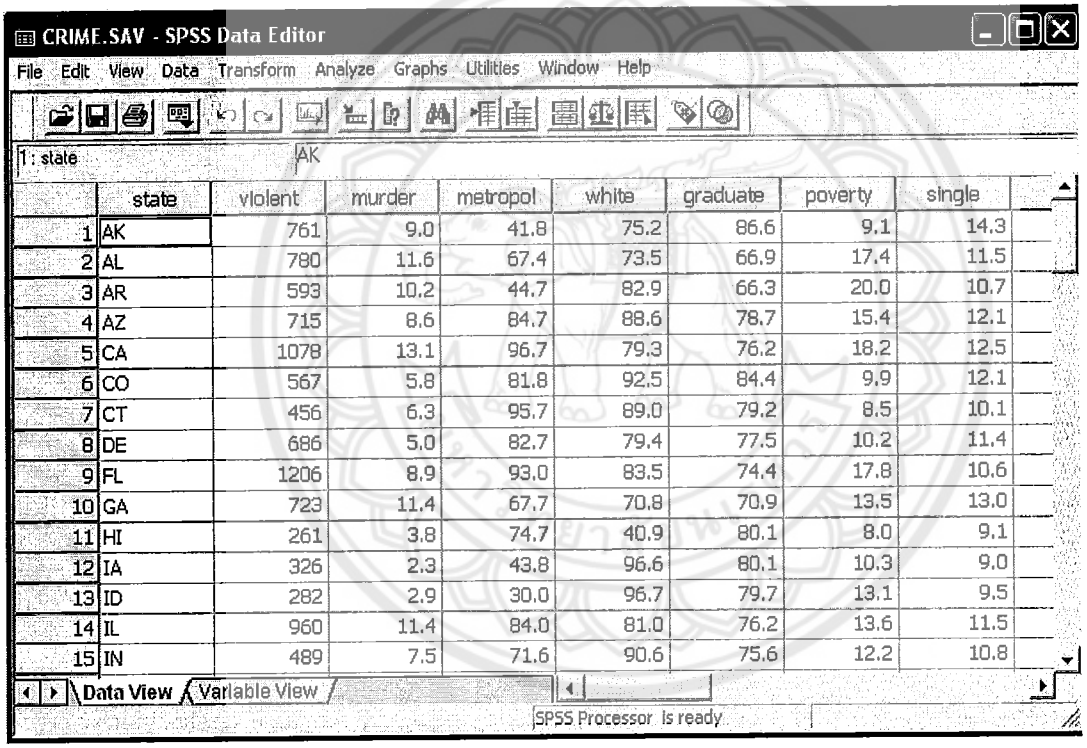
*** ภายในหน้าต่าง ประกอบด้วยพื้นที่ทำงานที่เรียกว่า Data sheet มีลักษณะเป็นตารางว่างเปล่ามีเส้นแบ่งเป็นคอลัมภ์และแถว แต่ละคอลัมภ์จะมีหัวคอลัมภ์แสดงข้อความในพื้นสีเทาว่า "var" และแต่ละแถวจะมีตัวเลขแสดงแถวอยู่บนพื้นสีเทา ตัวเลขจะเรียงลำดับกันตั้งแต่ แถวที่ 1, 2, 3, ...

วัตถุประสงค์ของมุมมอง (View Bar) คือ เป็นการแบ่งงานออกเป็นสองส่วนใหญ่คือ ส่วนออกแบบกำหนดโครงสร้างตัวแปรและส่วนกำหนดค่าชุดตัวแปร

โดยแบ่งมุมมองออกเป็น 2 มุมมอง ได้แก่

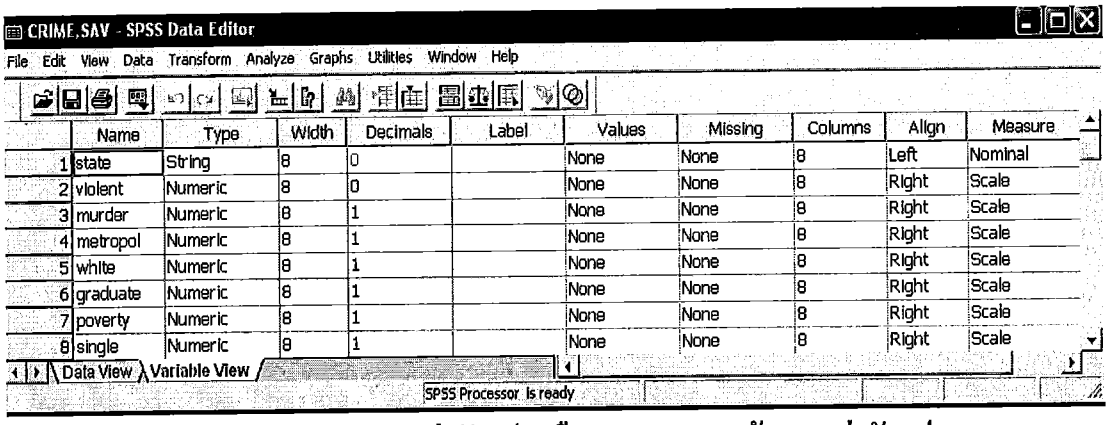
1. **Data view**
2. **Variable view**

Data view เป็นส่วนทำงานของโปรแกรมที่แตกต่างไปจากโปรแกรมอื่น มีลักษณะเป็น Data Sheet คือเป็นหน้ากระดาษคล้าย Worksheet ใน Microsoft Excel สำหรับแสดงข้อมูลจากแฟ้มข้อมูลปัจจุบันที่เปิดใช้งานอยู่ Data view เป็นส่วนกำหนดค่าชุดของตัวแปรในแต่ละชุดหรือ Data entry ดังภาพ



	state	violent	murder	metropol	white	graduate	poverty	single
1	AK	761	9.0	41.8	75.2	86.6	9.1	14.3
2	AL	780	11.6	67.4	73.5	66.9	17.4	11.5
3	AR	593	10.2	44.7	82.9	66.3	20.0	10.7
4	AZ	715	8.6	84.7	88.6	78.7	15.4	12.1
5	CA	1078	13.1	96.7	79.3	76.2	18.2	12.5
6	CO	567	5.8	81.8	92.5	84.4	9.9	12.1
7	CT	456	6.3	95.7	89.0	79.2	8.5	10.1
8	DE	686	5.0	82.7	79.4	77.5	10.2	11.4
9	FL	1206	8.9	93.0	83.5	74.4	17.8	10.6
10	GA	723	11.4	67.7	70.8	70.9	13.5	13.0
11	HI	261	3.8	74.7	40.9	80.1	8.0	9.1
12	IA	326	2.3	43.8	96.6	80.1	10.3	9.0
13	ID	282	2.9	30.0	96.7	79.7	13.1	9.5
14	IL	960	11.4	84.0	81.0	76.2	13.6	11.5
15	IN	489	7.5	71.6	90.6	75.6	12.2	10.8

Variable view เป็นส่วนกำหนดคุณสมบัติของตัวแปร Variable ที่ผู้ใช้ได้ทำการสร้างไว้ เมื่อเริ่มใช้งานจะเป็นกระดาษว่างเปล่าจนกว่าผู้ใช้จะทำการกำหนดและสร้างชุดตัวแปรขึ้นดังภาพ


CRIME.SAV - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	state	String	8	0		None	None	8	Left	Nominal
2	violent	Numeric	8	0		None	None	8	Right	Scale
3	murder	Numeric	8	1		None	None	8	Right	Scale
4	metropol	Numeric	8	1		None	None	8	Right	Scale
5	white	Numeric	8	1		None	None	8	Right	Scale
6	graduate	Numeric	8	1		None	None	8	Right	Scale
7	poverty	Numeric	8	1		None	None	8	Right	Scale
8	single	Numeric	8	1		None	None	8	Right	Scale

Data View | Variable View

SPSS Processor is ready

- **Width = กำหนดจำนวนของค่าตัวแปรหรือจำนวนความกว้างของค่าตัวแปร**
- **Decimals = กำหนดจำนวนของจุดทศนิยม**
- **Label = กำหนดชื่อของตัวแปรจะมีผลเมื่อเราออกแบบรายงานเป็นกราฟ**
- **Value = กำหนดค่าตัวแปรโดยมิต้องไปกำหนดที่ Variable view**
- **Missing = กำหนดเมื่อไม่พบค่าตัวแปรของชุดตัวแปรนั้น**
- **Columns = กำหนดความกว้างของช่อง Columns สำหรับกรอกข้อมูล**
- **Align = จัดค่าของชุดตัวแปรให้ชิดซ้าย กลาง หรือ ขวา**
- **Measure = กำหนดมาตราวัดของตัวแปร**

2. การใช้คำสั่งโปรแกรม

เลือก โปรแกรม SPSS

การจัดการและการตรวจสอบลักษณะข้อมูล

การให้รหัสในแบบสอบถาม

การใช้คำสั่ง Recode

การเรียงลำดับข้อมูล (Sorting data)

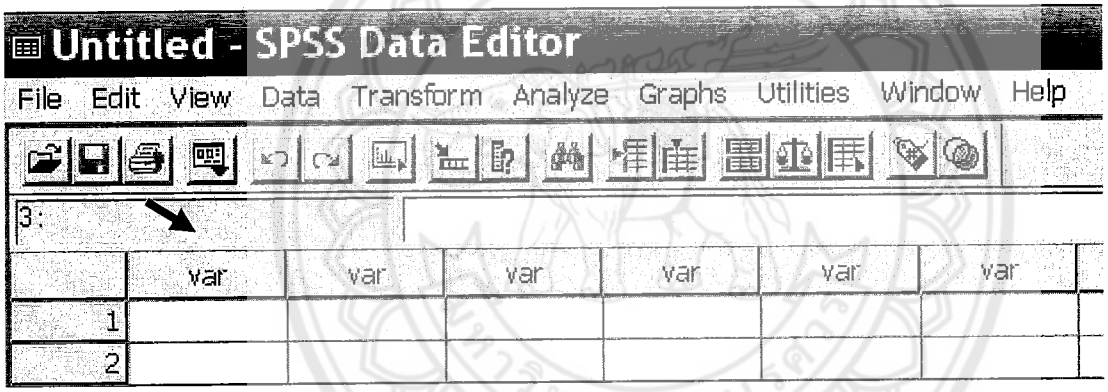
การเลือกบางส่วนของตัวอย่างสำหรับการวิเคราะห์

การตรวจสอบลักษณะการแจกแจงของข้อมูล

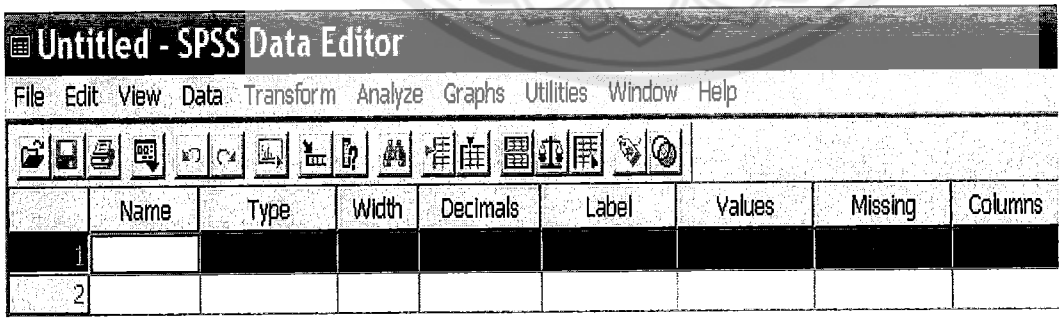
- การใช้คำสั่ง Explore

2.1 การคีย์ข้อมูลจากแบบสอบถาม

การสร้างตัวแปร



ดับเบิลคลิก ที่ Variable (var) จะปรากฏ ดังภาพ



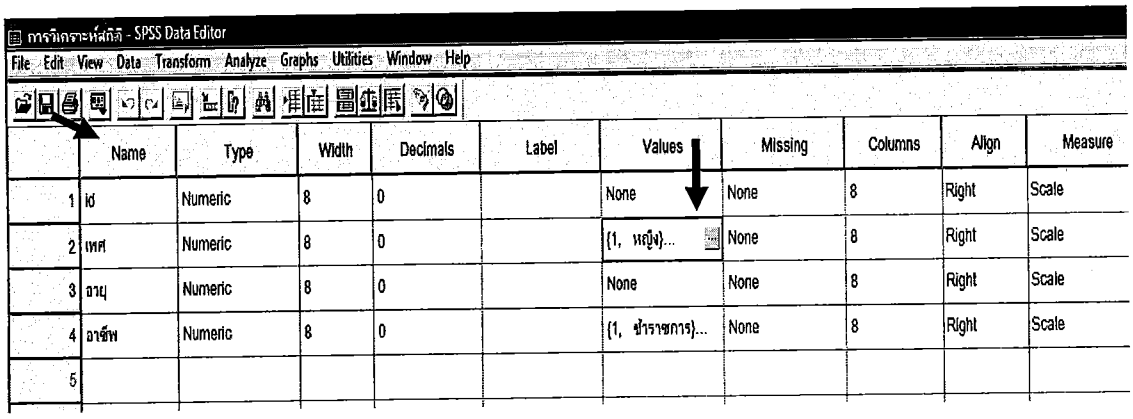
กำหนดตัวแปรต่างๆ ลงใน Name **คือ**

Id ลำดับที่ของแบบสอบถาม

เพศ (ข้อมูลเชิงคุณภาพ)

อายุ (ข้อมูลเชิงปริมาณ)

อาชีพ



แล้วกำหนด Values ต่าง ๆ โดยคลิกตรงเครื่องหมาย .... ในช่อง Values

เช่น เพศ 1 = หญิง 2 = ชาย

อาชีพ 1 = ข้าราชการ

2 = เกษตร

3 = ค้าขาย

จะปรากฏภาพแล้วกำหนดค่าลงในช่อง

Value

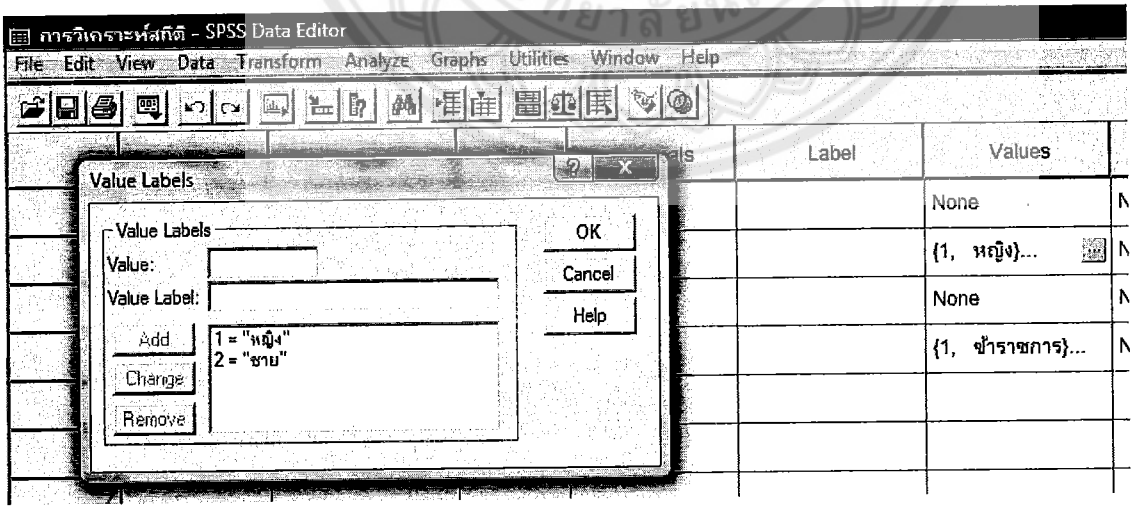
Value Label

ถ้าต้องการเพิ่ม คลิก add

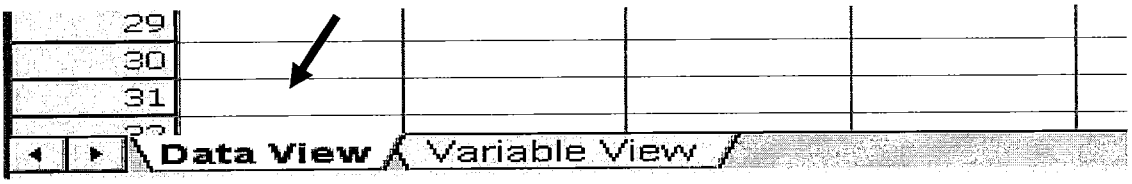
เปลี่ยน คลิก change

ไม่ต้องการ คลิก Remove

ถ้าถูกต้องแล้วคลิก OK



เรียบร้อยแล้วไปที่ Data view อยู่ด้านล่างซ้ายของหน้าจอ



คีย์ข้อมูลต่างๆ ลง ในแต่ละ Column โดยใช้รหัสต่างๆ แทน ตัวหนังสือ เช่น

1 แทน หญิง, 2 แทน ชาย

การวิเคราะห์สถิติ – SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window H

1 : id 1

	id	เพศ	อายุ	อาชีพ
1	1	1	35	
2	2	2	38	
3	3	2	42	
4	4	1	47	
5	5	1	51	

หาก

ต้องการแสดงข้อมูลจาก ตัวเลข ให้เป็น ตัวหนังสือ ให้คลิกที่ จะแสดงข้อมูลให้เห็นดังภาพ

ต้องการกลับเป็นเหมือนเดิมก็คลิก ที่เดิม

การวิเคราะห์สถิติ – SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window H

1 : id 1

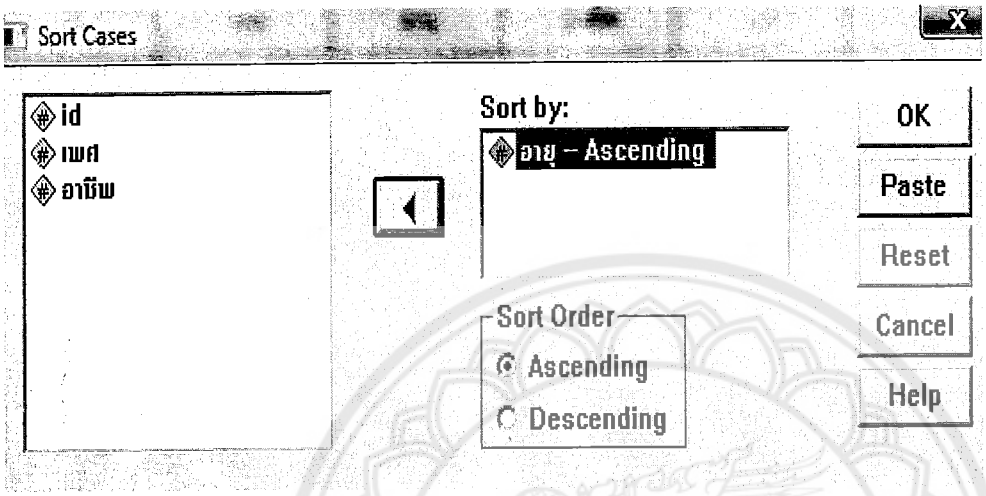
	id	เพศ	อายุ	อาชีพ
1	1	หญิง	35	ค้าขา
2	2	ชาย	38	ค้าขา
3	3	ชาย	42	เกษตรก
4	4	หญิง	47	เกษตรก
5	5	หญิง	51	ข้าราชกา
6	6	หญิง	38	ข้าราชกา

การเรียงข้อมูล

โดย เลือก Data > sort case คลิกข้อมูลที่ต้องการเรียง เช่น อายุ

1. Ascending เรียงข้อมูล จาก น้อย ไป มาก
2. Descending เรียงข้อมูล จาก มาก ไป น้อย

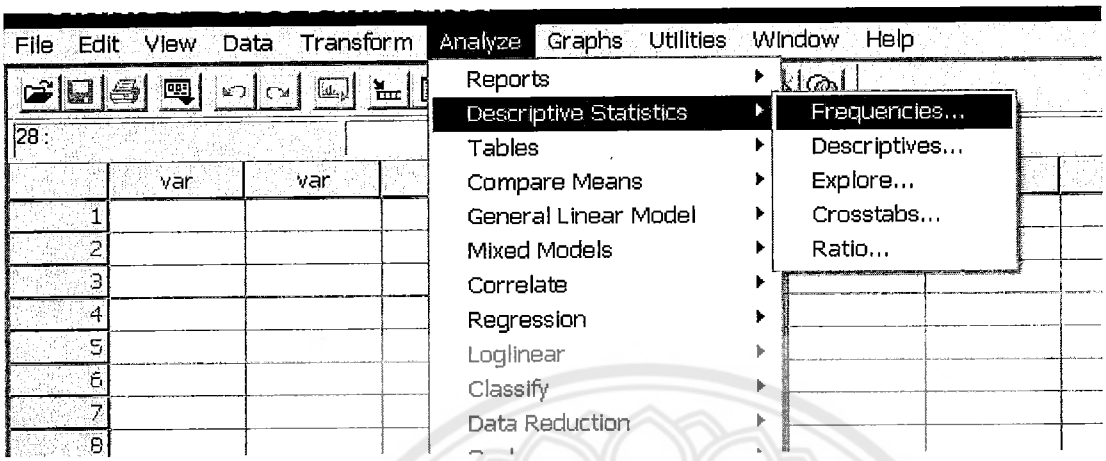
ต่อไปให้คลิก OK



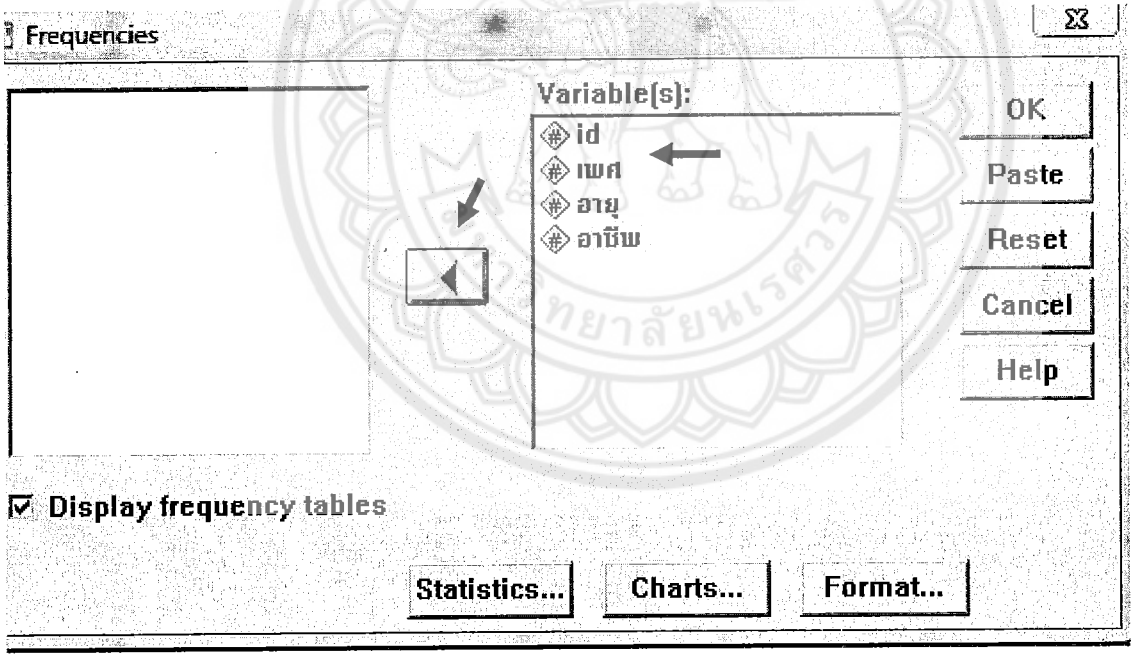
2.2 การวิเคราะห์ข้อมูลและ การคำนวณพื้นฐาน

<u>คำสั่ง Frequencies</u>

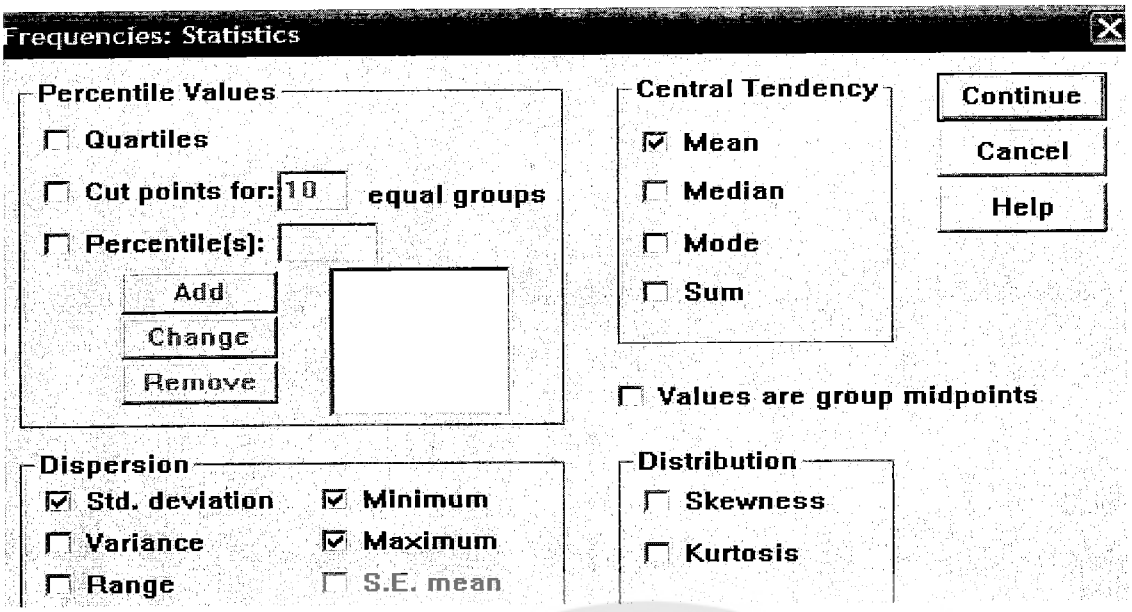
Analyze > Descriptive statistics > Frequencies



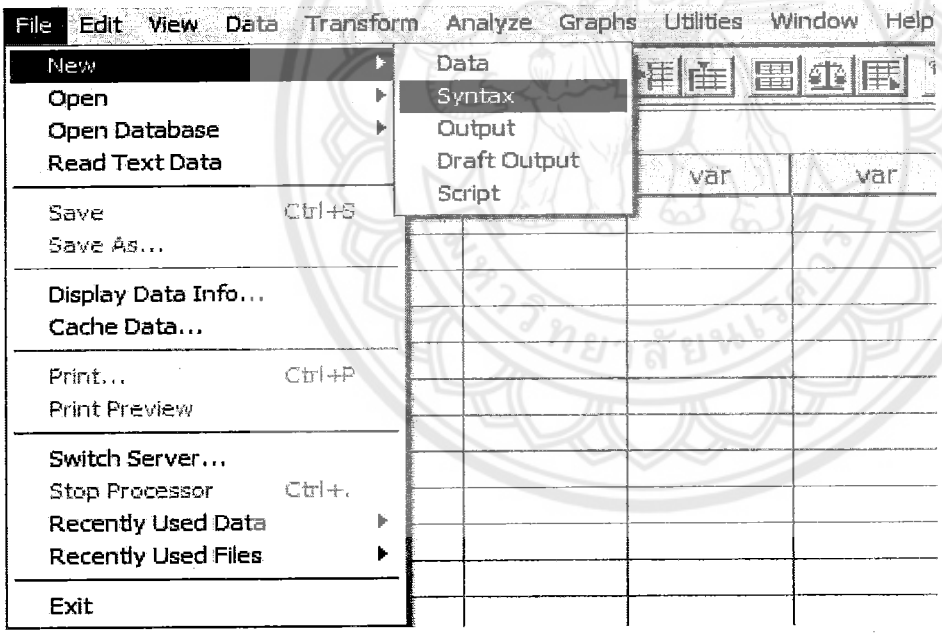
เลือกข้อมูลที่ต้องการ แล้วคลิกที่ลูกศร



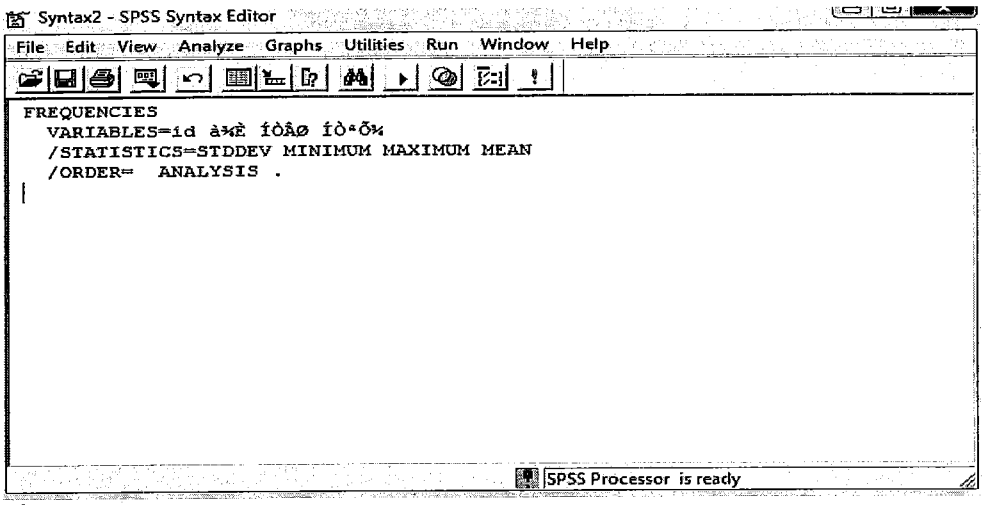
เลือกกำหนด Statistics ต่าง ๆ เช่น เลือก Std. deviation, mean, min, max ปรากฏดังภาพ



Syntax



คำสั่ง Syntax ใช้ในกรณีลืมคำสั่งในการวิเคราะห์ แล้วต้องการเข้าไปดูอีกครั้ง



เมื่อคลิกเข้าไปจะปรากฏ Out put ออกมาดังภาพ

และหากคลิก continue จาก frequencies จะปรากฏข้อมูล คือ

Frequencies

Statistics

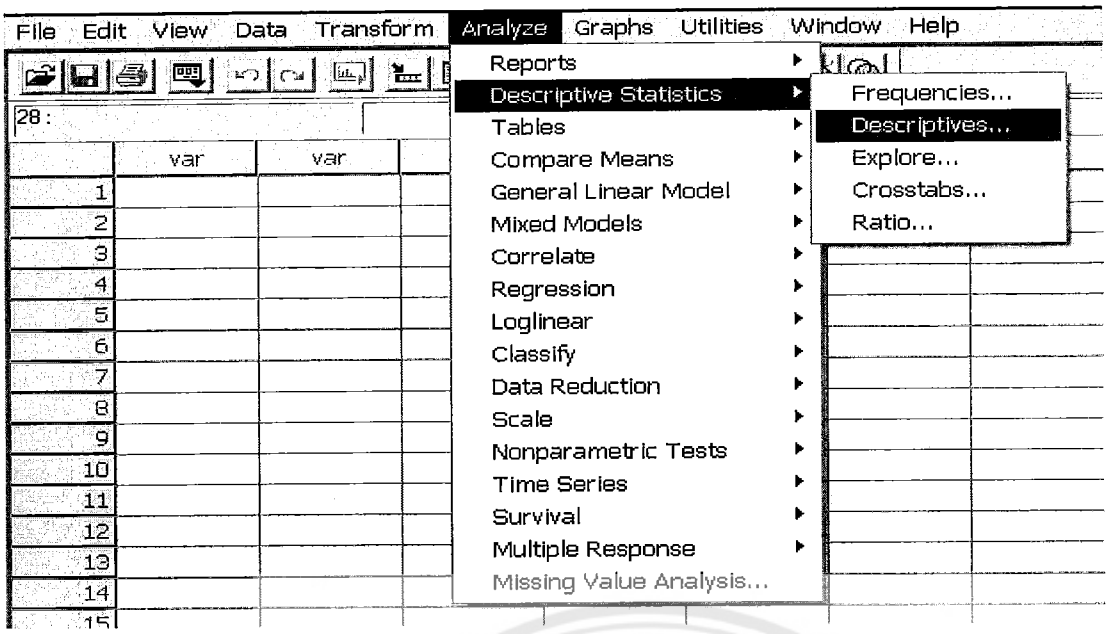
		เพศ	อายุ	อาชีพ
N	Valid	10	10	10
	Missing	0	0	0
Mean		1.50	43.70	2.10
Std. Deviation		.527	6.219	.876
Minimum		1	35	1
Maximum		2	52	3

Frequency Table

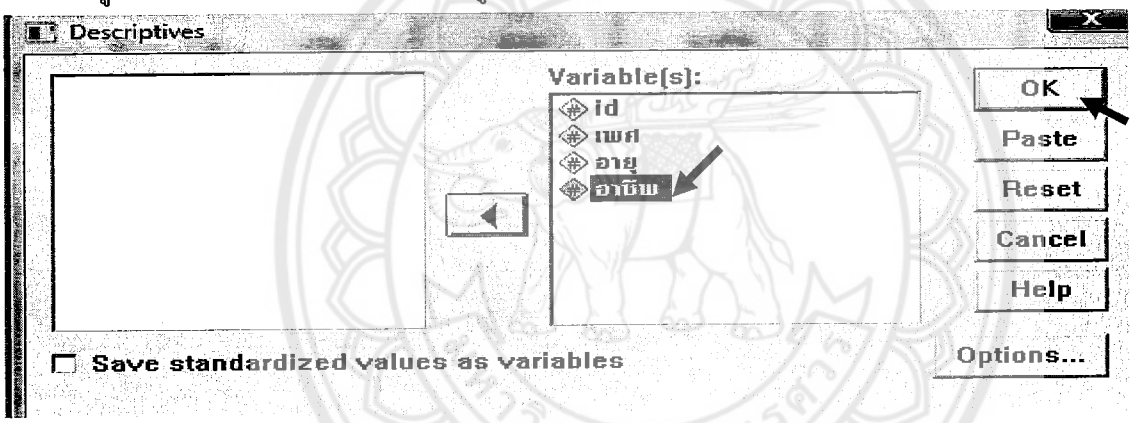
เพศ

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	หญิง	5	50.0	50.0	50.0
	ชาย	5	50.0	50.0	100.0
	Total	10	100.0	100.0	

Analyze > Descriptive statistics > Descriptive



เลือกข้อมูลที่ต้องการวิเคราะห์ แล้วคลิกที่ลูกศร เลือก OK



จะแสดง Out put ออกมา ดังภาพ อาชีพ ?

Descriptives

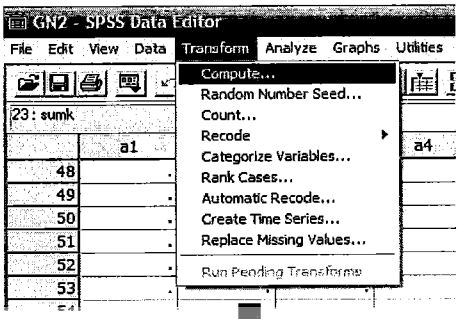
Descriptive Statistics

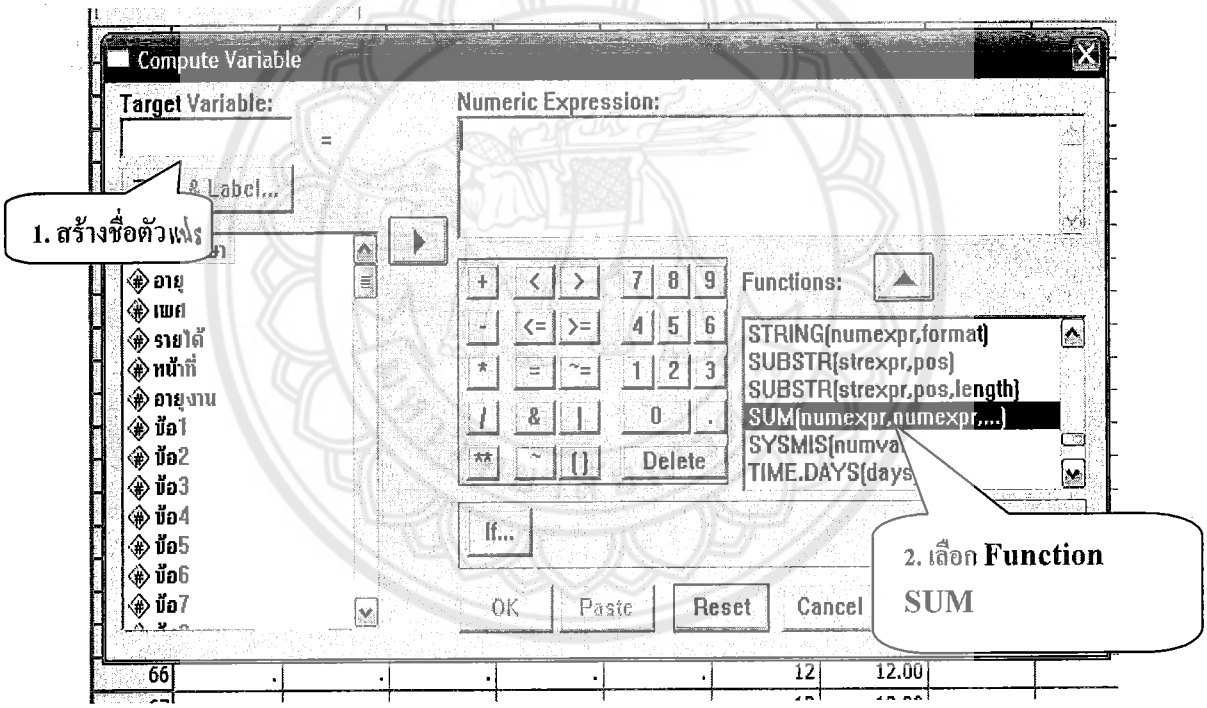
	N	Minimum	Maximum	Mean	Std. Deviation
เพศ	10	1	2	1.50	.527
อายุ	10	35	52	43.70	6.219
อาชีพ	10	1	3	2.10	.876
Valid N (listwise)	10				

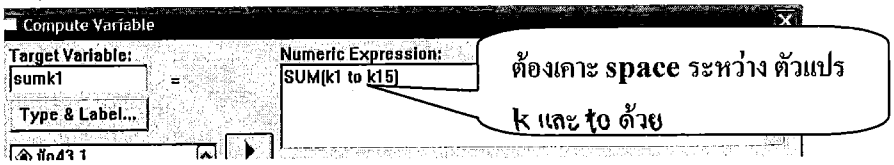
<u>การบวกค่า(คะแนน) ของตัวแปร ด้วยคำสั่ง(Compute)</u>

โดยการเลือกคำสั่ง Compute

จากเมนู เลือก Transform > Compute







ข้อควรระวัง

1. ก่อน SUM ต้อง clean ข้อมูลก่อน (อาจมีข้อมูลที่ไม่ได้กรอก)
2. การ SUM แบบใช้ Functions ต้องเป็นข้อมูลที่เรียงกัน

<u>คำสั่ง Recode ตัวแปร</u>

สามารถทำได้ 2 แบบ

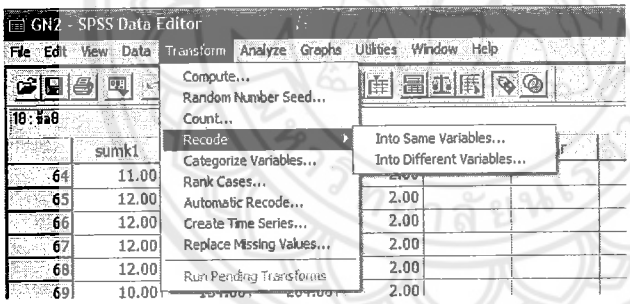
1. ใช้แทนค่า
2. ใช้จัดกลุ่มหรือแบ่งช่วงข้อมูล

มี 2 ตัวคือ

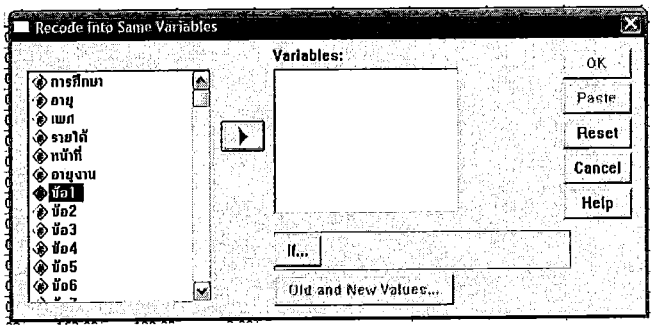
1. Into same ใช้ในการคำนวณคะแนนที่ดีกว่า ? เมื่อต้องการเปลี่ยนแปลงข้อมูลในตัวแปรเดิม
2. Into different ใช้ในการจัดข้อมูลเชิงปริมาณเป็นกลุ่ม กรณีที่เป็น 0 แท้ ได้แก่ อายุ รายได้ เป็นต้น ? เมื่อต้องการสร้างตัวแปรใหม่ โดยเปลี่ยนแปลงข้อมูลจากตัวแปรเดิม

การ Recode เพื่อใช้แทนค่า

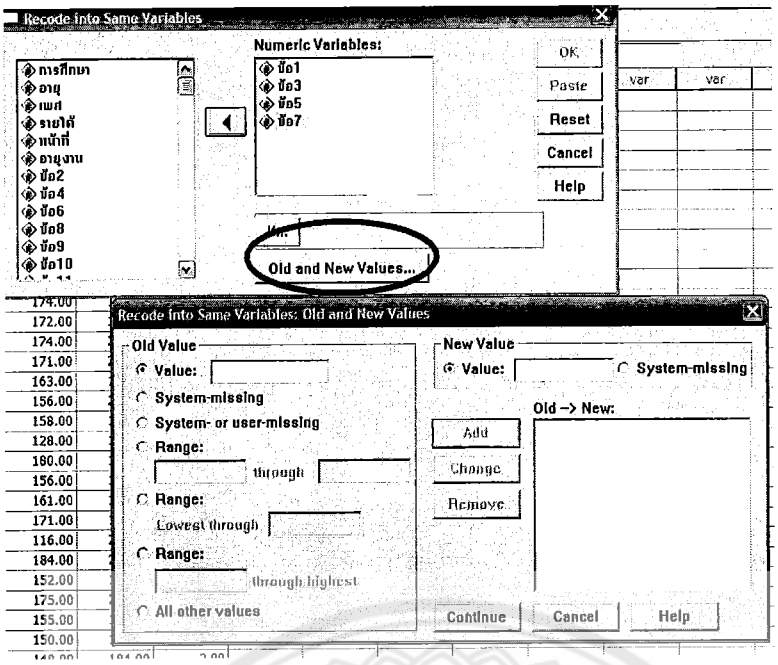
1. เลือก Transform > Recode
2. เลือก Recode > into same variables…



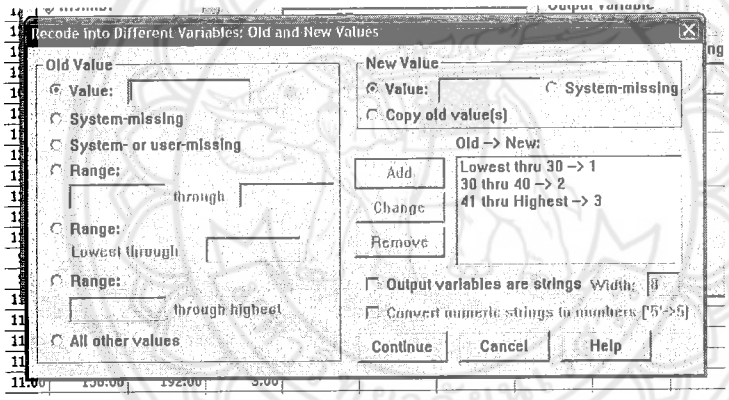
3.เลือกตัวแปรที่ต้องการ Recode

4. 

เลือก old and new values



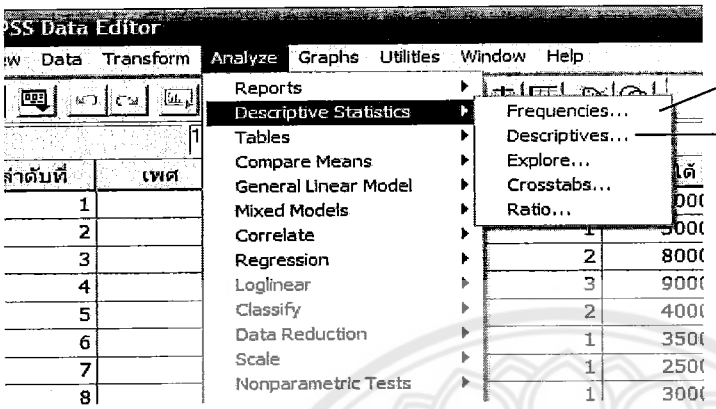
5.กำหนดค่าต่างๆที่ต้องการ recode



6. จากนั้นเลือก Continue > OK

2.3 การวิเคราะห์ข้อมูล

1. Analyze
2. Descriptive และเลือกตามต้องการ

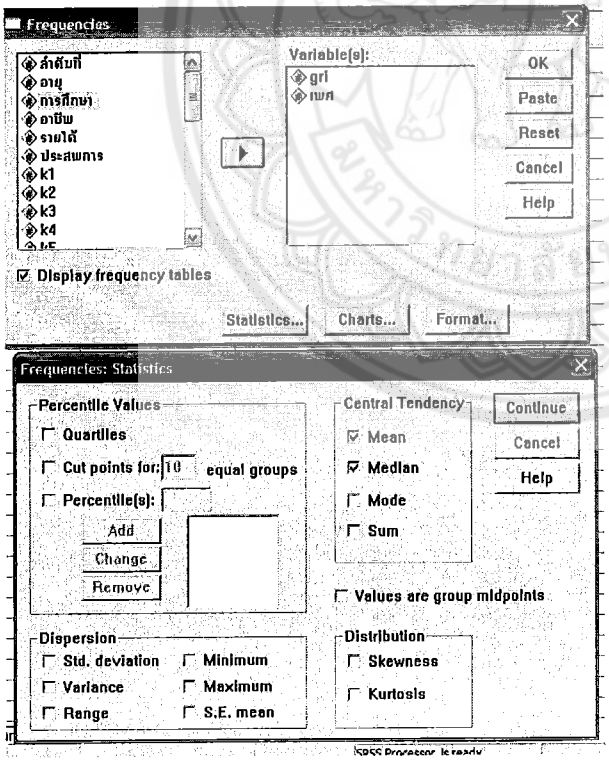


แจกแจงความถี่

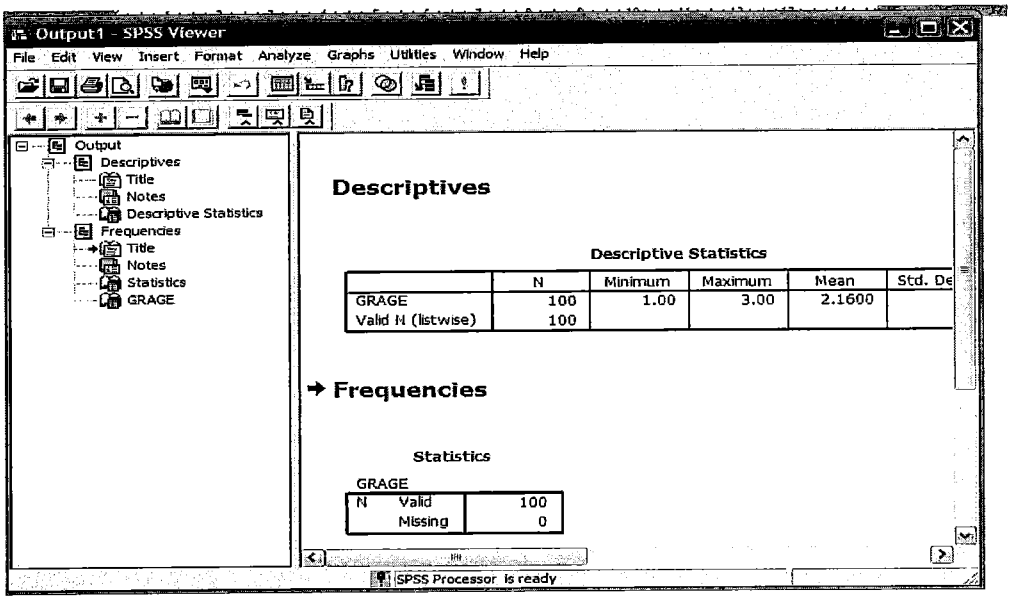
ใช้หาค่ามากสุด -น้อย และสถิติเชิงพรรณนาอื่นๆ

3.เลือกตัวแปรที่ต้องการข้อมูลและเลือก statistics...

4.เลือกค่าที่ต้องการให้คำนวณ เช่น ค่า mean , max



4. Continue และ ok.จะได้ดังรูป

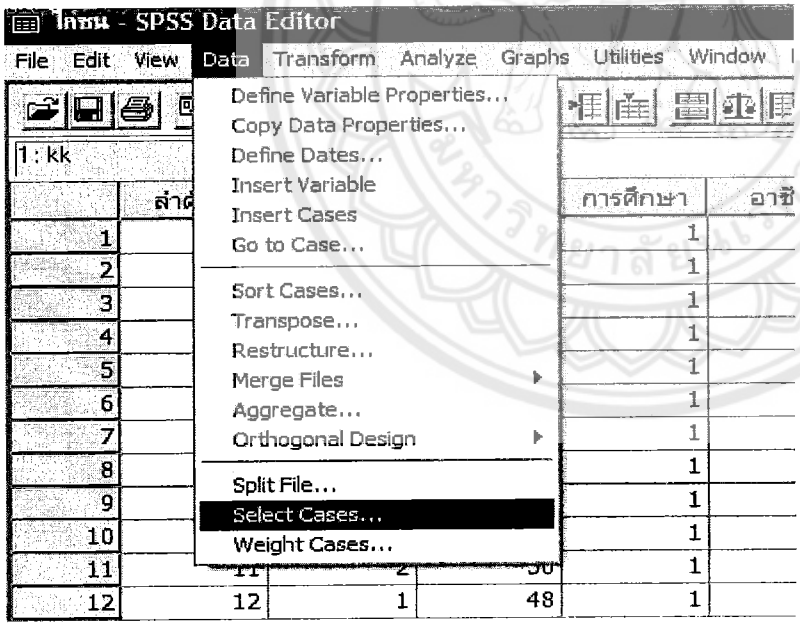


การใช้คำสั่ง select cases

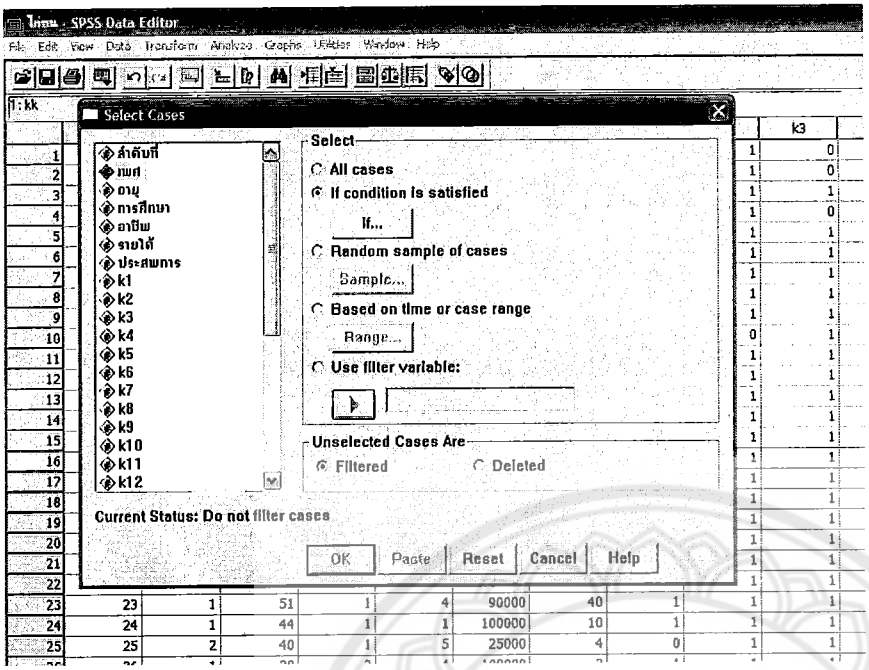
Select cases ใช้ในการที่เราต้องการดูค่าตัวแปรเฉพาะตัวแปร เช่นต้องการดูว่านักเรียนกลุ่มนี้ มีเด็กที่อายุ 10 ปีกี่คน

? ต้องการเลือกข้อมูลบางรายตามเงื่อนไขมาวิเคราะห์

1.เลือก Data เลือก Select cases



2. เลือกค่าที่ต้องการ เลือก if condition ....

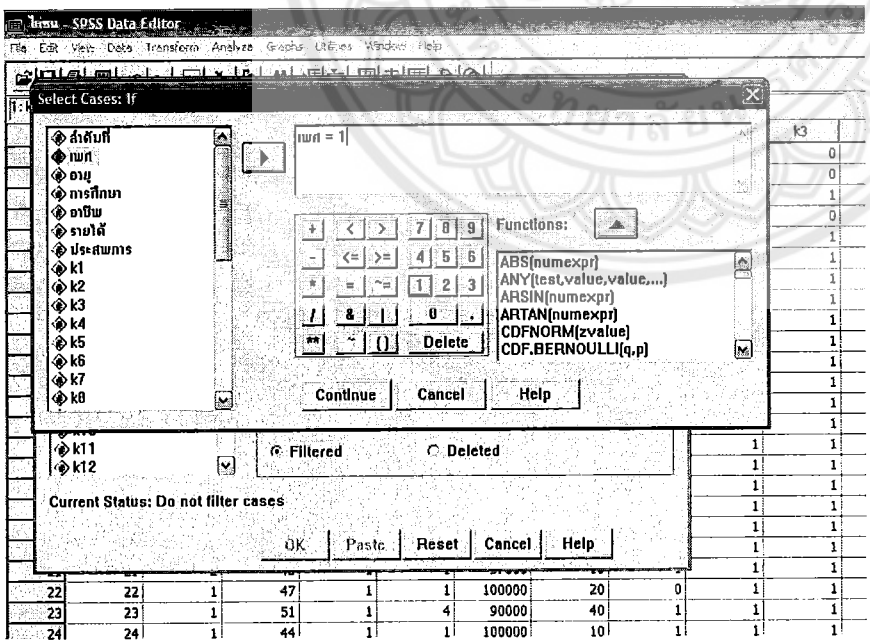


3.กด if และ OK

4.เลือก ตัวแปรที่ต้องการ มาไว้ที่ช่องขวามือ

5.เลือก เท่ากับ และเลขที่ต้องการในที่นี้เลือกเลข 1 ...ดูดังรูป

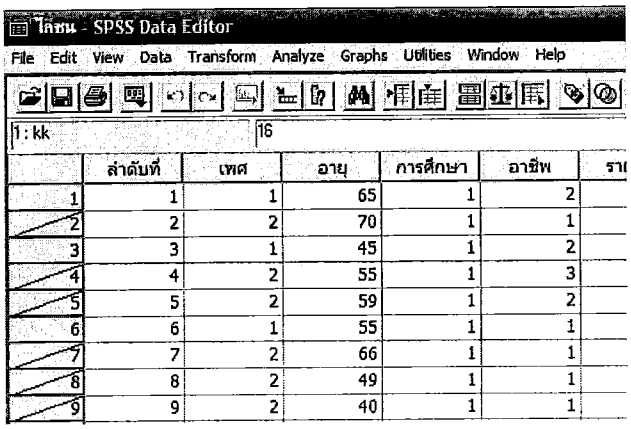
6. Continue และ ok



7. จะได้ดังรูป จะมีขีดๆ เลขที่เรากำหนดไว้ ขีดคือเลขที่ไม่ตรงกับที่เรากำหนดไว้

ห้องอ่านหนังสือคณะสาธารณสุขศาสตร์
มหาวิทยาลัยนเรศวร อ.เมือง จ.พิษณุโลก

WA20.5
ส581
2552
ฉ.2



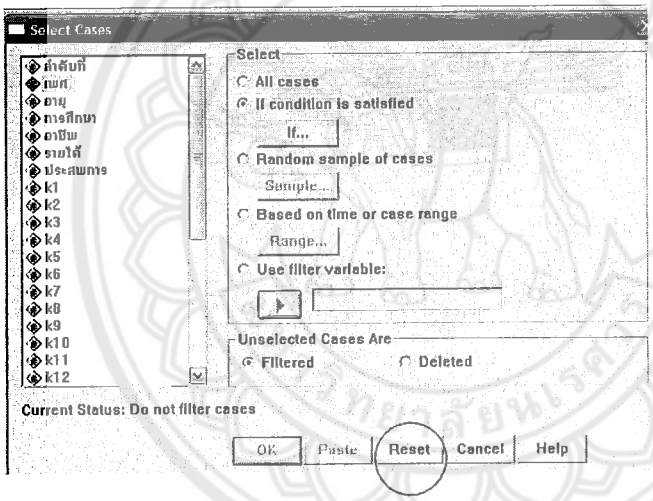
โสธน - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1 : kk 16

	ลำดับที่	เพศ	อายุ	การศึกษา	อาชีพ	ราย
1	1	1	65	1	2	
2	2	2	70	1	1	
3	3	1	45	1	2	
4	4	2	55	1	3	
5	5	2	59	1	2	
6	6	1	55	1	1	
7	7	2	66	1	1	
8	8	2	49	1	1	
9	9	2	40	1	1	

*** การยกเลิกการทำ select cases ....ทำตามขั้นตอนข้างต้น พอมาที่หน้าดังรูปให้เลือก reset

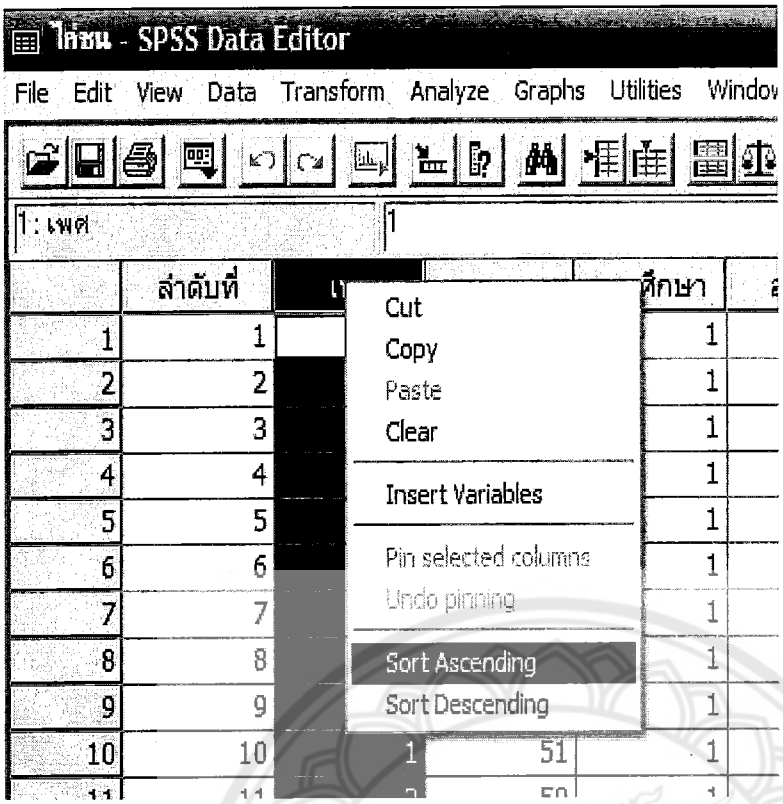


การใช้คำสั่ง Sort cases

เพื่อใช้ในการเรียงข้อมูล (จากข้อมูลที่มีค่ามากไปหาข้อมูลที่มีค่าน้อย และข้อมูลที่มีค่าน้อยไปหาข้อมูลที่มีค่ามาก)

1. เลือกตัวแปรที่ต้องการจัดเรียง คลิกให้เป็นแถบดำทั้งแถว
2. คลิกขวาเลือก
    2.1 sort Ascending น้อยไปมาก
    2.2 sort Descending มากไปน้อย

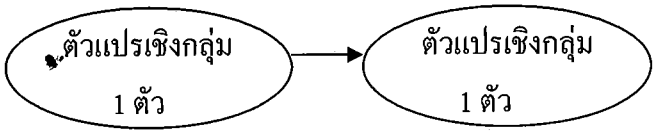
1.2788160


ไก่ชน - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Windov
1 : เพศ
1
ลำดับที่
ศึกษา
Cut
Copy
Paste
Clear
Insert Variables
Pin selected columns
Undo pinning
Sort Ascending
Sort Descending

การทดสอบความสัมพันธ์ไคสแควร์ Chi-square

เป็นการวิเคราะห์ความสัมพันธ์ของตัวแปร 2 ตัว โดยตัวแปรทั้งสองตัวแปรเป็นตัวแปรเชิงกลุ่ม

การทดสอบความสัมพันธ์ของตัวแปร 2 ตัว (อิสระ 1 , ตาม 1)



ขั้นตอนการทดสอบด้วยวิธี Chi-square

1. ทดสอบ Assumption
   - 1.1 ตัวแปรเป็นอิสระต่อกัน
   - 1.2 ค่าคาดหวัง (Ei) ในแต่ละ Cell เป็น 0 ไม่ได้
   - 1.3 ค่าคาดหวัง (Ei) ที่ < 5 ไม่ควรมีเกิน 20 % ของจำนวน Cell
2. ทดสอบสมมติฐาน
   - H0 : ตัวแปรอิสระและตัวแปรตามไม่มีความสัมพันธ์ต่อกัน (X=Y)
   - H1 : ตัวแปรอิสระและตัวแปรตามมีความสัมพันธ์ต่อกัน (X≠Y)
3. กำหนดระดับนัยสำคัญทางสถิติ
4. สร้างตารางไขว้ (Cross-tabulation) ระบุค่าคาดหวัง (Ei) ในแต่ละเซล % Column / Row และ % ยอดรวม
5. แปลผลค่าไคสแควร์
6. ผู้วิจัยจะปฏิเสธ H0 ก็ต่อเมื่อ ค่า Sig. ของการทดสอบไคสแควร์นั้นน้อยกว่าระดับนัยสำคัญที่กำหนด (แสดงว่าตัวแปรสัมพันธ์กัน)
7. การวิเคราะห์ขั้นต่อไป

   หลังจากปฏิเสธ H0 / ยอมรับ H1 แสดงว่าสองตัวแปรมีความสัมพันธ์กัน ผู้วิจัยควรเสนอดัชนีความสัมพันธ์เพื่อบอกปริมาณความสัมพันธ์ของสองตัวแปรด้วย

<u>กรณีที่ 1 ทั้งสองตัวแปรหรือตัวใดตัวหนึ่งเป็น Nominal</u>

- เสนอดัชนี Contingency, Phi, Lambda (0-1)

<u>กรณีที่ 2ทั้งสองตัวแปรเป็น Ordinal</u>

- เสนอดัชนี Gamma, Somers, Kendal's tau (-1 ถึง + 1)

จะนำ Exact Test ไปใช้ เมื่อใด

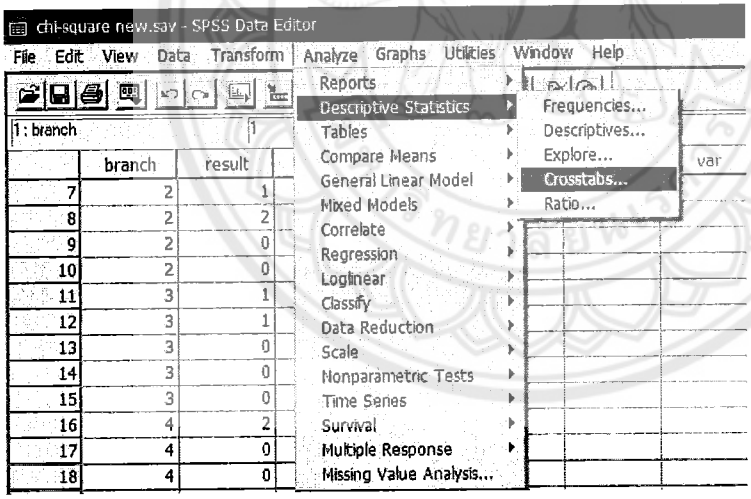
1. ค่า Ei (ค่าคาดหวัง) < 5 เกิน 20% ของจำนวน cell ทั้งหมด
2. ตารางขนาด 2 x 2 ?
3. ขนาดตัวอย่างน้อย
4. จำนวน row และ จำนวน column แตกต่างกันมาก เช่น 20 x3
5. มี cell ที่มีค่า Oi (ค่าสังเกต) เป็น 0 หลาย ๆ cell

<u>ตัวอย่าง</u> ขั้นตอนการทดสอบไคสแควร์

วิธีคำนวณค่า Chi-square จาก Spss

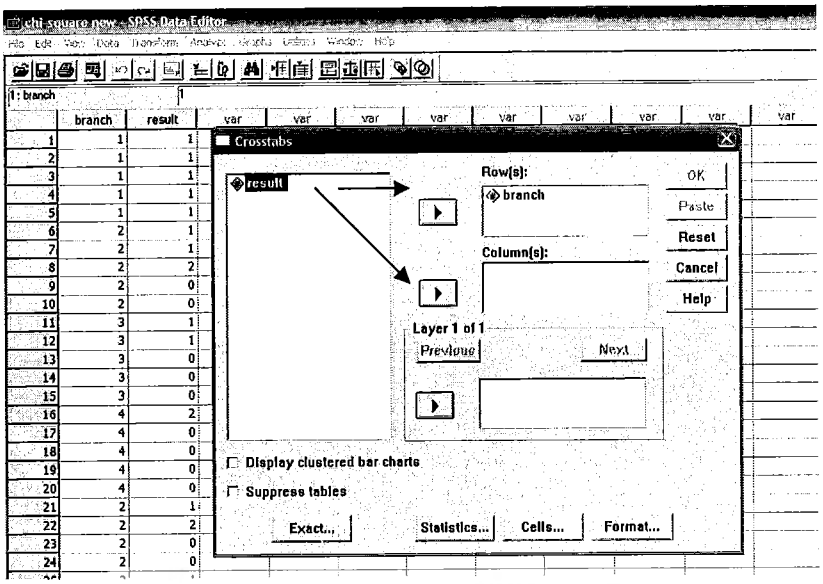
<u>ตัวอย่าง</u> ทดสอบความสัมพันธ์ระหว่างสาขาวิชาที่เรียนและผลการสอบเข้า N = 32

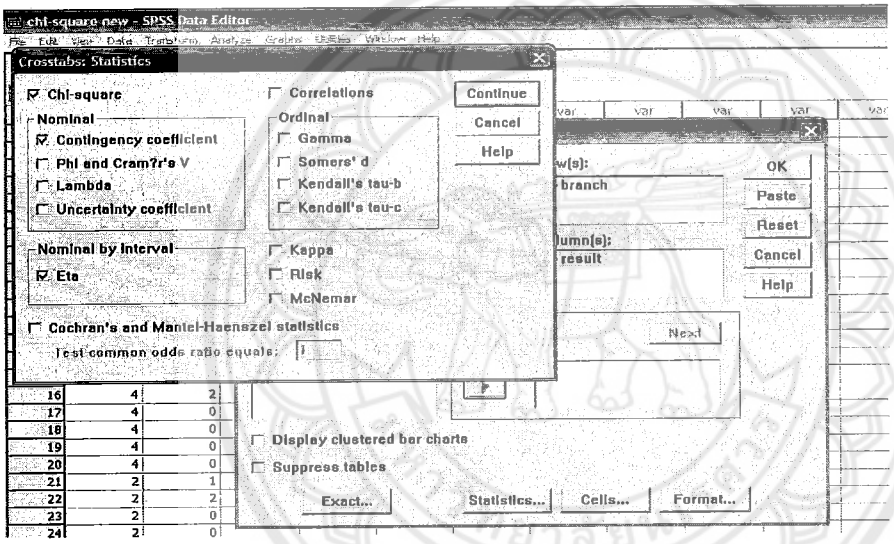
1. Analyze → Descriptive statistic → crosstabs



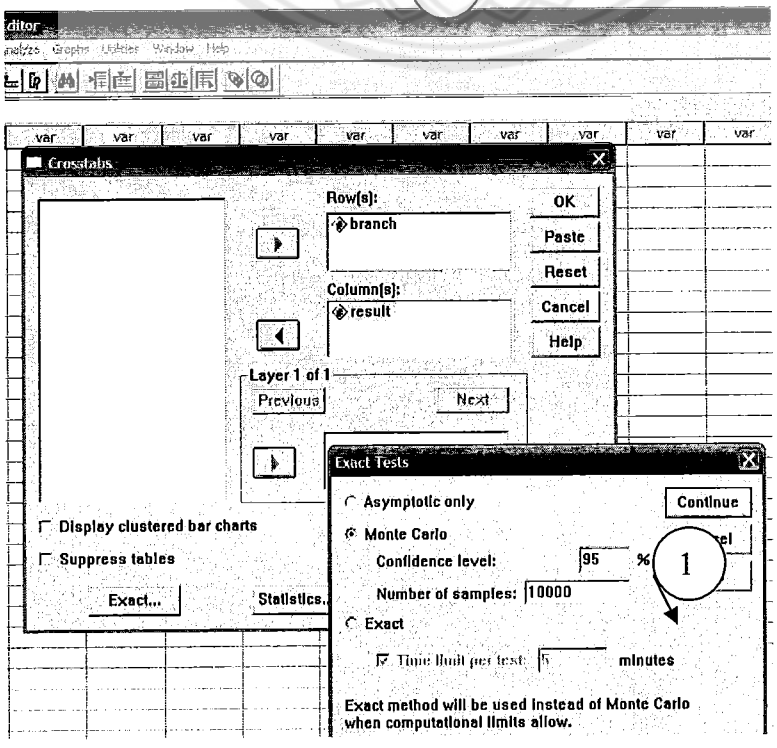
จะปรากฏหน้าจอดังภาพข้างล่าง

- เลือก branch ย้ายไปทางขวามือ (row)
- เลือก result ย้ายไปขวามือ (column)

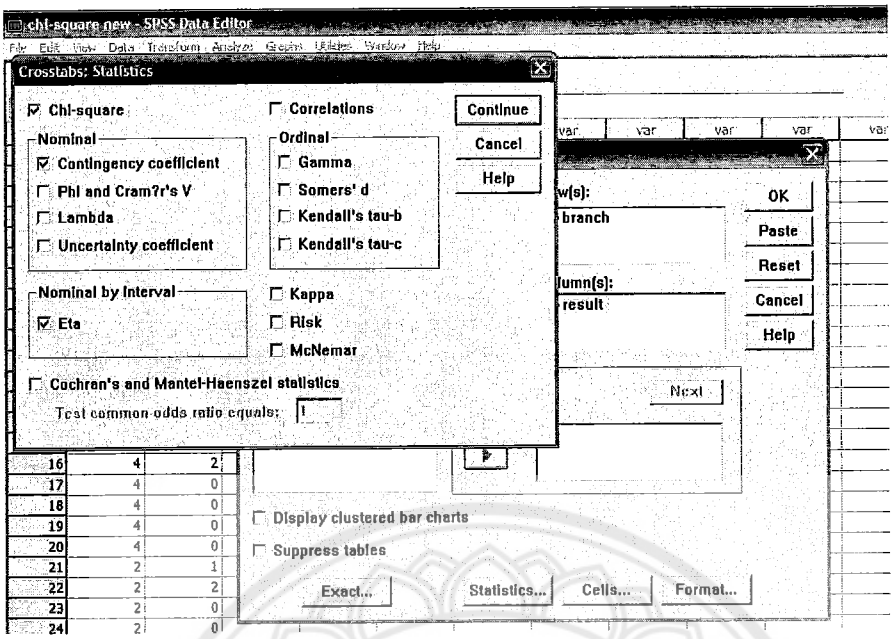




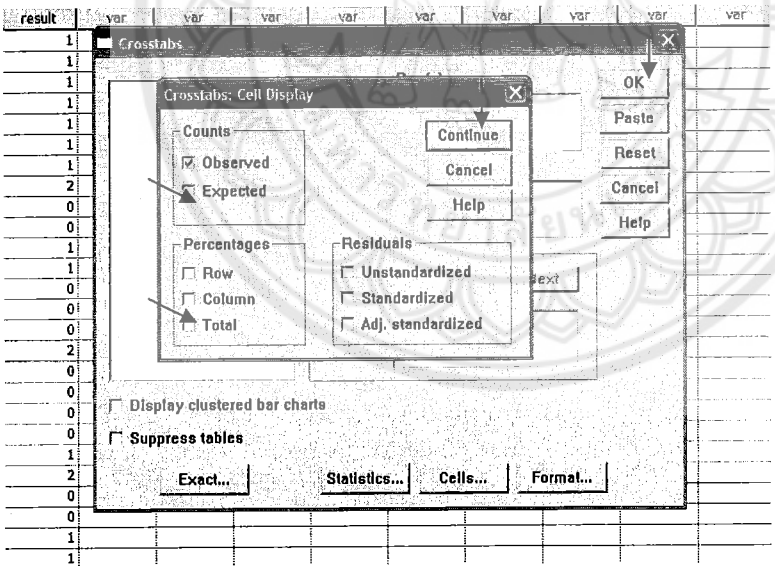
4. เลือกค่า สถิติ exact กำหนดค่านัยสำคัญทางสถิติ (1) และกด continue



5. Statistic เลือกค่าสถิติ chi – square, contingency coefficient (ค่าความสัมพันธ์) และกด continue



6. เลือก cells กำหนดให้การคำนวณแสดงค่า observed และ expected, total และกด continue และ OK



หน้าจอ out put จะแสดงผลดังนี้

→ **Crosstabs**

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
BRANCH * RESULT	32	100.0%	0	.0%	32	100.0%

BRANCH * RESULT Crosstabulation

			RESULT			Total
			ขาดสอบ	ผ่าน	ไม่ผ่าน	
BRANCH	รัฐศาสตร์	Count	0	5	0	5
		Expected Count	2.3	2.0	.6	5.0
		% of Total	.0%	15.6%	.0%	15.6%
	บัญชี	Count	5	4	3	12
		Expected Count	5.6	4.9	1.5	12.0
		% of Total	15.6%	12.5%	9.4%	37.5%
	วิศวกรรม	Count	3	4	0	7
		Expected Count	3.3	2.8	.9	7.0
		% of Total	9.4%	12.5%	.0%	21.9%
	คอม	Count	7	0	1	8
		Expected Count	3.8	3.3	1.0	8.0
		% of Total	21.9%	.0%	3.1%	25.0%
Total		Count	15	13	4	32
		Expected Count	15.0	13.0	4.0	32.0
		% of Total	46.9%	40.6%	12.5%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Monte Carlo Sig. (2-sided)			Monte Carlo Sig. (1-sided)		
				Sig.	95% Confidence Interval		Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound		Lower Bound	Upper Bound
Pearson Chi-Square	16.470[a]	6	.011	.009[b]	.007	.011			
Likelihood Ratio	21.336	6	.002	.003[b]	.002	.003			
Fisher's Exact Test	15.006			.005[b]	.004	.007			
Linear-by-Linear Association	4.670[c]	1	.031	.035[b]	.031	.038	.019[b]	.017	.022
N of Valid Cases	32								

a. 11 cells (91.7%) have expected count less than 5. The minimum expected count is .63.

b. Based on 10000 sampled tables with starting seed 2000000.

c. The standardized statistic is -2.161.

ดูว่าเป็นไปตามข้อตกลงเบื้องต้นของ chi-square หรือไม่ ถ้าไม่ ให้อ่านผล sig. ของ fisher's Exac

Symmetric Measures

		Value	Approx. Sig.	Monte Carlo Sig.		
				Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Nominal by Nominal	Contingency Coefficient	.583	.011	.009[c]	.007	.011
N of Valid Cases		32				

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on 10000 sampled tables with starting seed 2000000.

การแปรค่าระดับความสัมพันธ์

0.8-1 สัมพันธ์สูงมาก

0.7-0.79 สัมพันธ์กันมาก

0.3-0.69 สัมพันธ์ปานกลาง

0.2-0.29 สัมพันธ์น้อย

0-0.19 ไม่สัมพันธ์กันเลย

ตัวอย่าง จากตาราง 1-100 ข้อมูล เชิงกลุ่ม 2 ตัวคือ เพศ แทนตัวแปรต้น (X) และ ตำแหน่งงาน แทนตัวแปรตาม(Y)

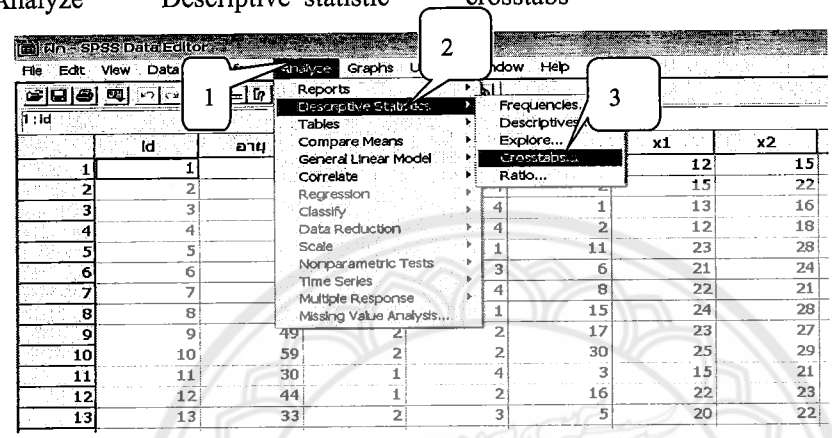
1. ตั้งสมมติฐาน

   H0 : เพศไม่มีความสัมพันธ์กับตำแหน่งงาน

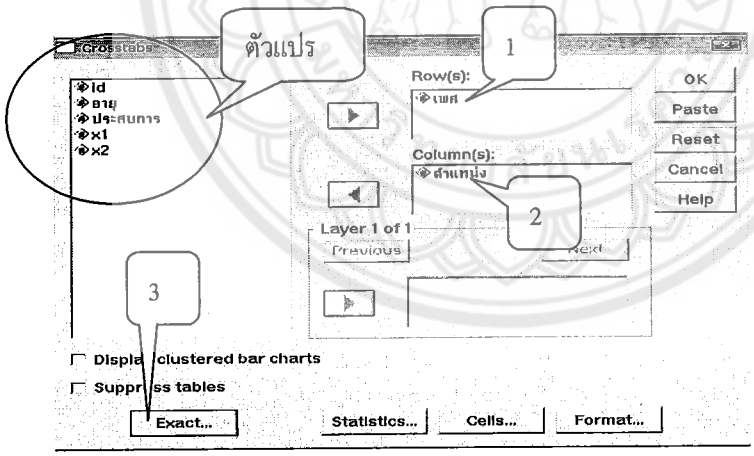
   H_1 : เพศมีความสัมพันธ์กับตำแหน่งงาน

2. กำหนดค่า ความคลาดเคลื่อน = 0.05

3. Analyze → Descriptive statistic → crosstabs

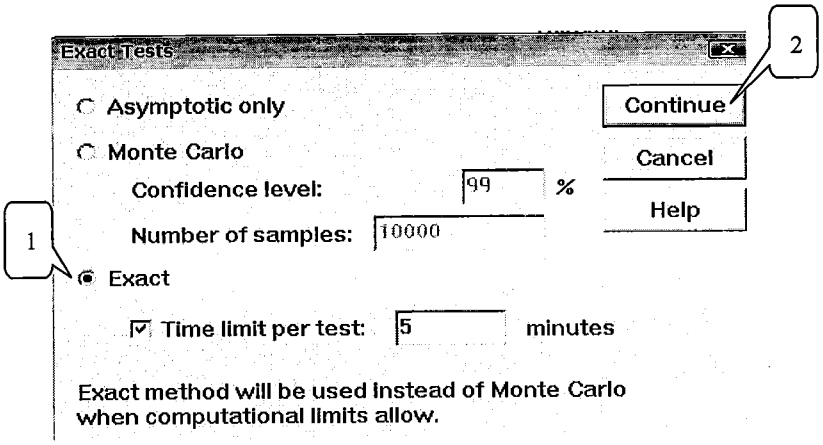


1. เลือกเมนู Analyze
2. เลือกคำสั่ง Descriptive Statistics
3. เลือกคำสั่ง Crosstabs...



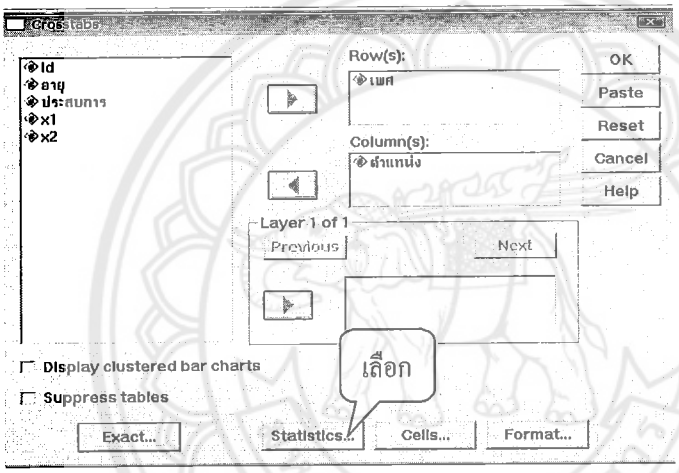
ที่หน้าต่าง Crosstabs จะพบตัวแปรทั้งหมด ในช่องซ้ายมือ แล้วทำการเลือกตัวแปรที่ต้องการหาความสัมพันธ์

1. เลือกตัวแปร เพศ มาไว้ในหน้าต่าง Row(s)
2. เลือกตัวแปร ตำแหน่ง มาไว้ในหน้าต่าง Column(s)
3. เลือก Exact...

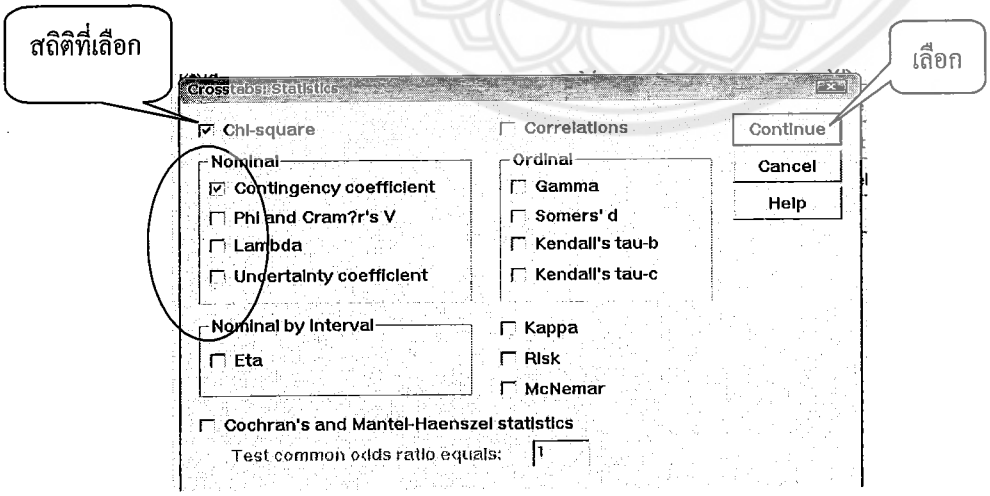


ที่หน้าต่าง Exact tests

1. เลือก Exact
2. เลือก Continue

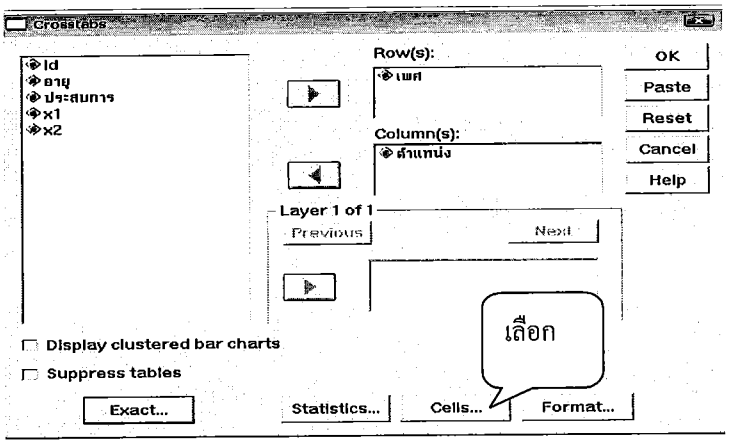


จะกลับมายัง หน้าต่าง Crosstabs เลือกที่ Statistics...

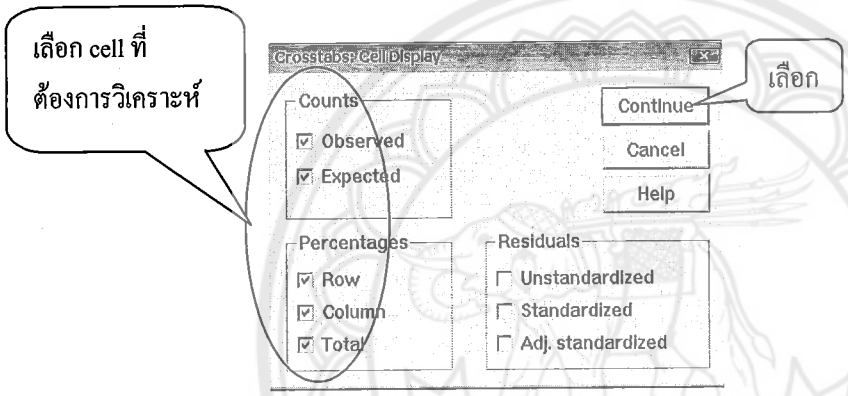


พบหน้าต่าง Crosstabs : Statistics

เลือกสถิติที่ต้องการนำมาวิเคราะห์ ได้แก่ Chi-square และ Contingency coefficient แล้ว เลือก Continue

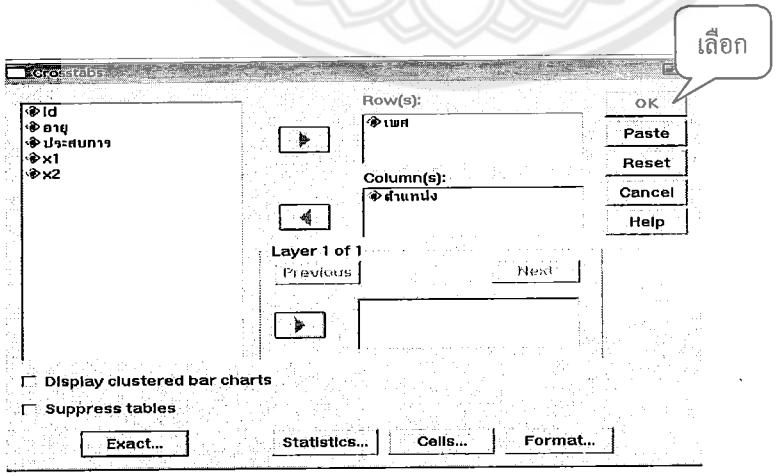


จะกลับมายัง หน้าต่าง Crosstabs อีกครั้ง ให้เลือกที่ cells... เพื่อเลือกว่าจะนำข้อมูล จาก cell ใดบ้างมาวิเคราะห์



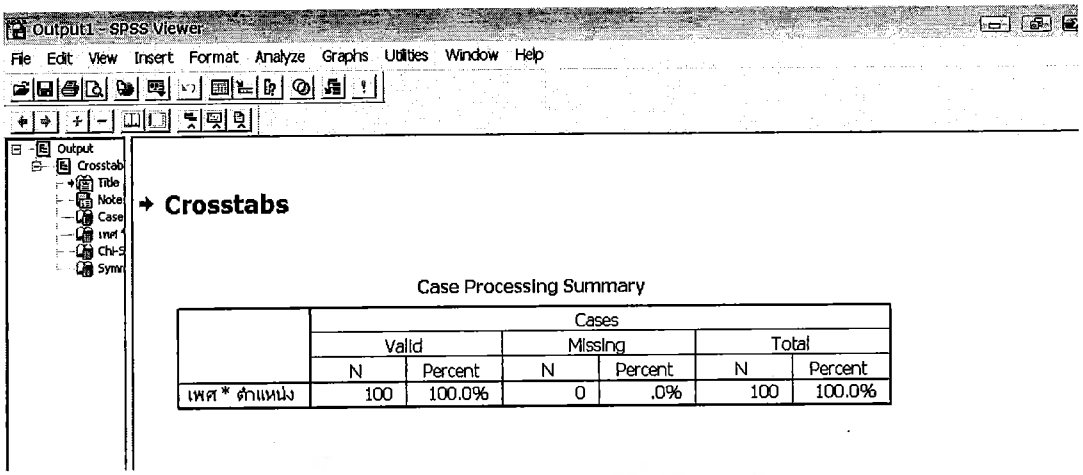
จะพบหน้าต่าง crosstabs : Cell Display

1. เลือก cell ที่ต้องการวิเคราะห์ (เลือกทั้งหมด)
2. เลือก Continue



เมื่อกลับมาที่หน้าต่าง Crosstabs ครั้งนี้ให้ เลือก OK เพื่อวิเคราะห์ข้อมูลได้เลย

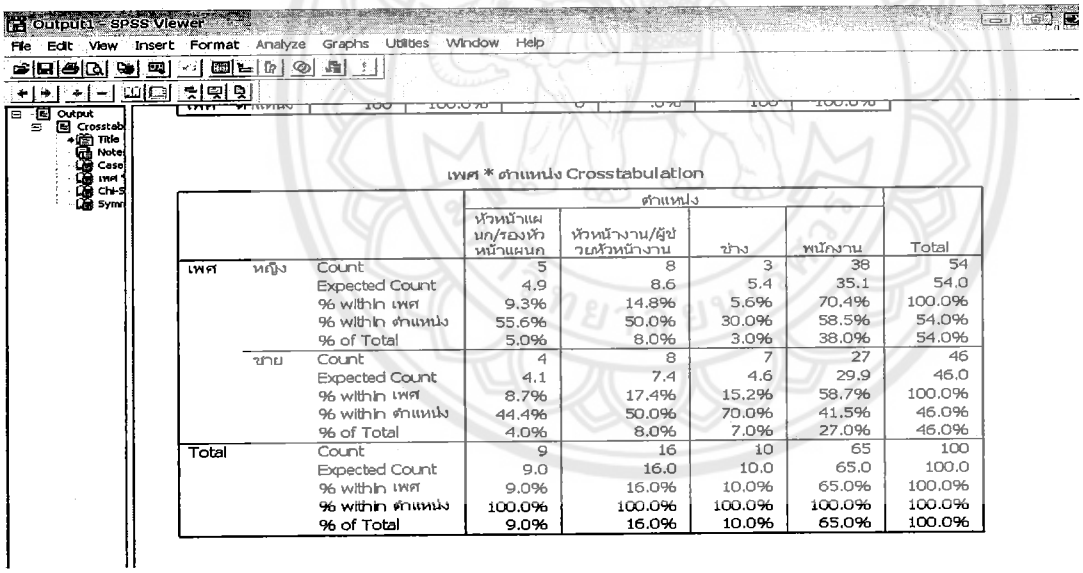
หน้าต่าง output



Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
เพศ * ตำแหน่ง	100	100.0%	0	.0%	100	100.0%

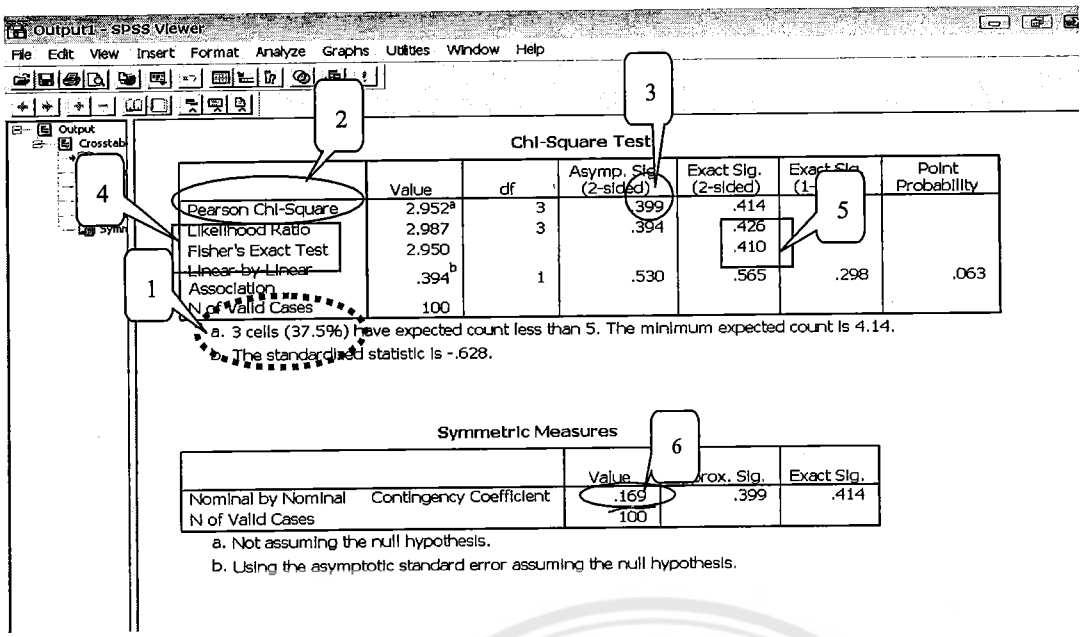
สามารถดูรายละเอียด ตัวแปรที่นำมาหาความสัมพันธ์ โดยมีรายละเอียดต่าง ๆ ตามที่เราเลือกไว้ใน Output ดังรูปข้างบนนี้ ตารางนี้ จะบอกว่าเราจะหาความสัมพันธ์ ระหว่างตัวแปรใดบ้าง ในที่นี้ คือ ตัวแปร เพศ กับ ตัวแปร ตำแหน่ง



เพศ * ตำแหน่ง Crosstabulation

			ตำแหน่ง				
			หัวหน้าแผนก/รองหัวหน้าแผนก	หัวหน้างาน/ผู้ช่วยหัวหน้างาน	ช่าง	พนักงาน	Total
เพศ	หญิง	Count	5	8	3	38	54
		Expected Count	4.9	8.6	5.4	35.1	54.0
		% within เพศ	9.3%	14.8%	5.6%	70.4%	100.0%
		% within ตำแหน่ง	55.6%	50.0%	30.0%	58.5%	54.0%
		% of Total	5.0%	8.0%	3.0%	38.0%	54.0%
	ชาย	Count	4	8	7	27	46
		Expected Count	4.1	7.4	4.6	29.9	46.0
		% within เพศ	8.7%	17.4%	15.2%	58.7%	100.0%
		% within ตำแหน่ง	44.4%	50.0%	70.0%	41.5%	46.0%
		% of Total	4.0%	8.0%	7.0%	27.0%	46.0%
Total		Count	9	16	10	65	100
		Expected Count	9.0	16.0	10.0	65.0	100.0
		% within เพศ	9.0%	16.0%	10.0%	65.0%	100.0%
		% within ตำแหน่ง	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	9.0%	16.0%	10.0%	65.0%	100.0%

ตาราง Crosstabulation จะแสดงการคำนวณหาค่า ความคาดหวัง (Expected) ของตัวแปร ในแต่ละเซล



ใน Output ส่วนของ Chi-Square Tests จะบอกรายละเอียดจากการหาความสัมพันธ์ด้วยสถิติต่าง ๆ

1. แสดงค่า คาดหวัง (Expected) ที่มีค่าน้อยกว่า 5 ในแต่ละเซล (มี 3 เซล) ร้อยละ 37.5% ซึ่งมากกว่า 20% จึงใช้สถิติ Chi-Square Tests ไม่ได้
2. แสดงสถิติ Chi-Square Tests กรณีที่ใช้สถิตินี้ได้ โดยอ่านผลที่ช่อง Asymp.sig (3)
3. กรณีที่ใช้สถิติ Fisher's Exact test ให้อ่านผลที่ช่อง Exact.sig (4)

การอ่านผล ให้เทียบกับค่า นัยสำคัญ α หรือระดับความเชื่อมั่น (1 - α)ที่ตั้งไว้ ส่วนใหญ่ คือ 0.05

ค่า df ไม่ต้องอ่าน จาก การวิเคราะห์ด้วยคอมพิวเตอร์เพราะไม่ต้องเอาไปเปิดตาราง Chi-Square ตาราง Symmetric Measues จะบอกถึงระดับความสัมพันธ์ของตัวแปร ว่ามีมากหรือน้อย มีค่า .169 แสดงว่ามีความสัมพันธ์กัน (เพียงเล็กน้อย)

Tip : การแปรความ 0 - 1 มีความสัมพันธ์ทางบวก

0 ไม่มีความสัมพันธ์

-1 – 0 มีความสัมพันธ์ทางลบ

<u>สรุปการแปรผลข้อมูล</u>

ค่า sig . ของ Fisher's Exact test = 0.41 ซึ่งมากกว่า 0.05 จึงปฏิเสธ H1 ยอมรับ H0 แสดงว่าเพศไม่มีความสัมพันธ์กับตำแหน่งงานที่ระดับนัยสำคัญทางสถิติที่ 0.05 โดยมีความสัมพันธ์ในระดับน้อยมาก(r= 0.169)

T-Test

กระบวนการทางสถิติ T-Test เป็นการเปรียบเทียบค่าเฉลี่ย 2 ค่า นอกจากนั้นยังแสดงค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานในแต่ละตัวนำเสนอ 3 ข้อดังนี้

1. One-sample T-Test การวิเคราะห์เปรียบเทียบค่าเฉลี่ยของกลุ่มตัวอย่างกับประชากร (กลุ่มควบคุมหรือค่าเกณฑ์มาตรฐานที่กำหนด)
2. Independent T-Test การทดสอบสมมติฐานของกลุ่มตัวอย่าง 2 กลุ่มที่เป็นอิสระจากกัน
3. Dependent T-Test หรือ Paired T-Test การทดสอบสมมติฐานของกลุ่มตัวอย่าง 2 กลุ่มที่สัมพันธ์กัน

One Sample T-Test

การวิเคราะห์เปรียบเทียบค่าเฉลี่ยของกลุ่มตัวอย่างกับประชากร (กลุ่มควบคุมหรือค่าเกณฑ์มาตรฐานที่กำหนด) One-sample T-Test

สมมติฐานแบบไม่มีทิศทาง

H_0 : $\mu = 0$

H_1 : $\mu \neq 0$

สมมติฐานแบบมีทิศทาง

H_0 : $\mu \leq 0$ H_0 : $\mu \geq 0$

H_1 : $\mu > 0$ H_1 : $\mu < 0$

การทดสอบข้อตกลงเบื้องต้น ทดสอบว่าข้อมูลมีการแจกแจงแบบปกติหรือไม่

H_0 : ข้อมูลมีการแจกแจงแบบปกติ

H_1 : ข้อมูลมีการแจกแจงไม่ปกติ

ใช้เมนู Analyae→Descriptive statistic→Explore→ เข้าหน้าต่าง Dependent เลือกตัวแปร →Plot เลือก Histogram และ Normality plots

กรณีกลุ่มตัวอย่างมากกว่า 50 ขึ้นไปให้อ่านค่า Sig ที่ช่อง Kolmogorov-Smirnov(a)

กรณีกลุ่มตัวอย่างน้อยกว่า 50 ขึ้นไปให้อ่านค่า Sig ที่ช่อง Shapiro-Wilk

ดังตาราง

Tests of Normality

	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
รายได้	.088	100	**.052**	.972	100	**.031**

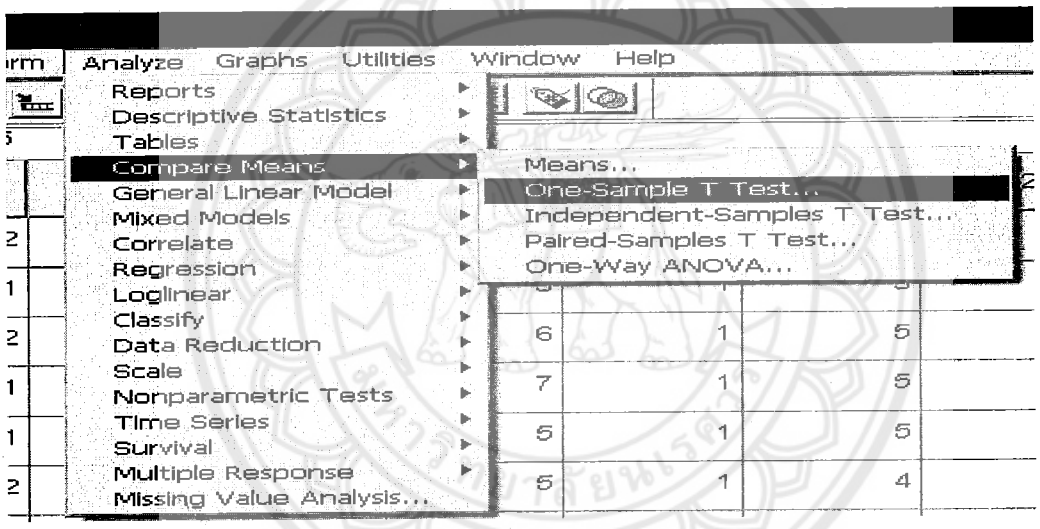
a Lilliefors Significance Correction

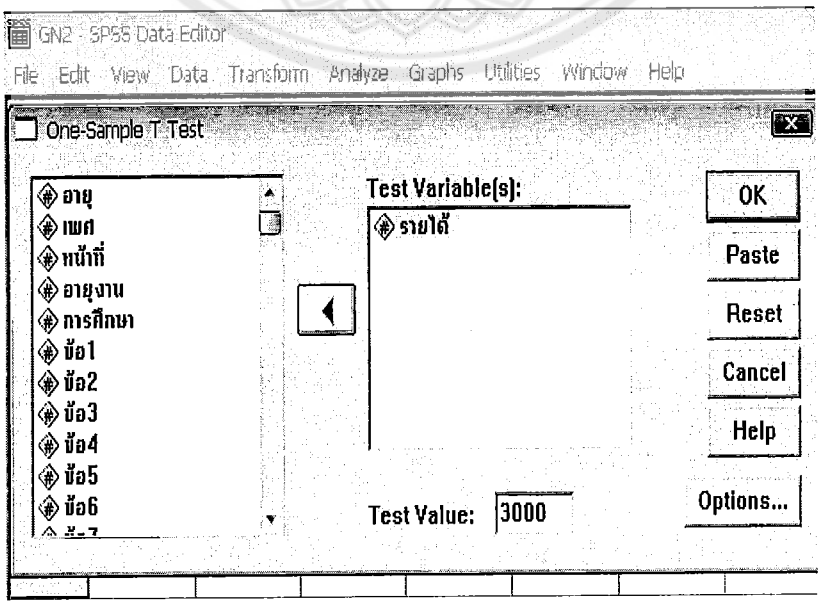
ตัวอย่างเช่น รายได้แท็กซี่ในกทม.มากกว่า 3000 บาท

H_0 : รายได้เฉลี่ยของแท็กซี่ในกทม.น้อยกว่าหรือเท่ากับ 3000 บาท

H_1 : รายได้เฉลี่ยของแท็กซี่ในกทม.มากกว่า 3000 บาท

ใช้เมนูหลัก Analyze → Compare means → One-sample T-Test จะปรากฏหน้าต่าง One-sample T-Test เลือกตัวแปรที่ต้องการใส่ช่อง Test Variable และพิมพ์ค่าเกณฑ์ที่กำหนดหรือค่ามาตรฐานใส่ช่อง Test Value ดังภาพ





ผลการวิเคราะห์ (Output)

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
รายได้	100	8571.48	451.371	45.137

One-Sample Test

	Test Value = 3000					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
รายได้	123.435	99	.000	5571.48	5481.92	5661.04

ตารางแรกแสดงค่าสถิติพื้นฐานของตัวแปร และตารางที่สองจะแสดงค่าสถิติ T-Test สามารถนำเสนอผลการวิเคราะห์ได้ดังตารางต่อไปนี้

ค่าเฉลี่ยกลุ่มตัวอย่าง	ส่วนเบี่ยงเบนมาตรฐาน	t	sig
8571.48	451.371	123.435	0.000

ผลการทดสอบปรากฏว่า ค่า sig มีค่าเท่ากับ 0.00 น้อยกว่าระดับนัยสำคัญที่ 0.05 แสดงว่ายอมรับ H_1 รายได้เฉลี่ยของแท็กซี่ในกทม.มากกว่า 3000 บาท อย่างมีนัยสำคัญทางสถิติ

Independent T-Test

เป็นการทดสอบสมมติฐานของกลุ่มตัวอย่าง 2 กลุ่มที่เป็นอิสระจากกัน Independent T-Test ในการทดสอบสมมติฐานที่ต้องการหาความแตกต่างของค่าเฉลี่ยของกลุ่มตัวอย่างสองกลุ่มว่าแตกต่างกันหรือไม่ ในกรณีที่กลุ่มตัวอย่าง 2 กลุ่มเป็นอิสระจากกัน สามารถตั้งสมมติฐานได้ดังนี้

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

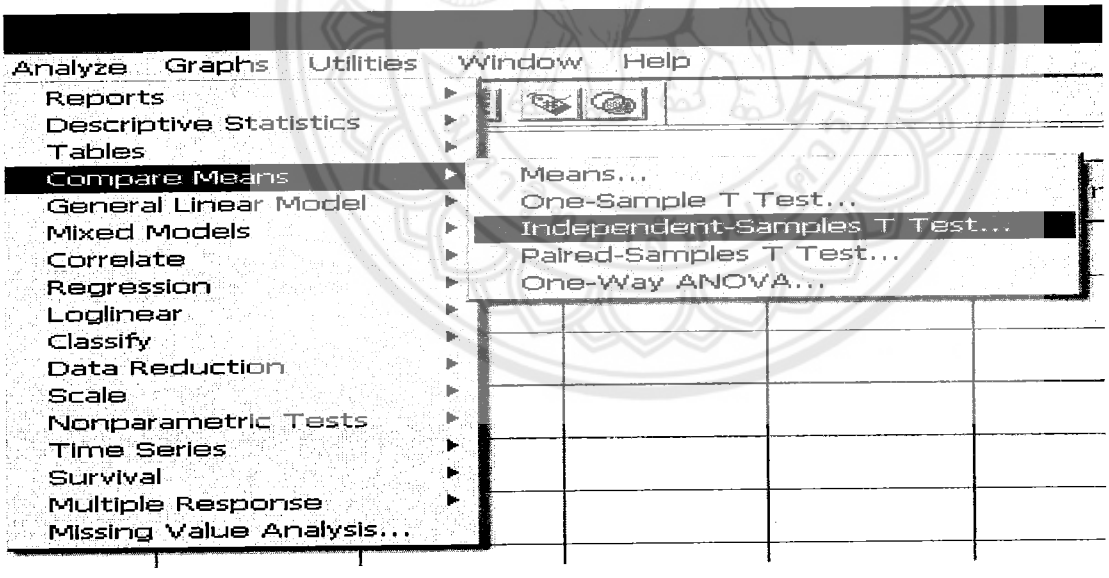
หรือ

$H_0 : \mu_1 - \mu_2 = 0$

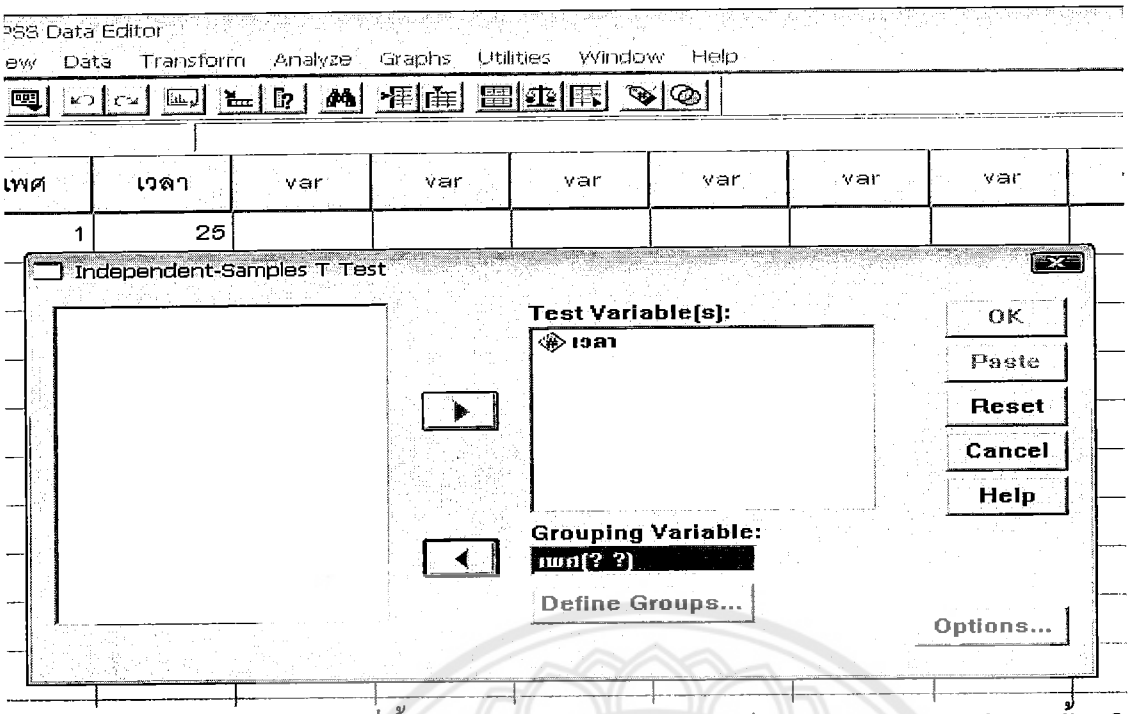
$H_1 : \mu_1 - \mu_2 \neq 0$

ตัวอย่างเช่น ต้องการทดสอบว่าเวลาที่ใช้ในการทานอาหารระหว่างเพศชายและเพศหญิงมีความแตกต่างกันหรือไม่

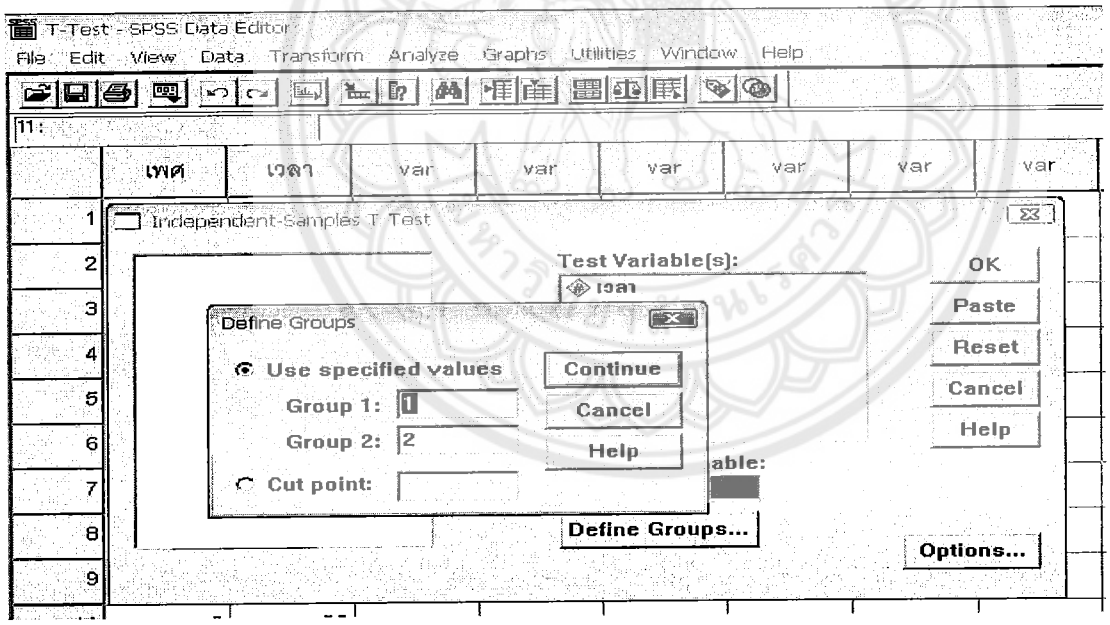
การวิเคราะห์ T-Test จะใช้ Menu หลัก Analyze → Compare means → Independent sample T-Test จะปรากฏหน้าต่าง Independent sample T-Test ดังภาพ



เลือกตัวแปรตามใส่ช่อง Test Variable และตัวแปรอิสระใส่ช่อง Grouping Variable ในที่นี้ตัวแปรตามคือเวลา ตัวแปรต้นคือ เพศ จากนั้นคลิกปุ่ม Define groups จะปรากฏหน้าต่างดังนี้



ใส่รหัสของตัวแปร ในที่นี้เราใช้รหัส 1 แทนเพศชาย ใช้รหัส 2 แทนเพศหญิงจากนั้นคลิกที่ปุ่ม Continue แล้วคลิกปุ่ม OK จะแสดงผลในหน้าต่าง Output



Group Statistics

	เพศ	N	Mean	Std. Deviation	Std. Error Mean
เวลา	หญิง	16	23.63	1.544	.386
	ชาย	14	23.36	1.737	.464

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
เวลา	Equal variances assumed	.128	.724	.447	28	.658	.27	.599	-.959	1.494
	Equal variances not assumed			.444	26.296	.661	.27	.604	-.972	1.508

การทดสอบความแปรปรวน

H_0 : ความแปรปรวนของเพศชายและหญิงเท่ากัน

H_1 : ความแปรปรวนของเพศชายและหญิงไม่เท่ากัน

จากตารางให้อ่านค่า sig ของ F-Test ในที่นี้คือ 0.724 มีค่ามากว่าระดับนัยสำคัญที่ 0.05 แสดงว่ายอมรับ H_0 นั่นคือความแปรปรวนของเพศชายและหญิงเท่ากัน ให้อ่านค่า sig ของ t-test แถวบน Equal variances assumed คือ 0.658

9 ไม่ได้ทดสอบ การแจกแจงปกติของข้อมูลเวลา ในกลุ่มเพศชาย และหญิง

Paired-T Test

การทดสอบสมมติฐานของกลุ่มตัวอย่าง 2 กลุ่มที่สัมพันธ์กัน Dependent T-Test ในการเปรียบเทียบค่าเฉลี่ย 2 ค่าว่ามีความแตกต่างกันหรือไม่ โดยค่าเฉลี่ย 2 ค่านี้วัดมาจากกลุ่มตัวอย่าง 2 กลุ่มที่สัมพันธ์กัน โดยอาจวัดมาจากกลุ่มตัวอย่างกลุ่มเดียวกัน 2 ครั้งหรือวัดมาจากกลุ่มตัวอย่าง 2 กลุ่มที่ได้มาจากการจับคู่คุณลักษณะที่เท่าเทียมกันมีวิธีการคำนวณหาความแตกต่างของค่าเฉลี่ยดังนี้

สมมติฐาน

H_0 : $\mu_1 - \mu_2 = 0$ $\mu_d = 0$

H_1 : $\mu_1 - \mu_2 \neq 0$ $\mu_d \neq 0$

ตัวอย่างเช่น การทดสอบความแตกต่างของการอบรมว่าทำให้ผู้รับการอบรมมีความรู้เพิ่มขึ้นหรือไม่

สมมติฐานทางวิจัย

H_0 : คะแนนเฉลี่ยก่อนและหลังการอบรมไม่ต่างกัน

H_1 : คะแนนเฉลี่ยก่อนและหลังการอบรมต่างกัน

ใช้เมนูหลัก Analyze → Compare means → Paired-sample T-Test จะปรากฏหน้าต่าง Paired-sample T-Test ดังภาพจากนั้นให้เลือกคู่ตัวแปรที่ต้องการทดสอบให้อยู่ในช่อง Paired Variable แล้วคลิก OK จะได้ผลการวิเคราะห์ดังตาราง

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PRE	13.05	37	2.235	.367
	POST	13.68	37	2.310	.380

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	PRE & POST	37	.154	.362

9 ไม่ได้ทดสอบการแจกแจงแบบปกติ

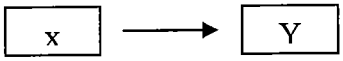
Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1 PRE - POST	-.62	2.957	.486	-1.61	.36	-1.279	36	.209

ตารางแรกเป็นการนำเสนอค่าสถิติพื้นฐานของทั้งสองตัวแปรตารางที่สองแสดงค่าสหสัมพันธ์ระหว่าง 2 ตัวแปรและตารางที่สามจะแสดงค่าสถิติ T-Test จากตัวอย่างค่า Sig มีค่าเท่ากับ 0.209 มีค่ามากกว่าระดับนัยสำคัญที่ 0.05 แสดงว่ายอมรับ H_0 นั่นคือ คะแนนเฉลี่ยก่อนและหลังการอบรมไม่ต่างกัน

การวิเคราะห์การถดถอย (Regression Analysis)

Regression

- เป็นการวิเคราะห์หาความสัมพันธ์ระหว่างตัวแปรอิสระหรือตัวแปรต้น กับ ตัวแปรตาม 

- จะนำเสนอผลการวิเคราะห์ในรูปแบบของสมการ

จุดประสงค์ของ Regression Analysis

เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปร ว่ามีความสัมพันธ์กันมากน้อยเพียงใด (ถ้า X และ Y มีความสัมพันธ์กันมากแสดงว่าถ้า X มีค่าเปลี่ยนแปลงไปจะมีผลกระทบต่อค่าของ Y เป็นอย่างมาก)
-ใช้ความสัมพันธ์ที่วิเคราะห์ได้ มาประมาณค่า หรือ พยากรณ์ ค่า Yในอนาคต

รูปแบบของRegression Analysis

- Simple Regression
- Multiple Regression
- Regression with Dummy Variable

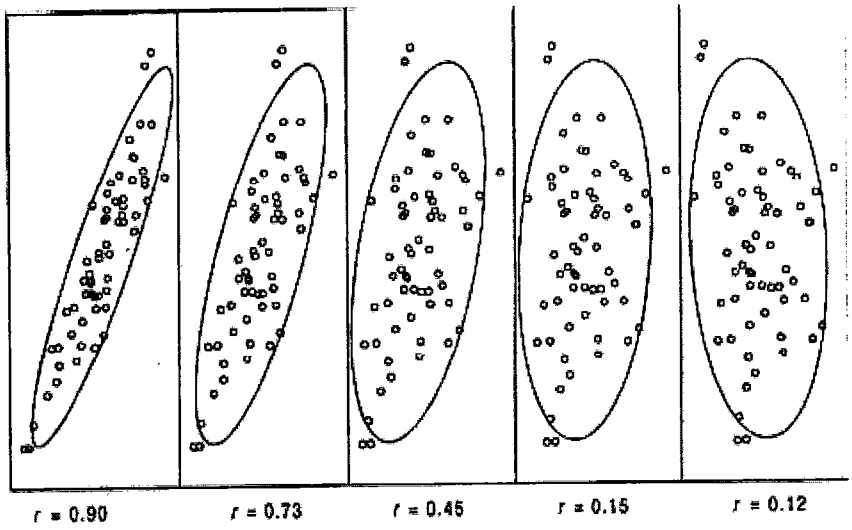
ค่าสถิติที่ควรรู้

1. ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)

•เป็นค่าที่บอกความสัมพันธ์ของตัวแปร X กับ Y
•สัญลักษณ์ที่ใช้คือ r_{xy} หรือ r
•ค่าที่คำนวณได้อยู่ในช่วงระหว่าง -1 ถึง 1

ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient = r)

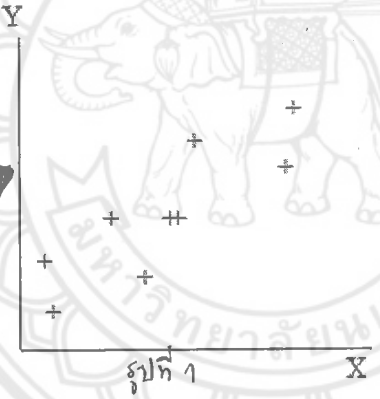
ถ้า r เป็น + มีค่าเข้าใกล้ 1 หมายถึงตัวแปร X และ Y มีความสัมพันธ์กันมาก ในรูปเชิงเส้น ในทิศทางเดียวกัน (ตัวแปรหนึ่งเพิ่ม อีกตัวแปรจะเพิ่มขึ้นด้วย)
ถ้า r เป็น - มีค่าเข้าใกล้ -1 หมายถึงตัวแปร X และ Y มีความสัมพันธ์กันมาก ในรูปเชิงเส้นแต่ทิศทางตรงข้าม (ตัวแปรหนึ่งเพิ่มขึ้น อีกตัวแปรหนึ่งจะลดลง)
ถ้า r มีค่าใกล้ 0 แสดงว่าตัวแปร X และ Y มีความสัมพันธ์กันน้อย
ถ้า r = 0 แสดงว่าตัวแปร X และ Y ไม่มีความสัมพันธ์กันเลย



การทดสอบความสัมพันธ์เชิงเส้นของตัวแปร ด้วยแผนภาพกระจัดกระจาย (Scatter Plot)

ความสัมพันธ์ทางเดียวกัน คือ เมื่อตัวแปร X มีค่าต่ำ ตัวแปร Y ก็มีค่าต่ำด้วย และเมื่อตัวแปร X มีค่าสูง ตัวแปร Y ก็มีค่าสูง เช่นเดียวกัน เมื่อตัวแปรมีความสัมพันธ์ทางเดียวกัน ลักษณะของกราฟจะ เป็นดังรูป

รูปแบบความสัมพันธ์ทางบวก



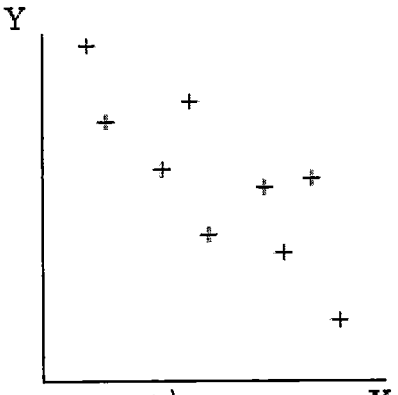
รูปที่ 1

การแสดงความสัมพันธ์ระหว่างตัวแปรต่อเนื่อง โดยการใช้แผนภูมิ

•**กำหนดให้ตัวแปรอิสระอยู่ในแนวนอน (แกน X) และตัวแปรตามอยู่ในแกนตั้ง (แกน Y)**

<u>รูปแบบความสัมพันธ์ทางลบ</u>

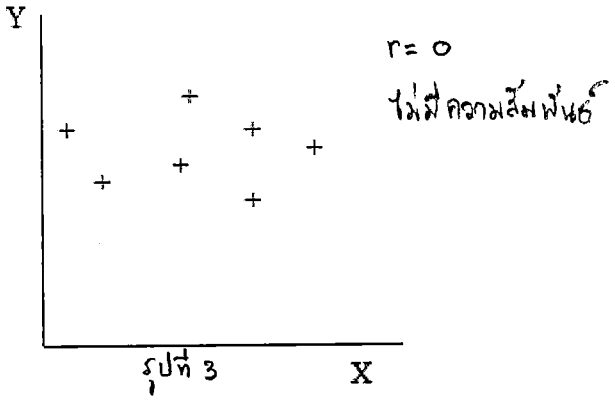
ความสัมพันธ์ทางเดียวกัน คือเมื่อตัวแปร X มีค่าต่ำ ตัวแปร Yก็มีค่าต่ำด้วย และเมื่อตัวแปร Xมีค่าสูง ตัวแปร Y ก็มีค่าสูง เช่นเดียวกัน เมื่อตัวแปรมีความสัมพันธ์ทางเดียวกัน ลักษณะของกราฟจะเป็นดังรูป (รูปที่ 1)



รูปที่ 2

รูปแบบแสดงความไม่สัมพันธ์

ความสัมพันธ์ผกผัน คือ เมื่อตัวแปร X มีค่าต่ำ ตัวแปร Yจะมีค่าสูง และเมื่อตัวแปร Xมีค่าสูง ตัวแปร Y จะมีค่าต่ำ
- เมื่อตัวแปรมีความสัมพันธ์ทางผกผัน ลักษณะของกราฟจะเป็น ดังรูป (รูปที่ 2)



2. ค่าสัมประสิทธิ์การตัดสินใจ (The Coefficient of Determination)

เกิดจากการนำค่าสัมประสิทธิ์สหสัมพันธ์มายกกำลังสองใช้แสดงความแปรผันที่เกิดขึ้นกับตัวแปร Y มีผลเนื่องมาจากตัวแปร X คิดเป็นกี่เปอร์เซ็นต์ ใช้ศึกษาว่า สมการการประมาณค่ามีความเหมาะสมที่จะนำไปใช้ได้มากหรือน้อย ค่าที่คำนวณได้จะอยู่ในช่วงระหว่าง 0 ถึง 1 สัญลักษณ์ที่ใช้คือ r^2

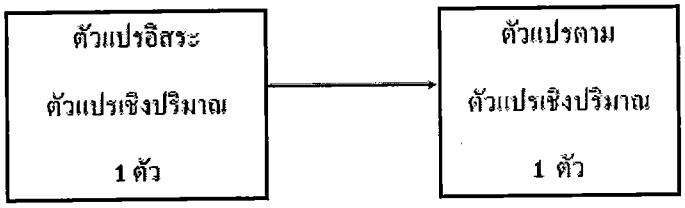
-ในกรณีที่ค่า r^2 มีค่าเข้าใกล้ 1 แสดงว่าตัวแปร X มีอิทธิพลต่อตัวแปร Y อย่างมาก หมายความว่า สมการการประมาณค่าจะมีความเหมาะสมที่จะนำไปใช้งานได้มาก

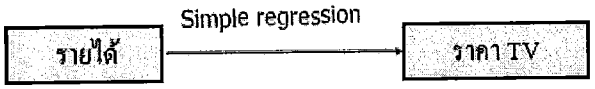
-ในกรณีที่ค่า r^2 มีค่าเข้าใกล้ 0 แสดงว่า ตัวแปร X มีอิทธิพลต่อตัวแปร Y น้อยมาก หมายความว่า สมการการประมาณค่าจะมีความเหมาะสมที่จะนำไปใช้งานได้น้อย

คุณสมบัติของ R2

- R2 จะไม่มีหน่วย
- ถ้า R2 มีค่าเข้าใกล้ 1 แสดงว่าเปอร์เซ็นต์ที่ X สามารถอธิบายการเปลี่ยนแปลงของ Y มีค่ามาก หรือ X และ Y มีความสัมพันธ์กันมาก
- ถ้า R2 มีค่าเข้าใกล้ 0 แสดงว่าเปอร์เซ็นต์ที่ X สามารถอธิบายการเปลี่ยนแปลงของ Y มีค่าน้อย หรือ X และ Y มีความสัมพันธ์กันน้อย
- ในการวิเคราะห์ความถดถอย อาจใช้ค่า Adjusted R2

Simple Regression





Assumption

1. กรณีที่ X และ Y มีความสัมพันธ์เชิงเส้นตรง
2. ค่าแปรปรวนของค่าคลาดเคลื่อนมีค่าคงที่
3. ค่าคลาดเคลื่อนต้องเป็นอิสระกัน
4. ค่าคลาดเคลื่อนต้องมีการแจกแจงแบบปกติ

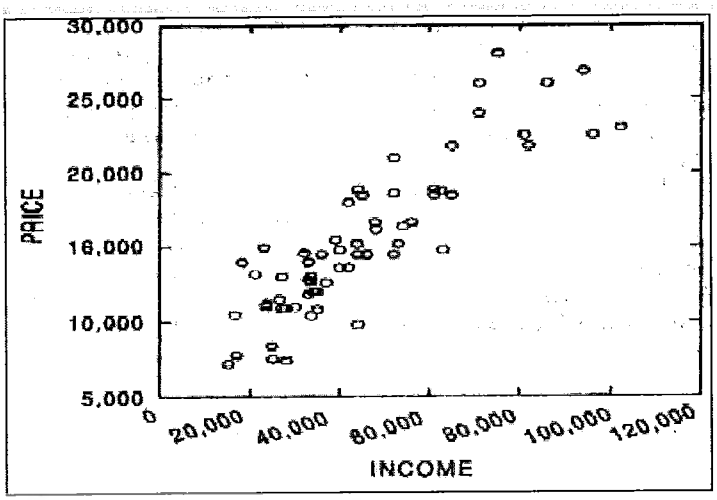
(ค่าคลาดเคลื่อน (ค่า e) มีค่าเฉลี่ยเท่ากับศูนย์) Standard deviation เท่ากับ 1

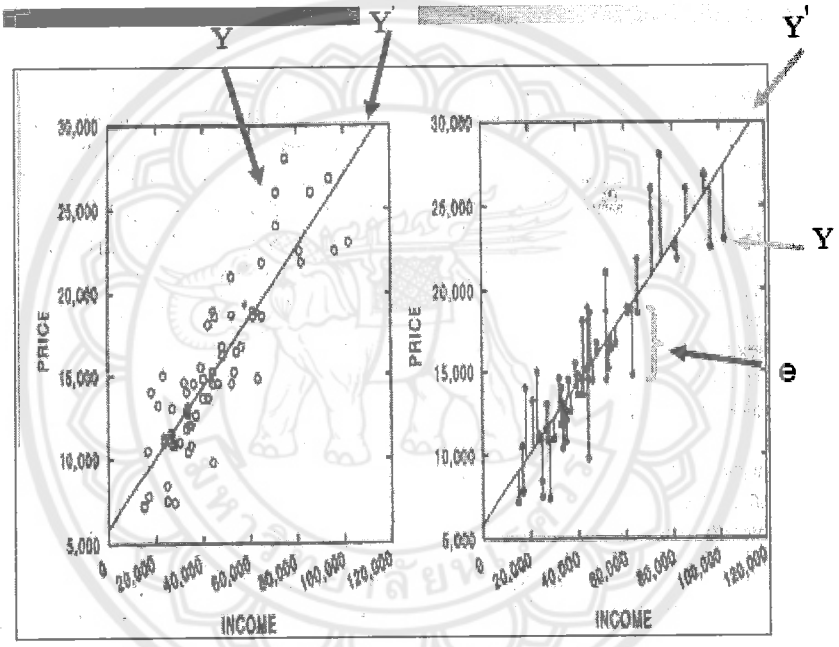
Assumption ทั้ง 4 ข้อจะต้องเป็นจริง จึงจะสามารถวิเคราะห์ simple regression ต่อได้ ทำให้สมการการถดถอยที่สร้างขึ้น สามารถนำไปใช้ในการพยากรณ์ได้อย่างมีประสิทธิภาพ

การทดสอบ Assumption

1. การตรวจสอบความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระ กับตัวแปรตาม ....จาก correlation และกราฟ
2. การตรวจสอบค่าแปรปรวนของค่าคาดเคลื่อน..ดูจากกราฟที่พล็อตระหว่าง ค่าคาดเคลื่อน (e) กับค่าพยากรณ์ (Y')
3. การตรวจสอบความเป็นอิสระกันของค่าคลาดเคลื่อน...ใช้สถิติทดสอบ Durbin-Watson
4. การตรวจสอบค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติหรือไม่....Histogram, Boxplot ของค่าคาดเคลื่อน

X และ Y มีความสัมพันธ์เชิงเส้นตรง





Simple Regression Equation

การศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัว ที่มีความสัมพันธ์อยู่ในรูปเชิงเส้น ซึ่งสามารถแสดงความสัมพันธ์ในรูปสมการเชิงเส้น ดังนี้

$$y' = b_0 + b_1 x_1 + e_i$$

เมื่อ y' = ค่าพยากรณ์ของตัวแปรตาม (y)

b0= ค่าจุดตัดแกน y (ค่าของ y เมื่อตัวแปรอิสระทุกตัว = 0)

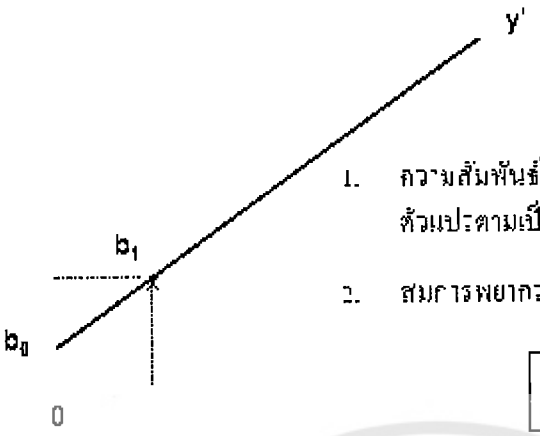
b1= ค่าสัมประสิทธิ์ถดถอย (เมื่อ x1 เปลี่ยนไป 1 หน่วย

แล้ว y' จะเปลี่ยนไปเท่าใด)

โดย ei = ค่าคลาดเคลื่อนในการพยากรณ์

กราฟพยากรณ์

X	Y
0	2
1	4



1. ความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามเป็นอย่างไร เชิงเส้นตรง ?
2. สมการพยากรณ์ ?

$$y' = b_0 + b_1 x_1$$

กฎข้อ 1 X และ Y มีความสัมพันธ์เชิงเส้นตรง

H 0 : x ไม่มีความสัมพันธ์กับ y ในรูปเชิงเส้น

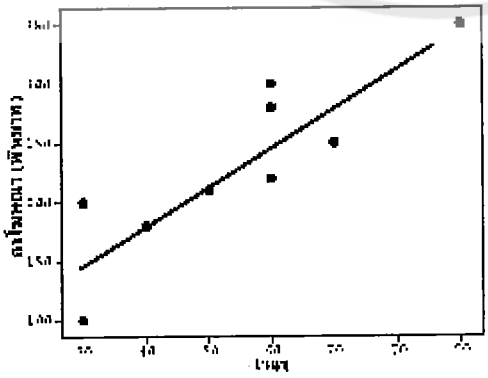
H1 : x มีความสัมพันธ์กับ y ในรูปเชิงเส้น

Correlations Matrix

Correlations

		College GPA	Highschool GPA
Pearson Correlation	College GPA	1.000	.656
	Highschool GPA	.656	1.000
Sig. (1-tailed)	College GPA	.	.000
	Highschool GPA	.000	.
N	College GPA	30	30
	Highschool GPA	30	30

การตรวจสอบ X และ Y ว่ามีความสัมพันธ์เชิงเส้นหรือไม่



Graph

ข้อมูลมีลักษณะการกระจายตัวเป็นในแนวเส้นตรง แสดงว่าเวลาที่ใช้ในการโฆษณามีความสัมพันธ์เชิงเส้นกับอัตราค่าโฆษณา

การตรวจสอบความสัมพันธ์เชิงเส้นระหว่าง x กับ y

- เป็นการทดสอบว่าสัมประสิทธิ์ถดถอยนั้นแตกต่างจากศูนย์อย่างมีนัยสำคัญหรือไม่
-ใช้การทดสอบ t-test สำหรับสัมประสิทธิ์ถดถอย
- หากปฏิเสธ Null Hypothesis แสดงว่าความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามมีอยู่จริง

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

ค่าแปรตามไม่ขึ้นกับค่าแปรอิสระ ตัวแปรอิสระไม่สามารถพยากรณ์ค่าแปรตาม

ค่าแปรตามขึ้นกับค่าแปรอิสระ ตัวแปรอิสระสามารถพยากรณ์ค่าแปรตาม

H_0: Y ไม่ขึ้นกับ X

H_1: Y ขึ้นกับ X

H_0: X ไม่พยากรณ์ Y

H_1: X พยากรณ์ Y

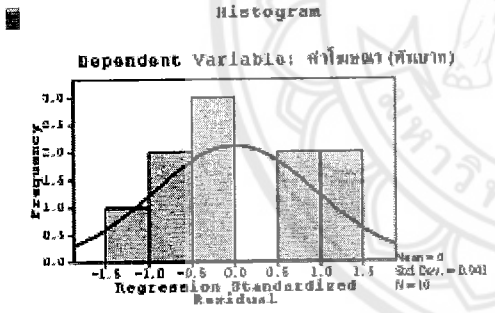
ด้วยการทดสอบนัยสำคัญของสัมประสิทธิ์ถดถอย

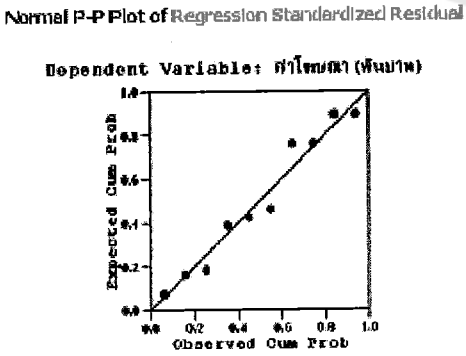
<u>กฎข้อ 2 ค่า e มีความแปรปรวนคงที่</u>

ดูจากลักษณะการกระจายตัวของข้อมูล โดยสังเกตจากกราฟ ซึ่งจะเป็นการสร้างกราฟระหว่างตัวแปร e และตัวแปร $\hat{Y}$

คำสั่ง

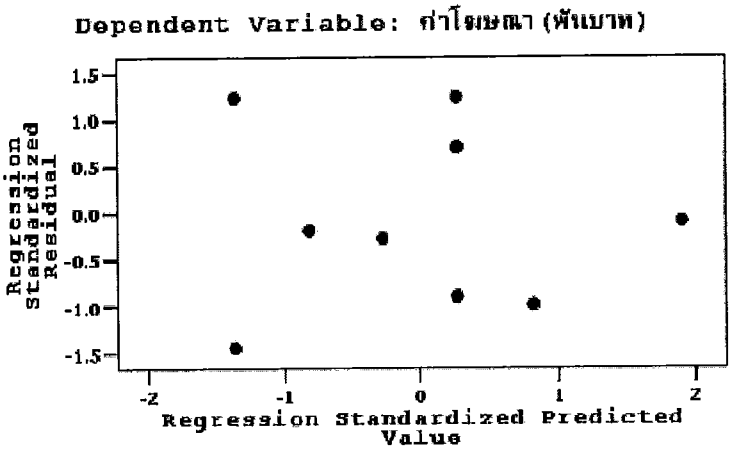
Analysis → Regression → Linear …



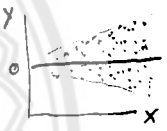


ทั้ง 2 รูป ใช้สำหรับดูการแจกแจงแบบปกติของค่า e

Scatterplot



ลักษณะของข้อมูลมีการกระจายตัวแบบไม่เป็นรูปแบบ แสดงว่าค่าความแปรปรวนของค่า e คงที่ (ถ้าลักษณะของข้อมูลมีการกระจายตัวแบบเป็นรูปแบบ คือเป็นแบบเชิงเส้น แสดงว่า ค่าความแปรปรวนของค่า e ไม่คงที่) ดูรูปหน้าถัดไป homoscedasticity ? มีการกระจายตัวรอบเส้น y=0 อย่างเป็นรูปแบบ เช่น ปากแตร



<u>กฎข้อ 3 ค่า e มีการแจกแจงแบบปกติ</u>

Residual Analysis

ใช้สมการพยากรณ์มาสร้างคะแนนพยากรณ์ (Y') และนำไปหาความคลาดเคลื่อน (ej) จากนั้นนำความคลาดเคลื่อนมาวิเคราะห์ต่อไปนี้

- Normal distribution (ดูจาก Normal plot หรือทดสอบทางสถิติ)
- outliers โดยดูจากค่า Standardized residual ซึ่งไม่ควรสูงกว่า -2.00 ถึง 2.00
- <u>Homoscedasticity</u> โดยนำ residual ไป plot ร่วมกับค่าพยากรณ์ (Y') → กฎข้อ 2

Tests of Normality

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.172	10	.200*	.932	10	.469

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Ho : ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

H1 : ค่าความคลาดไม่มีการแจกแจงแบบปกติ

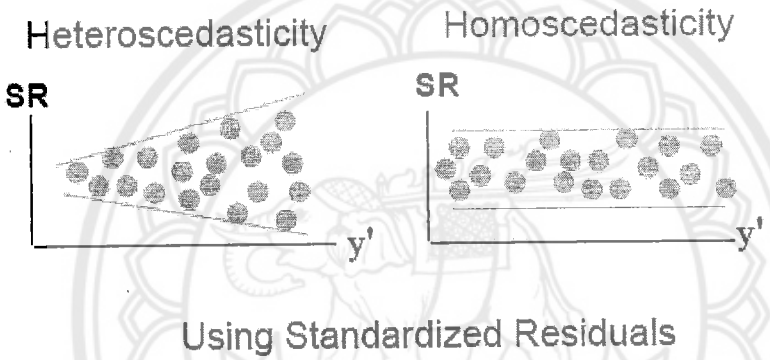
ดังนั้น จากการตรวจสอบ กฎข้อที่ 3 พบว่า ค่า e มีการแจกแจงแบบปกติ

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	153.80	353.48	237.00	61.263	10
Residual	-53.803	46.361	.000	35.327	10
Std. Predicted Value	-1.358	1.901	.000	1.000	10
Std. Residual	-1.436	1.237	.000	.943	10

a. Dependent Variable: ค่าโฆษณา(พันบาท)

Residual Analysis for Homoscedasticity



<u>กฎข้อ 4 ค่า e แต่ละตัว</u>จะต้องเป็นอิสระต่อกัน

โดยจะตรวจสอบจากค่าของ Durbin-Watson ดังนี้

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.866[a]	.750	.719	37.470	1.637

a. Predictors: (Constant), เวลา

b. Dependent Variable: ค่าโฆษณา(พันบาท)

จากผลลัพธ์พบว่าค่า Durbin-Watson = 1.637 เนื่องจากมีค่าเข้าใกล้ 2 แสดงว่าค่า e แต่ละตัวจะต้องเป็นอิสระต่อกัน ***Durbin-Watson ควรมีค่าอยู่ระหว่าง 1.5-2.5

<u>กฎ..... ค่า e มีค่าเฉลี่ยเท่ากับศูนย์</u>

ไม่ต้องทดสอบ เนื่องจากในการวิเคราะห์ค่าสัมประสิทธิ์การถดถอย จะใช้วิธีกำลังสองน้อยสุด (Least Square Method) จะทำให้ค่าผลรวมของ e มีค่าเท่ากับศูนย์ ดังนั้น ค่าเฉลี่ยของ e จะมีค่าเท่ากับศูนย์ตามไปด้วย กฎข้อที่ 1 จึงเป็นจริงเสมอ

ขั้นตอนในการวิเคราะห์ Simple Regression

- ☐ ทดสอบข้อตกลงเบื้องต้น
- ☐ ประเมินผลสมการ (Model testing)
  - -Coefficient of Determination (R2)
  - -Overall test
  - -Partial test
- ☐ Estimate ค่า b0, b1 และนำมาเขียนเป็นสมการคะแนนดิบ / สมการถดถอย
- ☐ นำเสนอและแปลความ

วิธีการสร้างสมการการถดถอย

- ☐ Enter
- ☐ Forward
- ☐ Backward
- ☐ Stepwise
- ☐ Remove

<u>ขั้นตอนการวิเคราะห์ด้วย Simple Regression</u>

คำสั่ง Analyze → Regression → Linear

Assumption

1. XและY มีความสัมพันธ์เชิงเส้นตรง
2. ค่าแปรปรวนของค่าคลาดเคลื่อนมีค่าคงที่
3. ค่าคลาดเคลื่อนต้องเป็นอิสระกัน
4. ค่าคลาดเคลื่อนต้องมีการแจกแจงแบบปกติ
   (ค่าคลาดเคลื่อน(e) มีค่าเฉลี่ย = 0)

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	เวลา[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: ค่าโฆษณา(พันบาท)

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.866[a]	.750	.719	37.470	1.637

a. Predictors: (Constant), เวลา

b. Dependent Variable: ค่าโฆษณา(พันบาท)

-สัมประสิทธิ์สหสัมพันธ์ (R)= 0.866

-สัมประสิทธิ์การตัดสินใจ (R2)= 0.750

-ค่าความคลาดเคลื่อนมาตรฐานของการประมาณ = 37.470

-หมายความว่าการประมาณค่าของอัตราค่าโฆษณามีความคลาดเคลื่อนเกิดขึ้นเท่ากับ 37.470 พันบาท(ใช้หน่วยเดียวกับหน่วยของตัวแปรตาม)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	33777.9	1	33777.9	24.058	.001[a]
	Residual	11232.1	8	1404.02		
	Total	45010.0	9			

a. Predictors: (Constant), เวลา

b. Dependent Variable: ค่าโฆษณา(พันบาท)

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

1. Ho : เวลาที่ใช้ในการโฆษณาไม่มีความสัมพันธ์กับอัตราค่าโฆษณาแบบเชิงเส้น

   H1 : เวลาที่ใช้ในการโฆษณามีความสัมพันธ์กับอัตราค่าโฆษณาแบบเชิงเส้น

2. ค่า Sig < α แสดงว่าปฏิเสธสมมติฐานหลัก นั่นคือ เวลาที่ใช้ในการโฆษณามีความสัมพันธ์กับอัตราค่าโฆษณาแบบเชิงเส้น ที่ระดับนัยสำคัญ 0.05

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	53.967	39.152		1.378	.205
	เวลา	3.328	.678	.866	4.905	.001

a. Dependent Variable: ค่าโฆษณา(พันบาท)

เขียนสมการพยากรณ์

$$\hat{Y} = b_0 + b_1 X$$

อัตราค่าโฆษณา = 53.967 + 3.328(เวลาในการโฆษณา)

วิธีการพยากรณ์ค่า

ในการพยากรณ์ค่า หรือการประมาณค่า จะต้องทราบค่า X เพื่อจะประมาณค่า Y ได้

เช่น เมื่อ เวลาในการโฆษณา = 100 นาที/สัปดาห์ อัตราค่าโฆษณาจะเป็นเท่าไร

$\hat{}$

อัตราค่าโฆษณา = 53.967 + 3.328 (100)

= 386.767 พันบาท/นาที

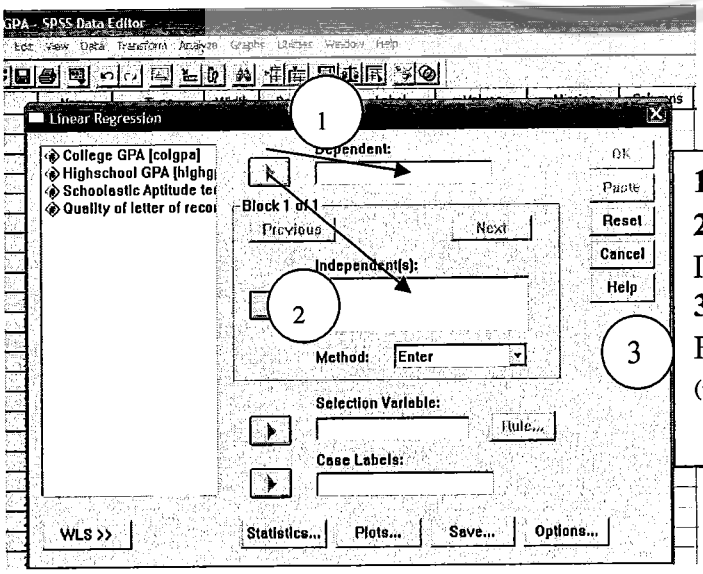
= 386,767 บาท/นาที

ตัวอย่าง การวิเคราะห์Simple Regression

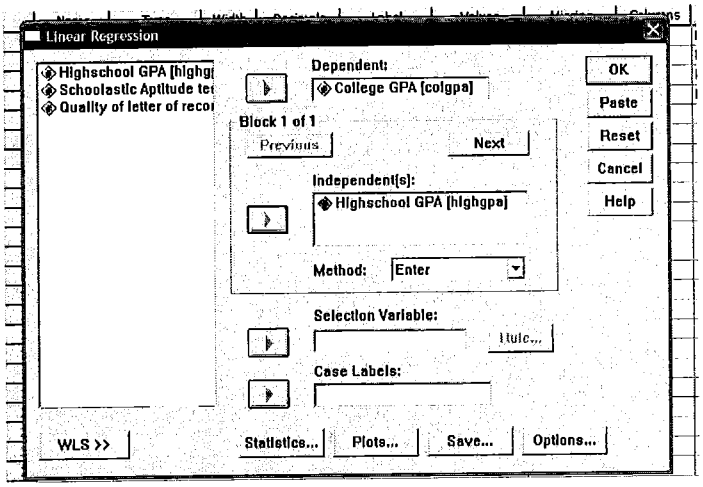
การศึกษาเพื่อพยากรณ์ว่าคะแนน Highschool GPAจะเป็นตัวพยากรณ์ที่ดีของคะแนน College GPA หรือไม่อย่างไร

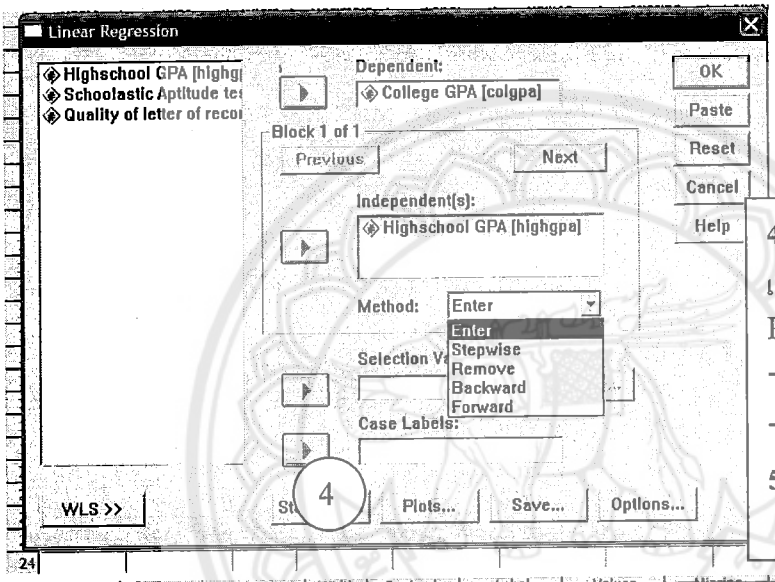
ขั้นตอนการวิเคราะห์ข้อมูล

1. Analysis → Regression → Linear …



1.เลือกตัวแปรตาม ไปที่ช่อง dependent
2. ตัวแปรต้น หรือตัวแปรอิสระ ย้ายไปที่ Independent
3. ในช่อง method เลือกวิธีการวิเคราะห์ Enter , Stepwise, Backward (กรณี ตัวแปรเดียว เลือก enter)

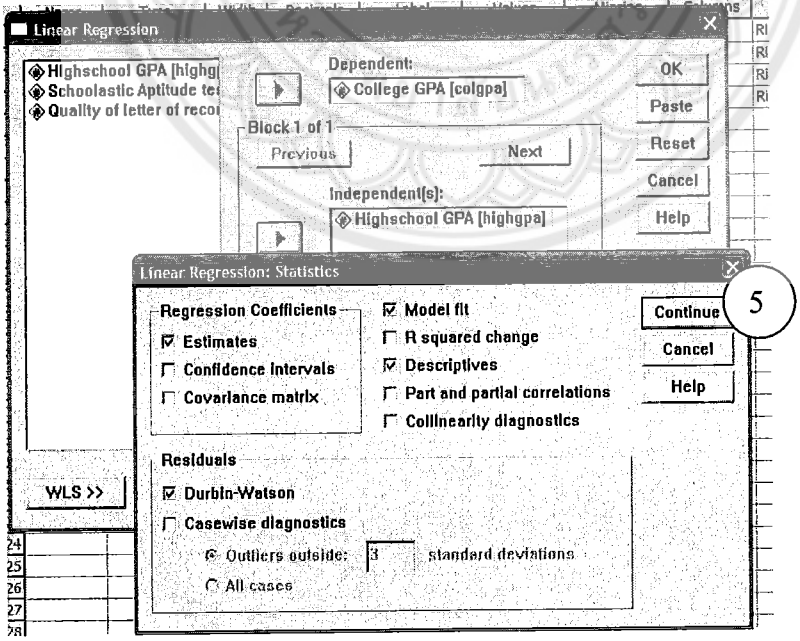


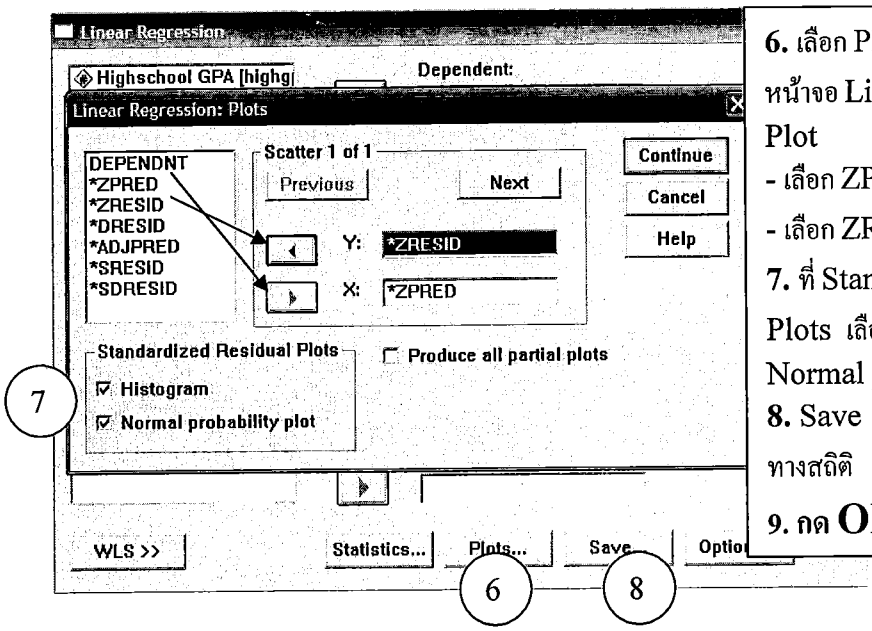


4. เลือก statistics โปรแกรมจะแสดงหน้าจอ Linear Regression: statistics

- เลือก Durbin-Watson
- เลือก Discriptive

5. กด Continue





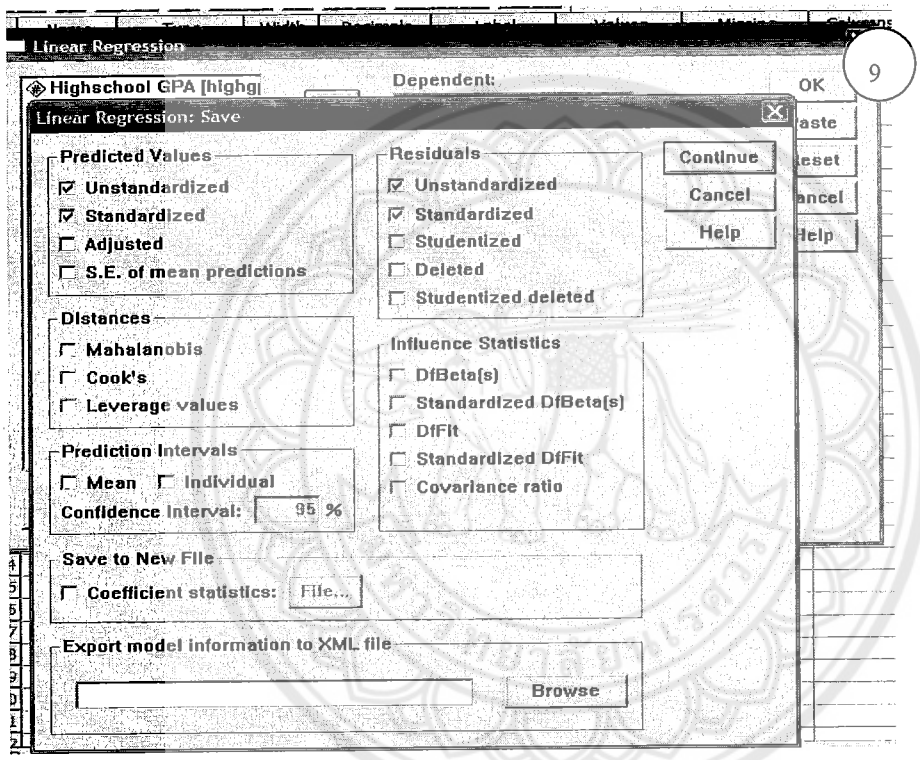
6. เลือก Plot โปรแกรมจะแสดงหน้าจอ Linear Regression: Plot

- เลือก ZPRED ไปที่แกน x
- เลือก ZRESID ไปที่แกน y

7. ที่ Standardized Residual Plots เลือก Histogram และ Normal probability plot

8. Save กำหนดระดับนัยสำคัญทางสถิติ

9. กด **OK**



การแปรผลจากหน้าจอ output

1.การตรวจสอบความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระกับตัวแปรตาม

X และ Y <u>ต้อง</u>มีความสัมพันธ์เชิงเส้นตรง (ตรวจสอบจาก **correlation** และ **กราฟ**)

Ho: Highschool GPA กับ College GPA **ไม่มีความสัมพันธ์เชิงเส้น**

H1: Highschool GPA กับ College GPA **มีความสัมพันธ์เชิงเส้น**

ทดสอบ Assumption นี้โดยดูจากตาราง

Correlations

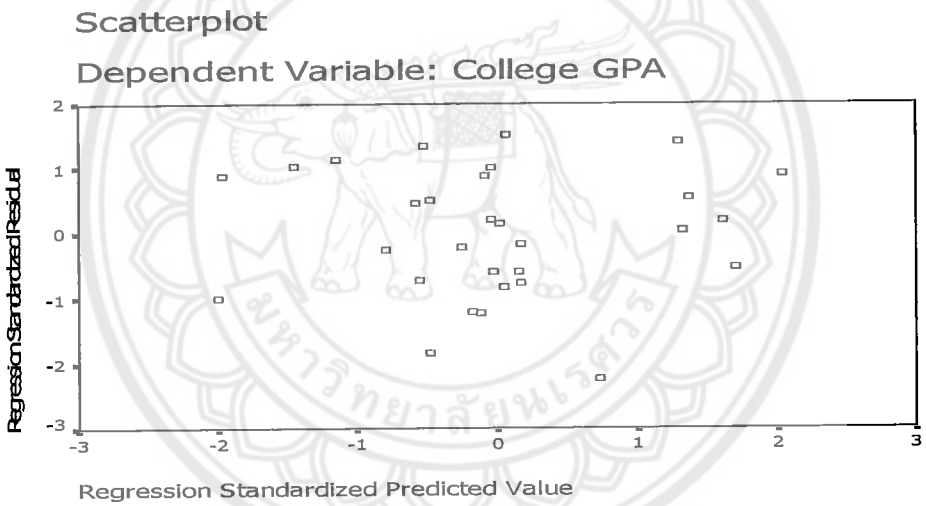
		College GPA	Highschool GPA
Pearson Correlation	College GPA	1.000	.656
	Highschool GPA	.656	1.000
Sig. (1-tailed)	College GPA	.	.000
	Highschool GPA	.000	.
N	College GPA	30	
	Highschool GPA		

ถ้า r เป็น + มีค่าเข้าใกล้ 1 แสดงว่ามีความสัมพันธ์กันมากในรูปเชิงเส้น

จากตารางจะพบว่า

ค่า Sig. = .000 < .05 ยอมรับ H1 แสดงว่า X มีความสัมพันธ์กับ Y ในรูปเชิงเส้น

2. ค่าแปรปรวนของค่าคาดเคลื่อน**ต้อง**มี**ค่าคงที่** (ตรวจสอบ ....ดูจากกราฟ ที่พล็อตระหว่าง ค่า**คาดเคลื่อน** (e) กับค่าพยากรณ์ (Y') โดยแปลผลจากกราฟ แสดงว่าค่าแปรปรวนของค่าคลาดเคลื่อน (e) มีค่าคงที่ (การกระจายตัวของ plot ไม่เป็นรูปแบบ)



3. ค่าคลาดเคลื่อน**ต้อง**เป็นอิสระกัน (ตรวจสอบ ... ใช้สถิติทดสอบ Durbin-Watson) โดยแปลผลจากตาราง

Model Summary [b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.656[a]	.430	.410	.69077	2.289

a. Predictors: (Constant), Highschool GPA

b. Dependent Variable: College GPA

จากตารางจะพบว่าค่า Durbin-Watson = 2.289 (อยู่ในช่วง 1.5-2.5) แสดงว่าค่าความคลาดเคลื่อนเป็นอิสระกัน

4. ค่าคลาดเคลื่อน**ต้อง**มีการแจกแจงแบบปกติ (ตรวจสอบ ...Histogram, Boxplot ของค่าคาดเคลื่อน)

โดยแปลผลจากตาราง

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	.7430	3.1291	1.9263	.59008	30
Residual	-1.5499	1.0419	.0000	.67876	30
Std. Predicted Value	-2.005	2.038	.000	1.000	30
Std. Residual	-2.244	1.508	.000	.983	30

a. Dependent Variable: College GPA

จากตารางจะพบว่าค่า Mean และ Std.Deviation ของ Std.Residual = .00 และ .983 ซึ่งมีค่าเข้าใกล้ 1 แสดงว่าค่าคลาดเคลื่อนมีการแจกแจงแบบปกติ

การทดสอบ Assumption ข้อตกลงทุกข้อต้องผ่าน ถ้าข้อใดข้อหนึ่งไม่ผ่าน จะนำมาวิเคราะห์ Regression ไม่ได้

1. **ทดสอบสมมติฐาน (มี 3 แบบ)**

   H_0: $\beta 1 = 0$ (ค่าสัมประสิทธิ์ถดถอยของตัวแปรมีค่าเป็น 0)

   H_1: $\beta 1 \neq 0$ (ค่าสัมประสิทธิ์ถดถอยของตัวแปรไม่มีค่าเป็น 0)

   H_0: Y ไม่ขึ้นกับ X

   H_1: Y ขึ้นกับ X

   H0: X ไม่พยากรณ์ Y

   H1: X พยากรณ์ Y

โดยแปลผลจากตาราง (ดูค่า F ของ Overall test จะบอกว่า X จะพยากรณ์ y ได้หรือไม่)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10.098	1	10.098	21.162	.000[a]
	Residual	13.361	28	.477		
	Total	23.458	29			

a. Predictors: (Constant), Highschool GPA

b. Dependent Variable: College GPA

จากตารางจะพบว่าค่า Sig. = .000 < .05 ยอมรับ H1 แสดงว่า X พยากรณ์ Y ได้ แต่ต้องการทราบต่อไปว่าพยากรณ์ได้กี่ %

โ ดยแปลผลจากตาราง (ดู Partial test จะบอกว่า X พยากรณ์ y ได้กี่ %)

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.656[a]	.430	.410	.69077	2.289

a. Predictors: (Constant), Highschool GPA

b. Dependent Variable: College GPA

จากตารางให้ดูที่ค่า Adjust R Square = .410 แสดงว่า X พยากรณ์ Y ได้ 41%

ค่า Adjust R^2 คือค่า R^2 ที่เอาไปถ่วงน้ำหนัก ทำให้มีความน่าเชื่อถือ ถ้ากลุ่ม ตัวอย่างน้อยอ่านค่านี้ แต่ถ้ากลุ่ม ตัวอย่างใหญ่ ค่า R^2 จะใกล้เคียงค่า Adjust นอกจากนี้ยังต้องสามารถเขียนสมการพยากรณ์ให้ได้อีก

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.508	.333		1.525	.139
	Highschool GPA	.712	.155	.656	4.600	.000

a. Dependent Variable: College GPA

จากตารางจะสามารถเขียนสมการได้ดับนี้ $Y' = b0 + b1x1$

$$Y' = .508 + .712 \text{ (Highschool GPA)}$$

<u>สรุป</u> คะแนน High school GPA เป็นตัวพยากรณ์ของคะแนน College GDP โดยมีอำนาจพยากรณ์ 41 % อย่างมีนัยสำคัญทางสถิติที่ 0.05

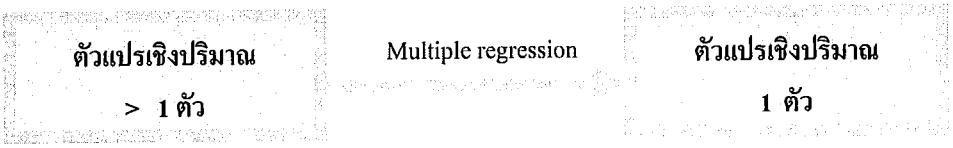
Multiple Regression Analysis (MRA)

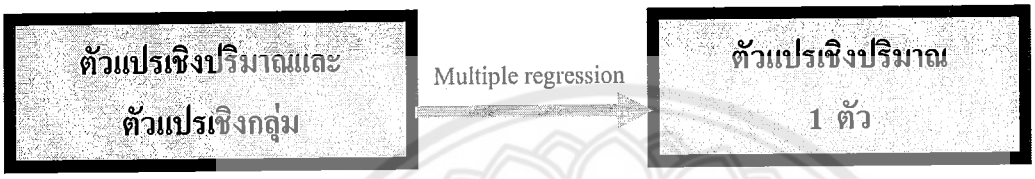
วัตถุประสงค์

- เพื่อศึกษาว่าตัวแปรอิสระ / ปัจจัยตัวใดบ้าง ที่ส่งผล (มีอิทธิพลต่อตัวแปรตาม)

- เพื่อพยากรณ์ตัวแปรตาม เมื่อทราบตัวแปรอิสระ

ใช้วิเคราะห์ ตัวแปรต้นเชิงปริมาณ <u>มากกว่า 1 ตัว</u> กับตัวแปรตามเชิงปริมาณ <u>1 ตัว</u>

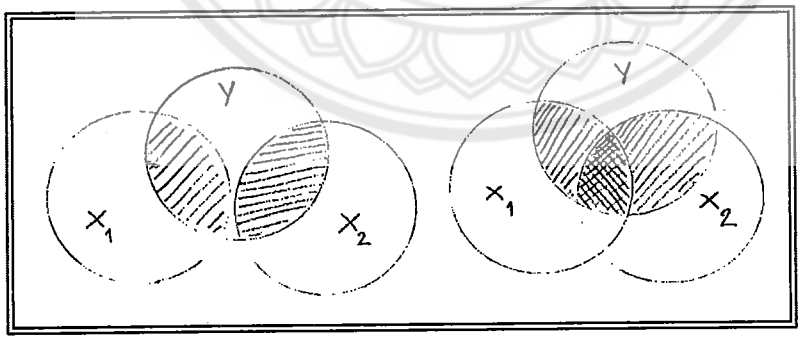




Assumptions

1. X และ Y มีความสัมพันธ์เชิงเส้นตรง (Linear Relationship)
2. ค่าคลาดเคลื่อน (Error /residual: e) มีการแจกแจงปกติ (Boxplot, Histogram.....)
3. ค่าแปรปรวนของค่าคลาดเคลื่อนมีค่าคงที่ (Homoscedasticity)
4. ค่าคลาดเคลื่อนที่ I และ j ต้องเป็นอิสระกัน (Durbin)
5. ตัวแปรอิสระต้องไม่มีความสัมพันธ์กันเอง (Multicollinearity) (Correlation) (เพิ่มจาก simple regression)

ผลของการเกิด Multicollinearity



- ทำให้การตีความอิทธิพลของตัวแปรอิสระตัวหนึ่ง ที่มีต่อตัวแปรตามเป็นไปได้ยาก เพราะเกิด การปนกัน (Confounding effect)
- ทำให้ค่าสัมประสิทธิ์ถดถอย (bj) แกว่ง (unstable) ซึ่งจะส่งผลให้การพยากรณ์มีความคลาดเคลื่อนด้วย

การทดสอบ Multicollinearity

- ผลการวิเคราะห์ Regression พบว่ามีค่า R2 ปานกลางถึงสูง แต่ไม่มีตัวแปรใด Significant เป็นข้อสังเกตประการแรก
- ดูจากค่า correlation ระหว่างตัวแปรอิสระกันเองใน correlation matrix
- ดูจากค่า VIF (variance inflation factors)
  - ค่า VIF ไม่ควรเกินกว่า 10 (Myers, Stevens, Hair และคณะ)
- ดูจากค่า Tolerance
  - ค่า Tolerance มีค่าสูงสุดไม่เกิน 1 ยิ่งมีค่ามากจะแสดงถึงความเป็นอิสระของตัวแปรนั้น ๆ ยิ่งมีค่าใกล้ 0 จะแสดงถึงการมีความสัมพันธ์ระหว่างตัวแปรอิสระกันเอง

การจัดการเมื่อเกิด Multicollinearity

- ตัดตัวแปรที่มี ความสัมพันธ์ระหว่างตัวแปรอิสระกันเองออก
- รวมตัวแปรที่มีความสัมพันธ์กันสูง (มากกว่า .8) เข้าด้วยกันด้วยเทคนิค Factor Analysis

Multiple Regression Equation

- **โครงสร้างเหมือน Simple Regression หากแต่เพิ่มตัวแปรอิสระเข้าไปมากกว่า 1 ตัว**

$$y' = b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n + e$$

เมื่อ y' = ค่าพยากรณ์ของตัวแปรตาม (y)

b0 = ค่าจุดตัดแกน y (ค่าของ y เมื่อตัวแปรอิสระทุกตัว = 0)

bj = ค่าสัมประสิทธิ์ถดถอย (เมื่อ xj เปลี่ยนไป 1 หน่วย แล้ว y' จะเปลี่ยนไปเท่าใด โดยที่ x ตัวอื่น ๆ มีค่าคงที่) โดยที่ eij = ค่าความคลาดเคลื่อนจากการพยากรณ์

* คำว่า "ค่าคงที่" หมายถึงค่าของตัวแปรอิสระตัวอื่น จะเป็นอะไรก็จะได้ผลเหมือนกัน (ผลของตัวแปรอื่นไม่มีต่ออิทธิพลของ xj กับ y)

* ถ้าผลของตัวแปรอื่นมีต่ออิทธิพลของ xj กับ y แล้ว จะเกิดปัญหากับ Regression อย่างมาก และไม่สามารถใช้โมเดลนี้วิเคราะห์ได้

ขั้นตอนในการวิเคราะห์ (enter method)

- ทดสอบข้อตกลงเบื้องต้นเกี่ยวกับ Normality, Linearity, Multicollinearity, Residual Analysis (nomality, homoscedasticity)
- ประเมินผลสมการ (Model testing)
  - Standard Error of Estimation

- Coefficient of Determination (R2)
- Overall test
- Partial test

- Estimate ค่า b0, bj และนำมาเขียนเป็นสมการคะแนนดิบ
- นำเสนอและแปลความ

Standard error of the estimate

- ค่าความแตกต่างโดยเฉลี่ย ระหว่าง y และ y‘ คำนวณโดย

$$s_\varepsilon = \sqrt{\frac{SSE}{n-k-1}}$$

- ถ้ามีค่ามาก บ่งบอกว่ามีความคลาดเคลื่อนของการพยากรณ์มาก (ตัวแปรอิสระอาจไม่มีอิทธิพลต่อตัวแปรตาม หรือ linear regression อาจไม่ใช่ วิธีที่เหมาะสม)
- ถ้ามีค่าน้อย บ่งบอกว่ามีความคลาดเคลื่อนของการพยากรณ์ต่ำ (ตัวแปรอิสระมีอิทธิพลสูงต่อตัวแปรตาม และ linear regression เป็นวิธีที่เหมาะสมในการศึกษาความสัมพันธ์)
- ไม่มี range ที่แน่นอกว่าค่าที่เหมาะสมควรอยู่ระหว่างเท่าใด

การทดสอบนัยสำคัญของสัมประสิทธิ์ถดถอย

- เป็นการทดสอบว่าสัมประสิทธิ์ถดถอยนั้นแตกต่างจาก 0 (ศูนย์)อย่างมีนัยสำคัญหรือไม่

$$H_o : \beta_i = 0$$
$$H_1 : \beta_i \neq 0$$

- ใช้การทดสอบ t-test สำหรับสัมประสิทธิ์ถดถอย
  - H_0: Y ไม่ขึ้นกับ X ใดๆ
  - H_1: Y ขึ้นกับ X อย่างน้อย 1 ตัว
- หากปฏิเสธ Null Hypothesis แสดงว่าความสัมพันธ์ระหว่างตัวแปรอิสระและ ตัวแปรตาม มีอยู่จริง
  - H_0: X ไม่มีส่วนร่วมในการพยากรณ์ Y
  - H_1: X มีส่วนร่วมในการพยากรณ์ Y อย่างน้อย 1 ตัว

วิธีการคัดเลือกสมการที่ดีที่สุด

Backward Elimination

- ทดสอบสมการด้วยตัวแปรอิสระทุกตัว
- ตัดตัวแปรที่มีความสัมพันธ์กับ y ต่ำสุดออก
- ทดสอบสมการด้วยตัวแปรที่เหลืออยู่
- ตัดตัวแปรที่มีความสัมพันธ์กับ y ต่ำ รองลงมาออก (partial out)
- ทำซ้ำจนไม่สามารถนำตัวแปรใดเข้าสมการได้
- ทดสอบสมการด้วยตัวแปรที่เหลืออยู่

Stepwise Procedure

- นำตัวแปรที่มีความสัมพันธ์กับ y สูงสุดเข้าสู่สมการ
- ทดสอบสมการด้วยตัวแปรที่มีอยู่
- นำตัวแปรที่มี partial correlation กับ y สูง รองลงมา เข้าสู่สมการ
- ทดสอบสมการด้วยตัวแปรที่มีอยู่
- ทำซ้ำจนตัวแปรที่เหลืออยู่ ไม่มีตัวใด มีความสัมพันธ์กับ y อีก จึงหยุด

ตัวอย่าง Multiple regression

ต้องการศึกษาว่าปัจจัยใดที่มีอิทธิพลต่อผลสัมฤทธิ์ทางการเรียนของนักเรียนระดับ College โดยมีตัวแปรอิสระ คือ ผลสัมฤทธิ์ทางการเรียนในระดับมัธยม (Highgpa) คะแนนความถนัดทางการเรียน (SAT) และคะแนนจดหมายแนะนำ (letter) ตัวแปรตาม คือ คะแนนผลสัมฤทธิ์ทางการเรียน (Colgpa)ของนักเรียนระดับวิทยาลัยจำนวน 30 คน

ขั้นตอนการวิเคราะห์ด้วย Multiple Regression

1. ทดสอบ Assumption

1.1 X และ Y มีความสัมพันธ์เชิงเส้นตรง (Linear Relationship)

Ho: ตัวแปรตามกับตัวแปรอิสระไม่มีความสัมพันธ์เชิงเส้นตรง

H1: ตัวแปรตามกับตัวแปรอิสระมีความสัมพันธ์เชิงเส้นตรง

โดยแปลผลจากตาราง

Correlations

		College GPA	Highschool GPA	Schoolastic Aptitude test	Quality of letter of recommendation
Pearson Correlation	College GPA	1.000	.656	.747	.443
	Highschool GPA	.656	1.000	.447	.570
	Schoolastic Aptitude test	.747	.447	1.000	.208
	Quality of letter of recommendation	.443	.570	.208	1.000
Sig. (1-tailed)	College GPA		.000	.000	.007
	Highschool GPA	.000	.	.007	.001
	Schoolastic Aptitude test	.000	.007	.	.135
	Quality of letter of recommendation	.007	.001	.135	.
N	College GPA	30	30	30	30
	Highschool GPA	30	30	30	30
	Schoolastic Aptitude test	30	30	30	30
	Quality of letter of recommendation	30	30	30	30

จากตารางจะพบว่า

- ค่า Pearson ของ Highschool GPA = .656

  Schoolastic Aptitude GPA = .747

  Quality of letter of recommendation = .443

- ค่า Sig. ของ Highschool GPA = .000

  Schoolastic Aptitude GPA = .000

  Quality of letter of recommendation = .007

ค่า Pearson > .4 มีความสัมพันธ์เชิงเส้นตรง

ค่า Sig. < .05 ยอมรับ H1

สรุปได้ว่า X1, X2, X3 มีความสัมพันธ์เชิงเส้นตรงกับ Y

1.2 ค่าคลาดเคลื่อนมีการแจกแจงปกติ

โดยแปลผลจากตาราง

Residuals Statistics[a]

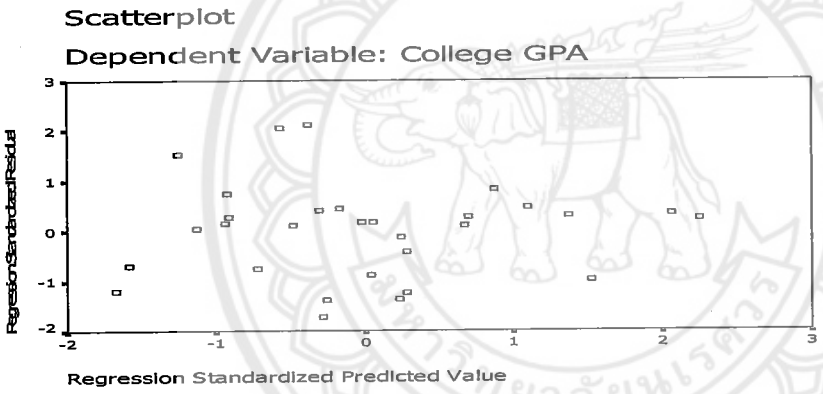
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	.6683	3.6250	1.9263	.7530	30
Residual	-.8989	1.1014	.0000	.4919	30
Std. Predicted Value	-1.671	2.256	.000	1.000	30
Std. Residual	-1.730	2.120	.000	.947	30

a. Dependent Variable: College GPA

จากตารางจะพบว่าค่า Mean = 0 และ Std. Deviation ของ Std.Residual = .947 ซึ่งมีค่าเข้าใกล้ 1 แสดงว่าค่าคลาดเคลื่อนมีการแจกแจงปกติ

1.3 ค่าแปรปรวนของค่าคลาดเคลื่อนมีค่าคงที่

โดยแปลผลจากกราฟ



จากกราฟจะพบว่าค่าแปรปรวนของค่าคลาดเคลื่อนมีค่าคงที่ (plot กระจายตัวแบบไม่เป็นรูปแบบ)

1.4 ค่าคลาดเคลื่อนต้องเป็นอิสระกัน

โดยแปรผลจากตาราง

Model Summary [b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.837[a]	.701	.666	.51948	1.661

a. Predictors: (Constant), Quality of letter of recommendation, Schoolastic Aptitude test, Highschool GPA

b. Dependent Variable: College GPA

จากตารางจะพบว่าค่า Durbin-Watson = 1.661 ซึ่งอยู่ในช่วง 1.5 – 2.5 แสดงว่าค่าคลาดเคลื่อนเป็นอิสระกัน

1.5 ตัวแปรอิสระต้องไม่มีความสัมพันธ์กันเอง (Multicollinearity correlation)

Ho: ตัวแปรอิสระไม่มีความสัมพันธ์กันเอง

H1: ตัวแปรอิสระมีความสัมพันธ์กันเอง

ซึ่งมีวิธีการอ่านค่าได้ 2 วิธี คือ

<u>วิธีที่ 1 แปลผลจากตาราง</u>

Correlations

		College GPA	Highschool GPA	Schoolastic Aptitude test	Quality of letter of recommendation
Pearson Correlation	College GPA	1.000	.656	.747	.443
	Highschool GPA	.656	1.000	.447	.570
	Schoolastic Aptitude test	.747	.447	1.000	.208
	Quality of letter of recommendation	.443	.570	.208	1.000
Sig. (1-tailed)	College GPA	.	.000	.000	.007
	Highschool GPA	.000	.	.007	.001
	Schoolastic Aptitude test	.000	.007	.	.135
	Quality of letter of recommendation	.007	.001	.135	.
N	College GPA	30	30	30	30
	Highschool GPA	30	30	30	30
	Schoolastic Aptitude test	30	30	30	30
	Quality of letter of recommendation	30	30	30	30

จากตารางจะพบว่า

- ค่า Pearson ระหว่าง Highschool GPA + Schoolastic Aptitude GPA = .447
- ค่า Pearson ระหว่าง Highschool GPA + Quality of letter of recommendation = .570
- ค่า Pearson ระหว่าง Schoolastic Aptitude GPA + Quality of letter of recommendation = .208

<u>สรุป</u> ค่า Pearson Correlation ไม่เกิน 0.8 แสดงว่าตัวแปรอิสระ (X)แต่ละตัวไม่มีความสัมพันธ์กันเอง

วิธีที่ 2 อ่านจากตาราง Coefficient

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-1.341	.495		-2.709	.012		
	Highschool GPA	.347	.155	.319	2.232	.034	.563	1.777
	Schoolastic Aptitude test	.002	.000	.575	4.786	.000	.797	1.254
	Quality of letter of recommendation	.091	.084	.141	1.076	.292	.672	1.488

a Dependent Variable: College GPA

ค่า Tolerance สูงสุดมีค่าไม่เกิน 1 แสดงว่าตัวแปรอิสระไม่มีความสัมพันธ์กันเอง

หรือค่า VIF ไม่เกิน 10 แสดงว่าตัวแปรอิสระไม่มีความสัมพันธ์กันเอง

สำหรับตาราง

Variables Entered/Removed b

Model	Variables Entered	Variables Removed	Method
1	Quality of letter of recommendation, Schoolastic Aptitude test, Highschool GPA	a	Enter

a. All requested variables entered.

b. Dependent Variable: College GPA

ใช้สำหรับดูว่า ตัวแปรต้นคืออะไร , ตัวแปรตามซึ่งในที่นี้ คือ College GPA และ ใช้วิธี ENTER ในการวิเคราะห์ผล

2 **ทดสอบสมมติฐาน** (เขียนสมมติฐานได้ 3 แบบ)

Ho: $\beta_1 = \beta_2 = \beta_3 = 0$ (ค่าสัมประสิทธิ์ถดถอยของตัวแปรมีค่าเป็น 0)

H1: $\beta_1 = \beta_2 = \beta_3 \neq 0$ (ค่าสัมประสิทธิ์ถดถอยของตัวแปรไม่มีค่าเป็น 0) $\beta_i \neq 0$ อย่างน้อย 1 ตัว

หรือ

H0: Y ไม่ขึ้นกับ X ใดๆ

H1: Y ขึ้นกับ X อย่างน้อย 1 ตัว

หรือ

H0: X ไม่มีส่วนร่วมในการพยากรณ์ Y

H1: X มีส่วนร่วมในการพยากรณ์ Y อย่างน้อย 1 ตัว

โดยแปลผลจากตาราง

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	16.442	3	5.481	20.310	.000[a]
	Residual	7.016	26	.270		
	Total	23.458	29			

a. Predictors: (Constant), Quality of letter of recommendation, Schoolastic Aptitude test, Highschool GPA

b. Dependent Variable: College GPA

จากตารางจะพบว่าค่า Sig. = .000 < .05ยอมรับ H1แสดงว่ามี X (อย่างน้อย 1 ตัว) มีส่วนร่วมในการพยากรณ์ Y

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.837[a]	.701	.666	.51948	1.661

a. Predictors: (Constant), Quality of letter of recommendation, Schoolastic Aptitude test, Highschool GPA

b. Dependent Variable: College GPA

และอ่านค่า R Square X สามารถพยากรณ์ Y ได้ 70.1%

ในที่นี้ให้อ่านที่ค่า R Square เนื่องจากมีเปอร์เซ็นต์การพยากรณ์มากกว่าค่า Adjust R Square และกำหนดให้เขียนเปอร์เซ็นต์ โดยมีทศนิยมเพียง 1 ตำแหน่งเท่านั้น)

และแปลผลจากตาราง

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-1.341	.495		-2.709	.012		
	Highschool GPA	.347	.155	.319	2.232	.034	.563	1.777
	Schoolastic Aptitude tes	.002	.000	.575	4.786	.000	.797	1.254
	Quality of letter of recommendation	.091	.084	.141	1.076	.292	.672	1.488

a. Dependent Variable: College GPA

อ่านค่า Beta จากตาราง จะพบว่าตัวแปรที่มีค่า Beta มาก จะมีค่าการพยากรณ์มากกว่าตัวอื่น ซึ่งค่า Beta เมื่อรวมกันแล้วจะเท่ากับ 1 ในที่นี้ จะพบว่าค่า Beta ของ SAT = .575 High school GPA = .319 และ Quality of Letter = .141

วิธี Enter ใช้เมื่อต้องการทราบความสัมพันธ์ระหว่าง y กับ x แต่ละตัว (โดย adjust x; อื่นๆ) ไม่ใช้สำหรับการพยากรณ์

เขียนสมการพยากรณ์ได้ดังนี้

$$y' = b0+b1x1+b2x2+b3x3$$

จากตารางจะพบว่าค่า Sig. ของ Highschool GPA = 0.034 < .05 ยอมรับ H1

Schoolastic Aptitude GPA = .000 < .05 ยอมรับ H1

Quality of letter of recommendation = .295 > .05 ยอมรับ H0

แสดงว่า Highschool GPA และ Schoolastic Aptitude GPA สามารถพยากรณ์ค่า College GPA

(ในการใช้วิธี Enter Method การเขียนสมการพยากรณ์ต้องนำค่า B ทุกตัวมาใส่ในสมการด้วยถึงแม้จะมีตัวที่ไม่ Sig. แต่ไม่ต้องนำค่า X ตัวที่ไม่ Sig. มาแทนค่าในสมการ)

เขียนสมการได้ดังนี้

College GPA = -1.341 + .347 (High school GPA) + .002 (SAT) + .091(QOL)

โดยในการแทนค่าตัวแปรจะไม่แทนค่า QOLเนื่องจากไม่ Sig.

การวิเคราะห์ MRA โดยวิธี Stepwise

ขั้นตอนการวิเคราะห์ แบบ stepwise

- **ต้องหาค่า r ว่าตัวแปรอิสระใดมีค่าใกล้ 1 มากที่สุด**
- **ให้นำตัวแปรนั้นเข้าขั้นตอนการวิเคราะห์**

Correlations

		College GPA	Highschool GPA	Schoolastic Aptitude test	Quality of letter of recommendation
College GPA	Pearson Correlation	1	.656**	.747**	.443*
	Sig. (2-tailed)	.	.000	.000	.014
	N	30	30	30	30
Highschool GPA	Pearson Correlation	.656**	1	.447*	.570**
	Sig. (2-tailed)	.000	.	.013	.001
	N	30	30	30	30
Schoolastic Aptitude test	Pearson Correlation	.747**	.447*	1	.208
	Sig. (2-tailed)	.000	.013	.	.269
	N	30	30	30	30
Quality of letter of recommendation	Pearson Correlation	.443*	.570**	.208	1
	Sig. (2-tailed)	.014	.001	.269	.
	N	30	30	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

จากตารางพบว่า School มีค่า r สูงที่สุด รองลงมาได้แก่ Highschool และ Quality of Letter

Variables Entered/Removed [a]

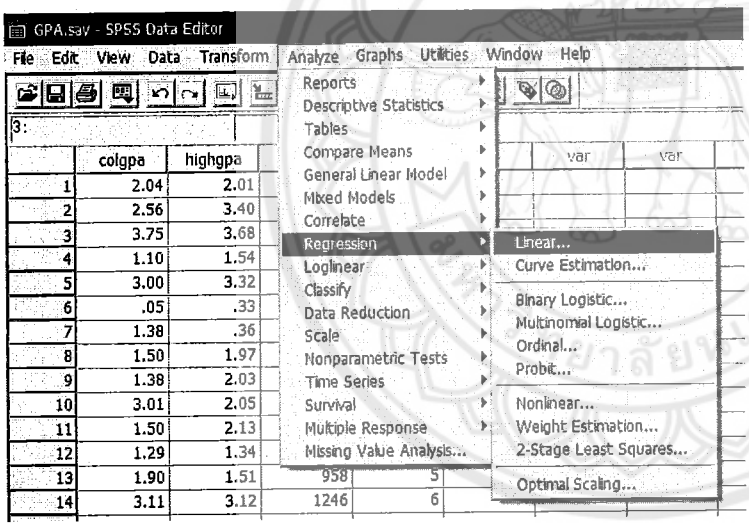
Model	Variables Entered	Variables Removed	Method
1	Schoolastic Aptitude test	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	Highschool GPA	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

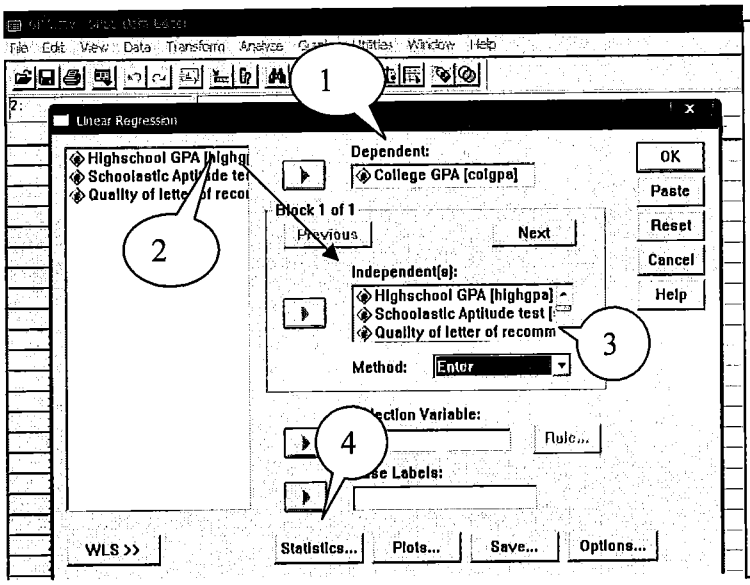
a. Dependent Variable: College GPA

โปรแกรมจะดึงข้อมูลของ X ตัวที่มีค่าความสัมพันธ์กับ Y (r) สูงสุดเข้ามาก่อนเป็นอันดับแรก และตัวที่มีค่าความสัมพันธ์รองลงมา จะถูกดึงเข้าเป็นอันดับต่อมา และตัดตัวที่ต่ำสุดออกไป

ขั้นตอนการวิเคราะห์ข้อมูล

1. Analyze → Regression → Linear





1. เลือกตัวแปรตาม ไปที่ช่อง dependent

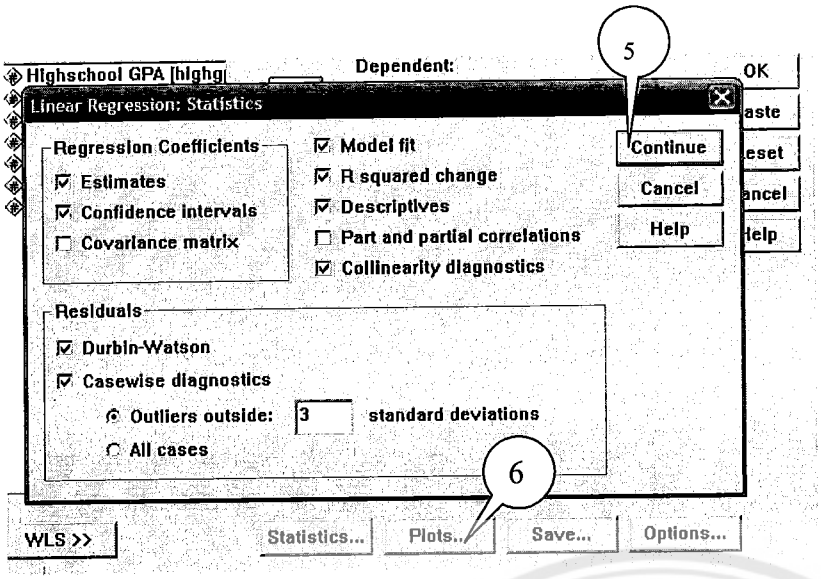
2. ตัวแปรต้นทุกตัว หรือตัวแปรอิสระ ย้ายไปที่ Independent

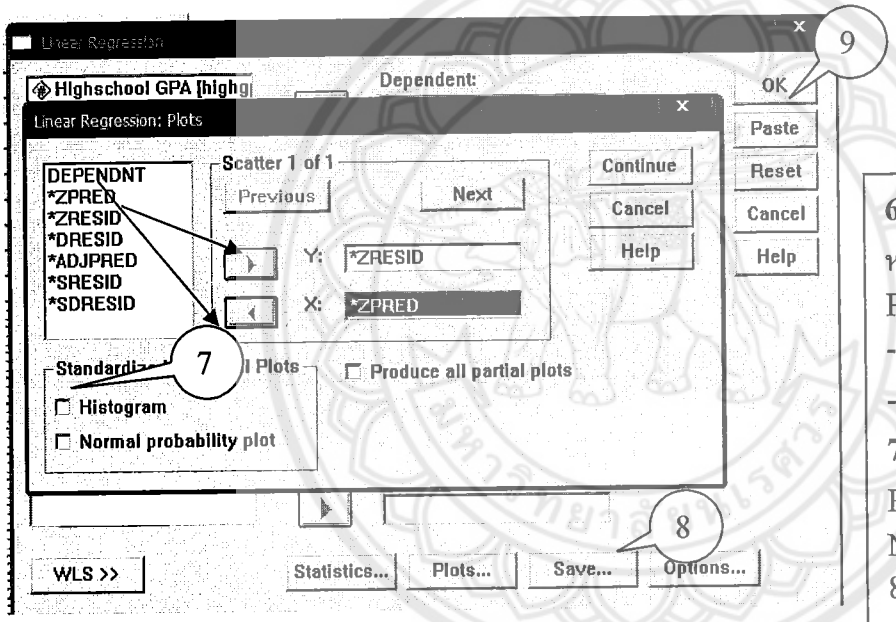
3. ในช่อง method เลือกวิธีการวิเคราะห์ **Stepwise**

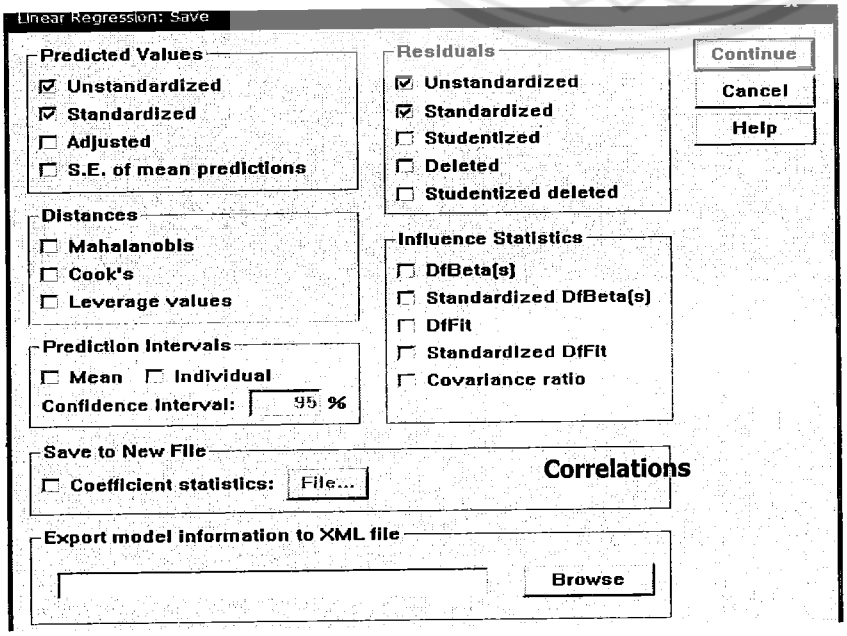
4. เลือก statistics โปรแกรมจะแสดงหน้าจอ Linear Regression: statistics

- เลือก Durbin-Watson
- เลือก Discriptive

5. กด Continue







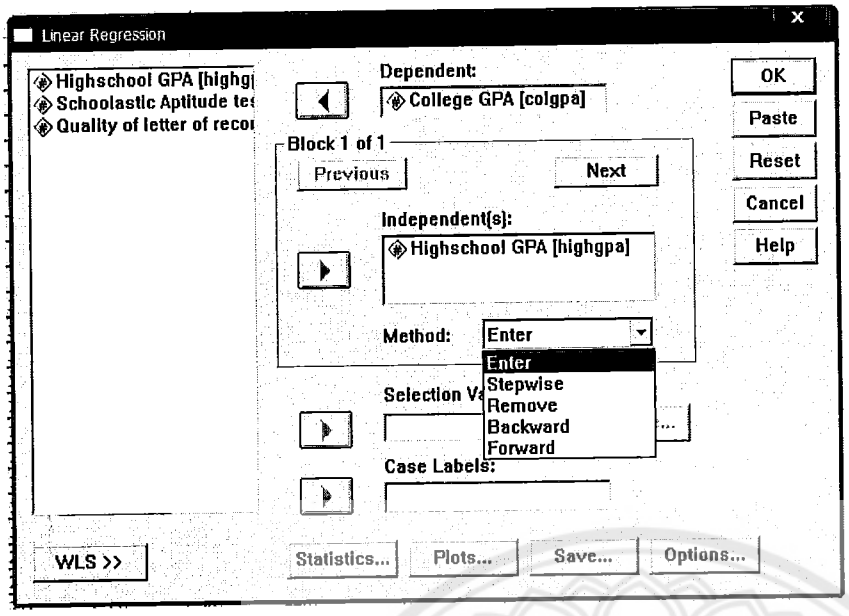
6. เลือก Plot โปรแกรมจะแสดงหน้าจอ Linear Regression: Plot

- เลือก ZPRED ไปที่แกน x
- เลือก ZRESID ไปที่แกน y

7. ที่ Standardized Residual Plots เลือก Histogram และ Normal probability plot

8. Save กำหนดระดับนัยสำคัญทางสถิติ

9. กด OK



การแปรผลจากหน้าจอ output

1. ทดสอบ Assumption ทำเหมือนวิธี Enter ข้อ 1.1-1.5

การตรวจสอบความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระกับตัวแปรตาม

Correlations

		College GPA	Highschool GPA	Schoolastic Aptitude test	Quality of letter of recommendation
Pearson Correlation	College GPA	1.000	.656	.747	.443
	Highschool GPA	.656	1.000	.447	.570
	Schoolastic Aptitude test	.747	.447	1.000	.208
	Quality of letter of recommendation	.443	.570	.208	1.000
Sig. (1-tailed)	College GPA	.	.000	.000	.007
	Highschool GPA	.000	.	.007	.001
	Schoolastic Aptitude test	.000	.007	.	.135
	Quality of letter of recommendation	.007	.001	.135	.
N	College GPA	30	30	30	30
	Highschool GPA	30	30	30	30
	Schoolastic Aptitude test	30	30	30	30
	Quality of letter of recommendation	30	30	30	30

a Dependent Variable: College GPA

2. ค่าคลาดเคลื่อน<u>ต้อง</u>เป็นอิสระกัน

Model Summary(c)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.747(a)	.558	.542	.60885	.558	35.283	1	28	.000	
2	.829(b)	.688	.664	.52099	.130	11.240	1	27	.002	1.513

a Predictors: (Constant), Schoolastic Aptitude test
b Predictors: (Constant), Schoolastic Aptitude test, Highschool GPA
c Dependent Variable: College GPA

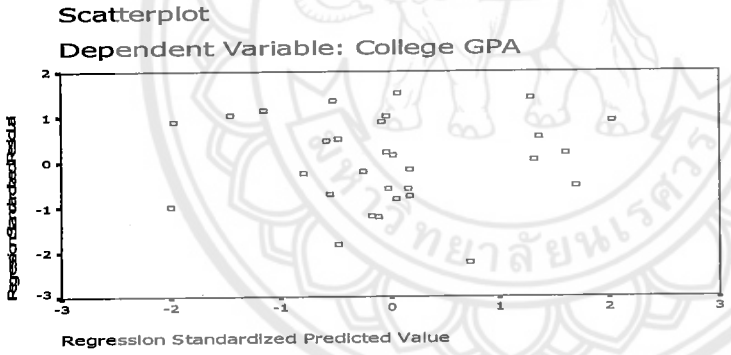
3.ค่าคลาดเคลื่อน<u>ต้อง</u>มีการแจกแจงแบบปกติ

Residuals Statistics(a)

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	.6169	3.6648	1.9263	.74579	30
Residual	-.9514	1.0901	.0000	.50270	30
Std. Predicted Value	-1.756	2.331	.000	1.000	30
Std. Residual	-1.826	2.092	.000	.965	30

a Dependent Variable: College GPA

4. ค่าแปรปรวนของค่าคาดเคลื่อน<u>ต้อง</u>มีค่าคงที่



5. 5. ตัวแปรอิสระต้องไม่มีความสัมพันธ์กันเอง (Multicollinearity) (Correlation) ดูค่า r ในพื้นที่สามเหลี่ยมชายธง (ในตาราง **Correlations**) หรือดูค่า Toleranc หรือ VIF

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	-.845	.480		-1.762	.089	-1.827	.137		
	Schoolastic Aptitude test	.003	.000	.747	5.940	.000	.002	.004	1.000	1.000
2	(Constant)	-1.048	.415		-2.527	.018	-1.900	-.197		
	Schoolastic Aptitude test	.002	.000	.567	4.714	.000	.001	.003	.801	1.249
	Highschool GPA	.438	.131	.403	3.353	.002	.170	.705	.801	1.249

2 ทดสอบสมมติฐาน ตั้งสมติฐานเหมือนกับวิธี Enter

แปลผลโดยการอ่านค่า Sig. จากตาราง ANOVA และ Coefficient ให้อ่านที่ Model สุดท้าย ของตาราง

ANOVA(c)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	13.079	1	13.079	35.283	.000(a)
	Residual	10.379	28	.371		
	Total	23.458	29			
2	Regression	16.130	2	8.065	29.713	.000(b)
	Residual	7.329	27	.271		
	Total	23.458	29			

a Predictors: (Constant), Schoolastic Aptitude test
b Predictors: (Constant), Schoolastic Aptitude test, Highschool GPA
c Dependent Variable: College GPA

จากตารางจะพบว่าค่า Sig. = .000 < .05ยอมรับ H1 แสดงว่ามี X (อย่างน้อย 1 ตัว) มีส่วนร่วมในการพยากรณ์ Y และอ่านค่า R Square(ในตาราง Model Summary(c)) X สามารถพยากรณ์ Y ได้ 68.8%

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	-.845	.480		-1.762	.089	-1.827	.137		
	Schoolastic Aptitude test	.003	.000	.747	5.940	.000	.002	.004	1.000	1.000
2	(Constant)	-1.048	.415		-2.527	.018	-1.900	-.197		
	Schoolastic Aptitude test	.002	.000	.567	4.714	.000	.001	.003	.801	1.249
	Highschool GPA	.438	.131	.403	3.353	.002	.170	.705	.801	1.249

a Dependent Variable: College GPA

อ่านค่า Beta จากตาราง จะพบว่าตัวแปรที่มีค่า Beta มาก จะมีค่าการพยากรณ์มากกว่าตัวอื่น Beta ของ SAT = .567 , High school GPA = .403 (**โดยโปรแกรมจะตัดตัวที่ไม่ Sig. ออกไป**)

จากตารางจะพบว่าค่า Sig. ของ Schoolastic Aptitude GPA = .000 < .05 ยอมรับ H1

High school GPA = 0.002 < .05 ยอมรับ H1

แสดงว่า Schoolastic Aptitude GPA และ High school GPA สามารถพยากรณ์ค่า College GPA

เขียนสมการพยากรณ์ได้ดังนี้

$$y' = b0+b1x1+b2x2$$

College GPA = -1.048 + .002(SAT) + .438(High school GPA)

การวิเคราะห์ MRA โดยวิธี Backward

วิธี Backward นำตัวแปรเข้าทั้งหมด โปรแกรมจะตัดออกทีละตัว

Variables Entered/Removed(b)

Model	Variables Entered	Variables Removed	Method
1	Quality of letter of recommendation, Schoolastic Aptitude test, Highschool GPA(a)	.	Enter
2	.	Quality of letter of recommendation	Backward (criterion: Probability of F-to-remove >= .100).

a All requested variables entered.
b Dependent Variable: College GPA

โปรแกรมจะตัดข้อมูลของ X ตัวที่มีค่าความสัมพันธ์กับ Y (r) ต่ำสุดออกก่อนเป็นอันดับแรก (โปรแกรมจะสลับเอาตัวที่มีความสัมพันธ์น้อยกว่าเข้ามาก่อน) ๆ นำเข้าทั้งหมด แล้ว จึงตัดตัวที่มีความสัมพันธ์น้อยสุดออกทีละตัว

ขั้นตอนการวิเคราะห์ข้อมูล

1. ทดสอบ Assumption ทำเหมือนวิธี Enter ข้อ 1.1-1.5 แต่ให้ระบุวิธีการวิเคราะห์เป็น BackWard

การตรวจสอบความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระกับตัวแปรตาม
- ค่า r เข้าใกล้ 1
- ค่า sig. < 0.05

ตัวแปรอิสระต้องไม่มีความสัมพันธ์กันเอง (Multicollinearity) (Correlation) ดูค่า r ในพื้นที่สามเหลี่ยมชายธง < 0.8

Correlations

		College GPA	Highschool GPA	Schoolastic Aptitude test	Quality of letter of recommendation
Pearson Correlation	College GPA	1.000	.656	.747	.443
	Highschool GPA	.656	1.000	.447	.570
	Schoolastic Aptitude test	.747	.447	1.000	.208
	Quality of letter of recommendation	.443	.570	.208	1.000
Sig. (1-tailed)	College GPA	.	.000	.000	.007
	Highschool GPA	.000	.	.007	.001
	Schoolastic Aptitude test	.000	.007	.	.135
	Quality of letter of recommendation	.007	.001	.135	.
N	College GPA	30	30	30	30
	Highschool GPA	30	30	30	30
	Schoolastic Aptitude test	30	30	30	30
	Quality of letter of recommendation	30	30	30	30

Residuals Statistics(a)

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	.6169	3.6648	1.9263	.74579	30
Residual	-.9514	1.0901	.0000	.50270	30
Std. Predicted Value	-1.756	2.331	.000	1.000	30
Std. Residual	-1.826	2.092	.000	.965	30

a Dependent Variable: College GPA

ค่าคลาดเคลื่อนต้องมีการแจกแจงแบบปกติ

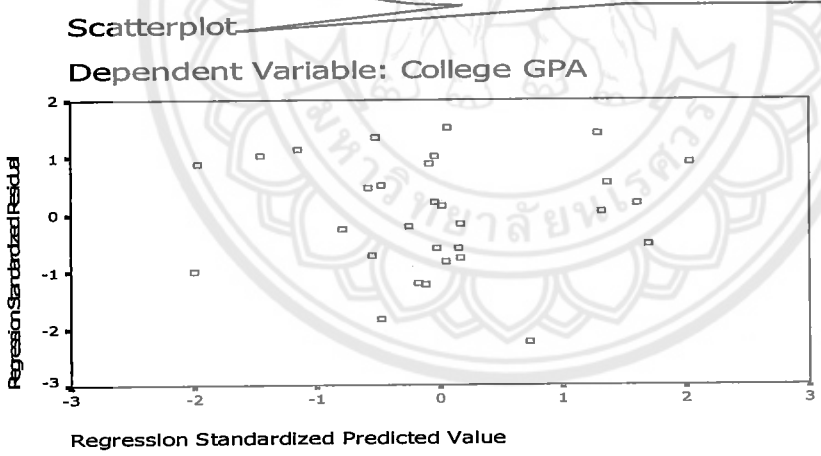
Model Summary(c)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.837(a)	.701	.666	.51948	.701	20.310	3	26	.000	
2	.829(b)	.688	.664	.52099	-.013	1.157	1	26	.292	1.513

a Predictors: (Constant), Quality of letter of recommendation, Schoolastic Aptitude test, Highschool GPA
b Predictors: (Constant), Schoolastic Aptitude test, Highschool GPA
c Dependent Variable: College GPA

ค่าคลาดเคลื่อนต้องเป็นอิสระกัน

ค่าแปรปรวนของค่าคาดเคลื่อนต้องมีค่าคงที่



2 . ทดสอบสมมติฐาน ตั้งสมติฐานเหมือนกับวิธี Enter

แปลผลโดยการอ่านค่า Sig. จากตาราง ANOVA และ Coefficient ให้อ่านที่ Model สุดท้าย ของตาราง

ANOVA(c)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	16.442	3	5.481	20.310	.000(a)
	Residual	7.016	26	.270		
	Total	23.458	29			
2	Regression	16.130	2	8.065	29.713	.000(b)
	Residual	7.329	27	.271		
	Total	23.458	29			

a Predictors: (Constant), Quality of letter of recommendation, Schoolastic Aptitude test, Highschool GPA
b Predictors: (Constant), Schoolastic Aptitude test, Highschool GPA
c Dependent Variable: College GPA

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	-1.341	.495		-2.709	.012	-2.358	-.323		
	Highschool GPA	.347	.155	.319	2.232	.034	.027	.666	.563	1.777
	Schoolastic Aptitude test	.002	.000	.575	4.786	.000	.001	.003	.797	1.254
	Quality of letter of recommendation	.091	.084	.141	1.076	.292	-.082	.263	.672	1.488
2	(Constant)	-1.048	.415		-2.527	.018	-1.900	-.197		
	Highschool GPA	.438	.131	.403	3.353	.002	.170	.705	.801	1.249
	Schoolastic Aptitude test	.002	.000	.567	4.714	.000	.001	.003	.801	1.249

a Dependent Variable: College GPA

สรุป

จากตาราง ANOVA จะพบว่าค่า Sig. = .000 < .05ยอมรับ H1 แสดงว่ามี X (อย่างน้อย 1 ตัว) มีส่วนร่วมในการพยากรณ์ Y และอ่านค่า R Square X สามารถพยากรณ์ Y ได้ 68.8%

อ่านค่า Beta จากตาราง จะพบว่าตัวแปรที่มีค่า Beta มาก จะมีค่าการพยากรณ์มากกว่าตัวอื่น Beta ของ SAT = .567 High school GPA = .403 **(โดยโปรแกรมจะตัดตัวที่ไม่ Sig. ออกไป)**

จากตารางจะพบว่าค่า Sig. ของ High school GPA = 0.002 < .05 ยอมรับ H1

Schoolastic Aptitude GPA = .000 < .05 ยอมรับ H1

แสดงว่า High school GPA และ Schoolastic Aptitude GPA สามารถพยากรณ์ค่า College GPA

เขียนสมการพยากรณ์ได้ดังนี้

$$y' = b0+b1x1+b2x2$$

College GPA = -1.048 +.438(High school GPA) + .002(SAT)

ตัวแปรหุ่นในการวิเคราะห์การถดถอยพหุคูณ
(Dummy Variables in Multiple Regression Analysis)

หลักการสร้างตัวแปรหุ่น

1. ตัวแปรหุ่น คือ ตัวแปรที่มีค่าสองค่า (dichotomy) เป็น 1 และ 0 จะใช้เมื่อเป็นตัวแปรต้นเป็นตัวแปรเชิงกลุ่ม
2. ตัวแปรอิสระเดิมที่มีระดับการวัดเป็นกลุ่ม จะถูกแทนที่ด้วยตัวแปรหุ่น จำนวน k-1 ตัวแปร เช่น เพศ มี สองกลุ่ม จะสร้างตัวแปรหุ่น 1 ตัวแปร อาชีพ มี 6 กลุ่ม จะสร้างตัวแปรหุ่น 5 ตัวแปร
3. นำตัวแปรหุ่นเข้าไปวิเคราะห์ MRA ร่วมกับตัวแปรอื่นๆ ทั้งที่เป็นตัวแปรหุ่นด้วยกัน (0 , 1) และตัวแปรเชิงปริมาณอื่นๆ
4. ตัวแปรตัวไหนที่จะใช้อ้างอิงให้เป็น 0 (ขึ้นอยู่กับผู้วิจัยจะตั้ง)
5. การวิเคราะห์และตีความ กลุ่มอ้างอิง คือ กลุ่มที่มีค่าเป็น 0

ตัวอย่าง ที่ 1

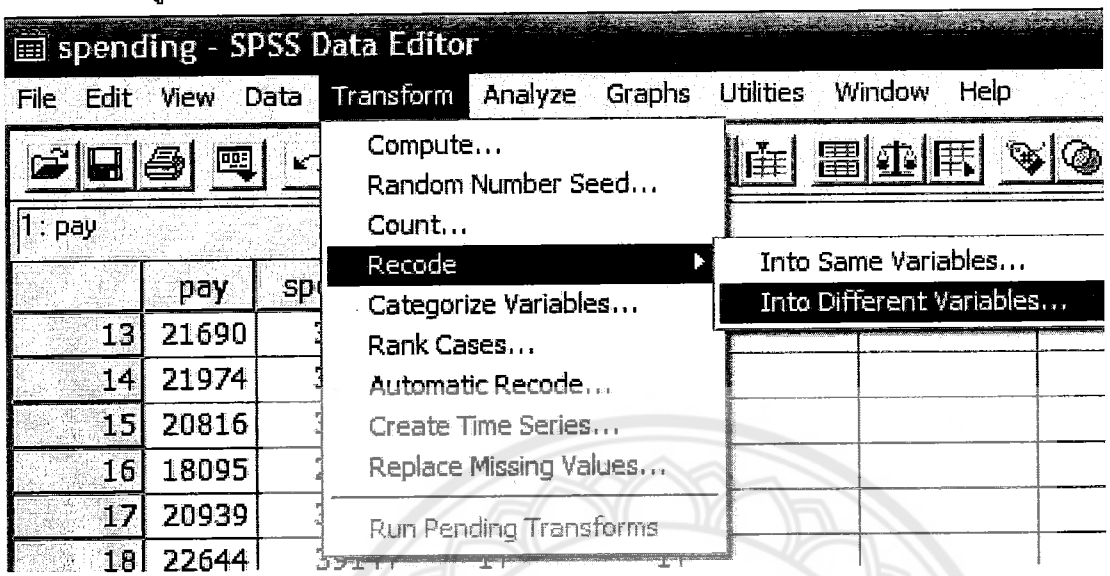
ตัวแปรดั้งเดิม คือ area มีรหัสเป็น 1= North 2 = South และ 3 = West

สร้างตัวแปรหุ่น ตั้งตัวแปรใหม่ 2 ตัว คือ north และ south ต้อง recode ตัวที่สร้างเป็นตัวแปรให้ตั้งค่าเป็น 1 File: Spending กำหนดให้ West เป็นกลุ่มอ้างอิง

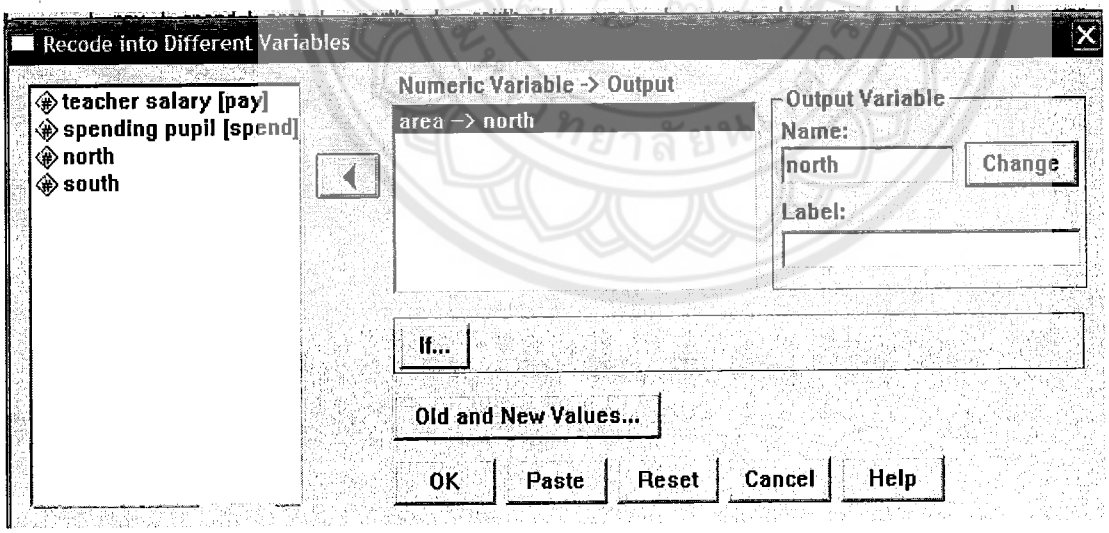
pay	spend	area
19583	3346	1
20263	3114	1
24624	4517	2
27186	4349	2
20969	2509	3
27224	5440	3

วิธีการสร้างตัวแปรหุ่น

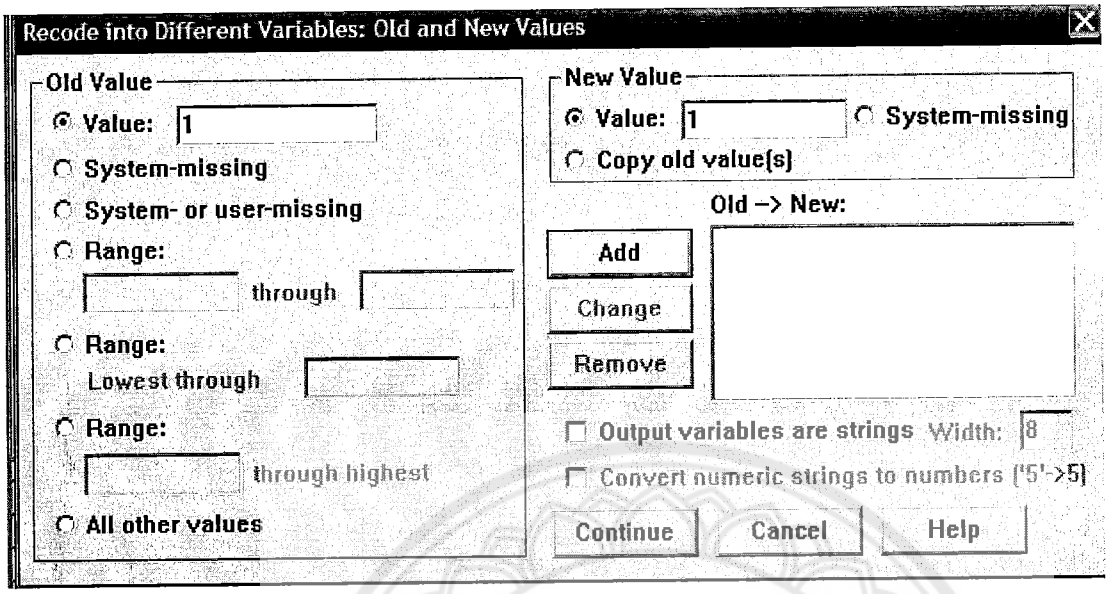
1. เข้าเมนู Transform > Recode > Into Different Variables…



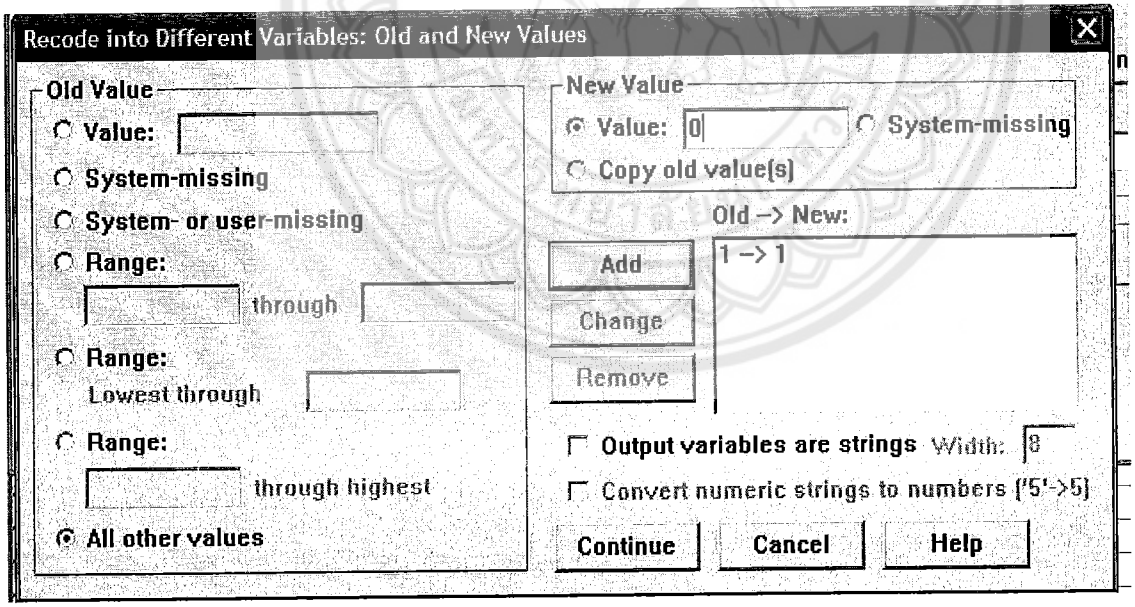
2. สร้างตัวแปรหุ่น คือ north โดยเลือก area เข้ามาในช่อง Numeric Variable แล้วพิมพ์ชื่อตัวแปรใหม่เป็น north ในช่อง Output Variable แล้วกด Change

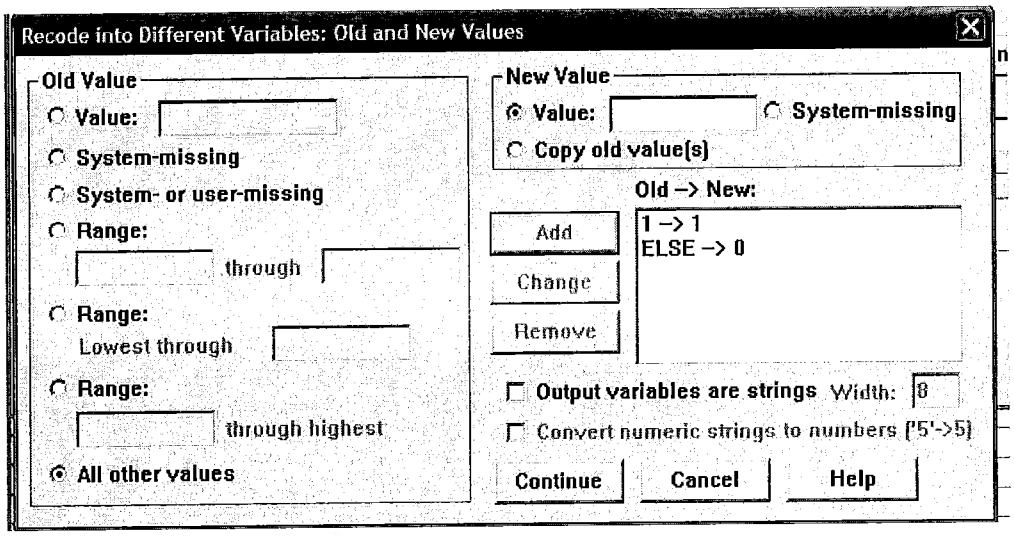


3. เข้าไปที่ Old and New Values... เติม 1 ในช่อง Value ในด้าน Old Value และเติม 1 ในช่อง Value ในด้าน New Value แล้วกด Add (North เป็น 1 อื่นๆเป็น 0)

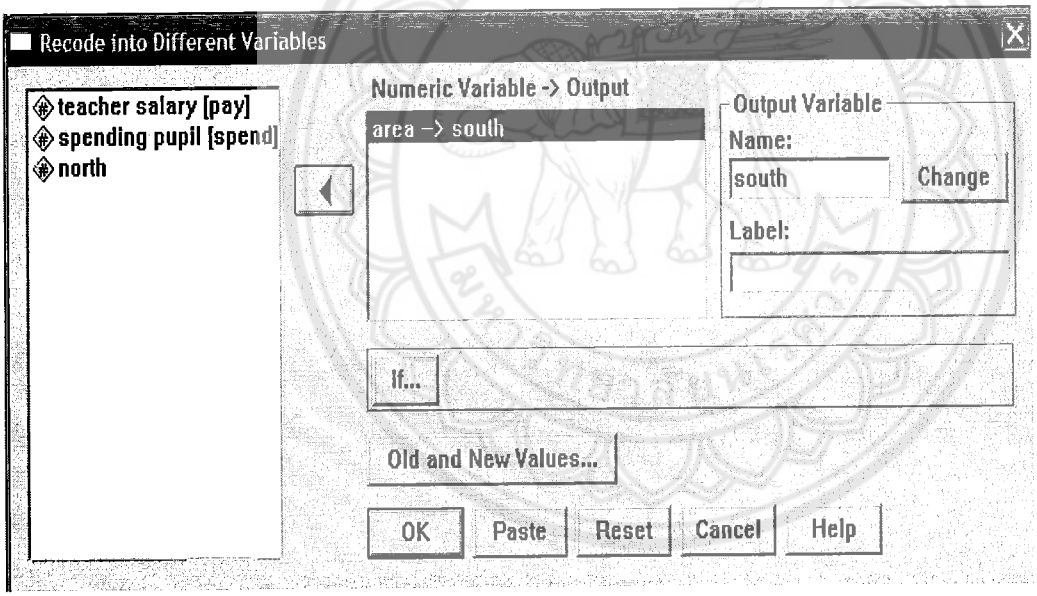


4. เลือก All other values ในด้าน Old Value และเติม 0 ในช่อง Value ในด้าน New Value แล้วกด Add (หมายถึงถ้าเป็นตัวอื่นที่ไม่ใช่ north จะให้แทนค่าเป็น 0)

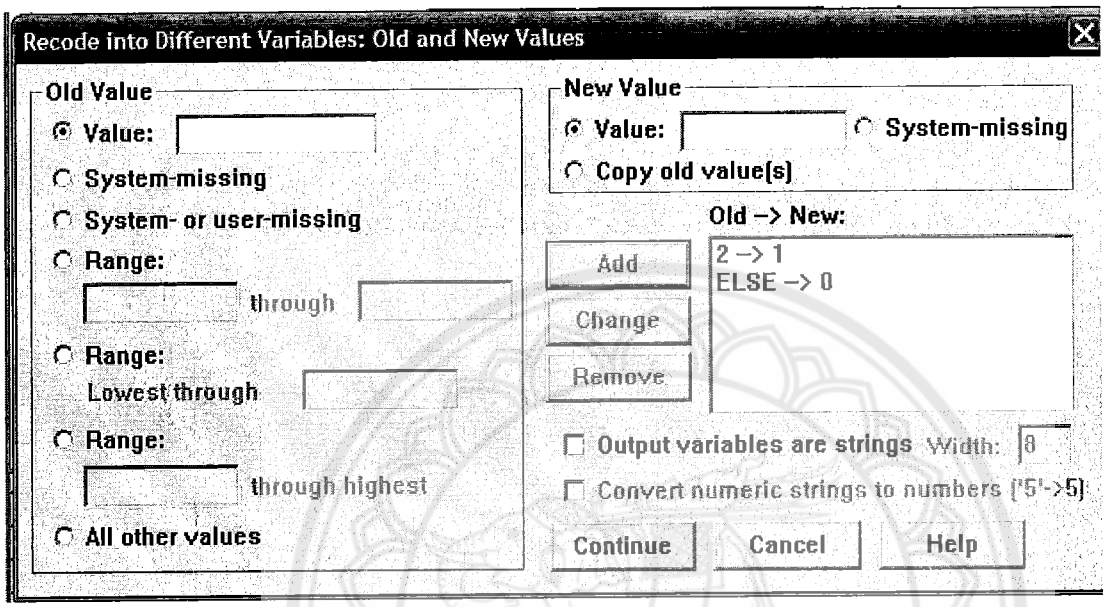




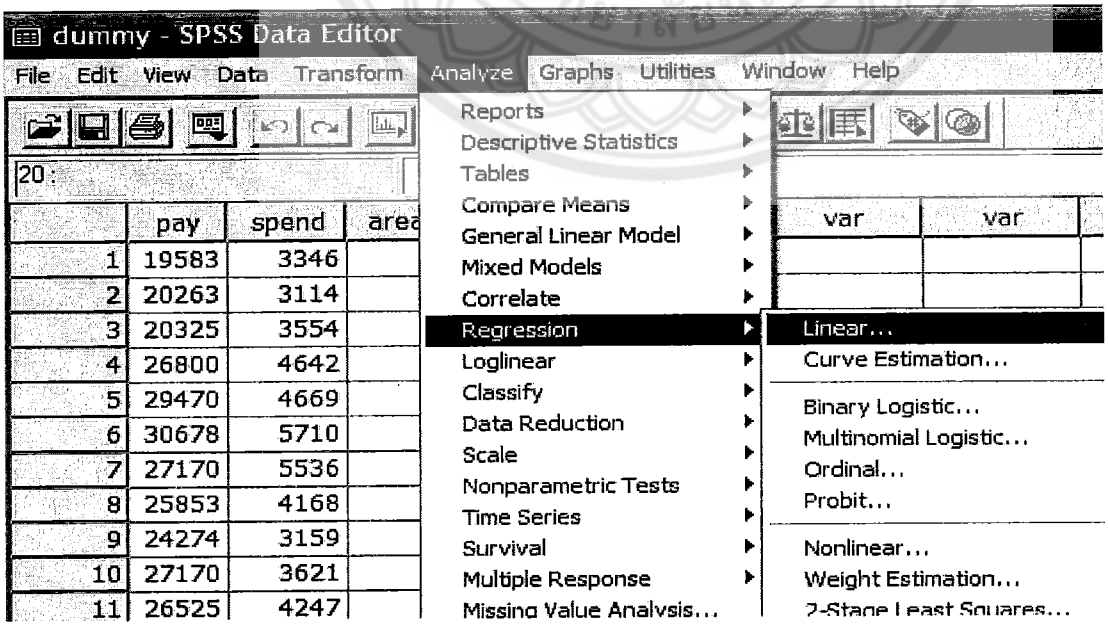
5. สร้างตัวแปรหุ่น คือ south โดยเลือก area เข้ามาในช่อง Numeric Variable แล้วพิมพ์ชื่อตัวแปรใหม่เป็น south ในช่อง Output Variable แล้วกด Change

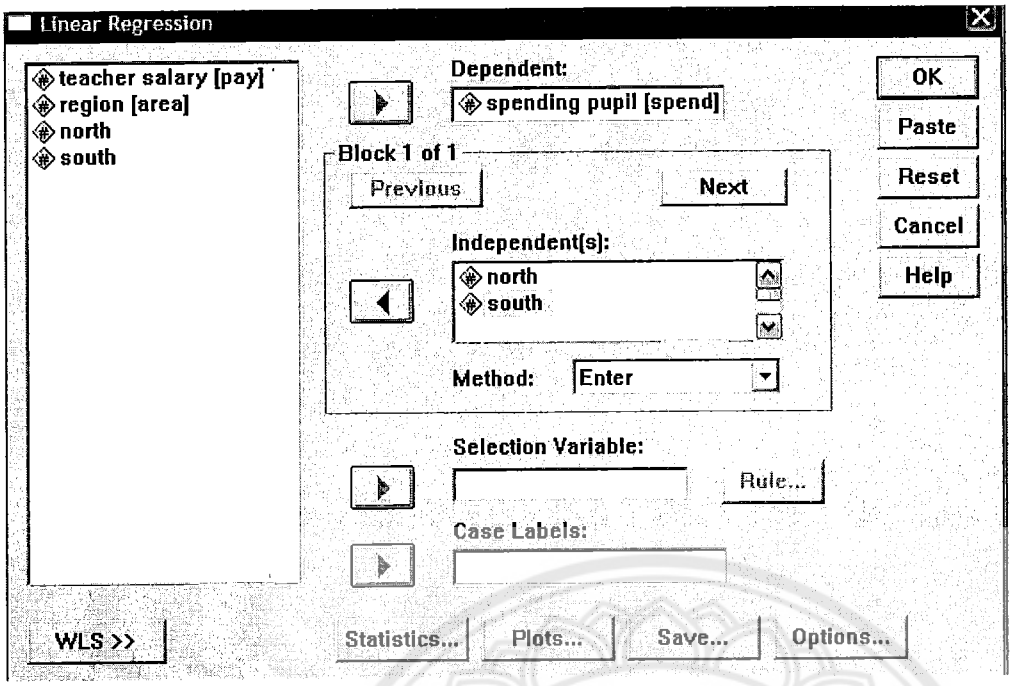


6. เข้าไปที่ Old and New Values... เติม 2 ในช่อง Value ในด้าน Old Value และเติม 1 ในช่อง Value ในด้าน New Value แล้วกด Add (south เป็น 1 อื่นๆเป็น 0) เลือก All other values ในด้าน Old Value และเติม 0 ในช่อง Value ในด้าน New Value แล้วกด Add (หมายถึงถ้าเป็นตัวอื่นที่ไม่ใช่ south จะให้แทนค่าเป็น 0)



วิเคราะห์ด้วย MRA ไปที่ Analyze → Regression → Linear...ถ้าเลือกวิธี Enter ผลจะออกตารางออกมาให้ดูเยอะ ถ้าเลือก stepwise ตารางจะออกน้อย ตัวแปรเดิมไม่ต้องเอาเข้า (area ไม่ต้องเลือก)





ผลการวิเคราะห์

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3061837	2	1530918.682	1.269	.292[a]
	Residual	4.8E+07	40	1206841.883		
	Total	5.1E+07	42			

a. Predictors: (Constant), SOUTH, NORTH

b. Dependent Variable: spending pupil

Sig > 0.05 ตัวแปรพื้นที่ ไม่สามารถทำนายได้ (x ไม่ทำนาย y)

เมื่อนำตัวแปรหุ่นทั้งสองไปวิเคราะห์ร่วมกับตัวแปรอื่นอีก 1 ตัวแปร ได้สมการพยากรณ์ดังนี้

$$y' = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3$$

$$y' = b_0 + b_1 x_1 + b_2 \text{ north} + b_3 \text{ south}$$

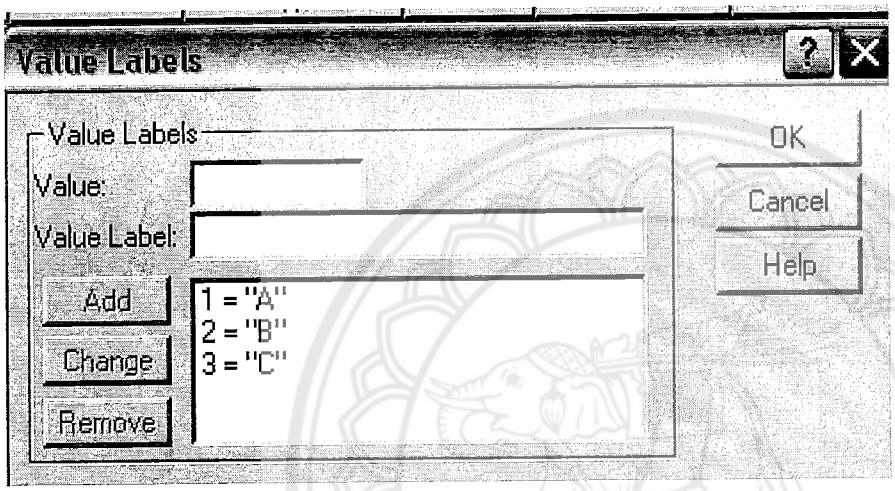
สัมประสิทธิ b_2 จะเปรียบเทียบค่าเฉลี่ยของตัวแปรตาม ระหว่างคนที่อยู่ภาคเหนือกับ ภาคตะวันตก และ b_3 จะเปรียบเทียบค่าเฉลี่ยของตัวแปรตามระหว่างคนที่อยู่ภาคใต้กับภาคตะวันตก

ตัวอย่างที่ 2

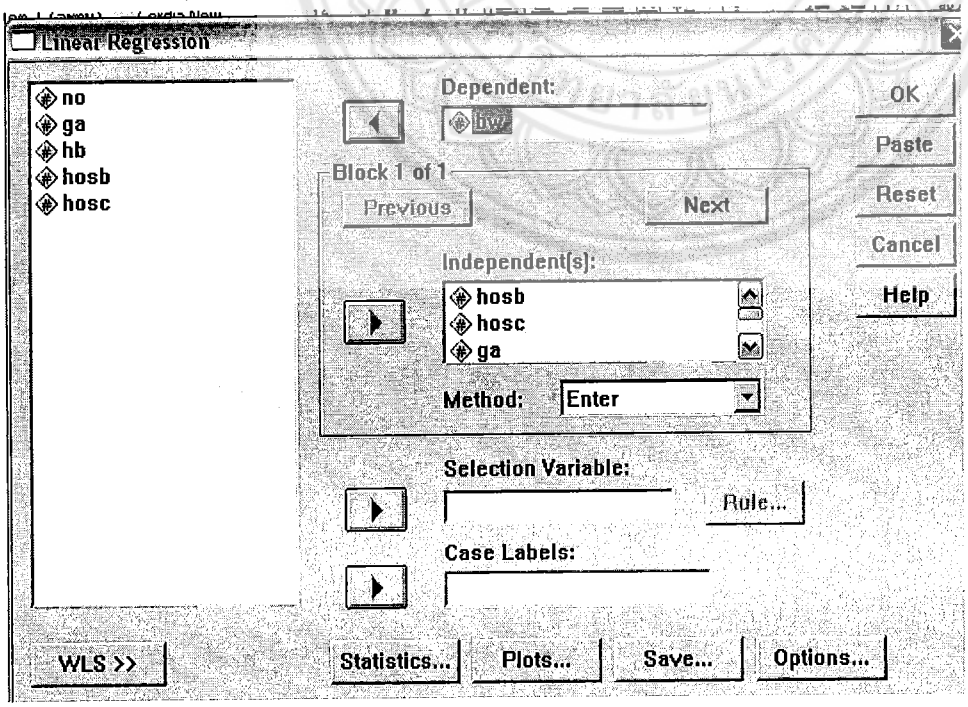
ข้อมูลจากตารางเป็นส่วนหนึ่งของการศึกษาเกี่ยวกับทารกที่คลอดก่อนกำหนด ซึ่งมีน้ำหนักแรกเกิดต่ำกว่าเกณฑ์ (< 2.5 kg) ของโรงพยาบาล 3 แห่ง จงวิเคราะห์ข้อมูลโดยเลือกใช้สถิติตามความเหมาะสม กำหนดให้ รพ. A เป็น Reference group

ศึกษาปัจจัยที่ส่งผลต่อ low birth weigh

1. แทนค่าให้ รพ. A เป็น 1 , รพ. B เป็น 2 และรพ. C เป็น 3 แล้วกด OK



2. สร้างตัวแปรหุ่น HOSB, HOSC แล้วนำไปวิเคราะห์ MRA ด้วยวิธี Enter



ผลการวิเคราะห์

1. ทดสอบข้อตกลงเบื้องต้น

1.1 x และ y มีความสัมพันธ์เชิงเส้นตรง

H_0: x กับ y ไม่มีความสัมพันธ์เชิงเส้นตรง

H_1: x กับ y มีความสัมพันธ์เชิงเส้นตรง

โดย x_1 คือ รพ. B เทียบกับกับ A , x_2 คือ รพ. C เทียบกับ A , x_3 คือ Gestation Age และ y คือ น้ำหนักแรกคลอดของทารก ที่ $\alpha < 0.05$

Variables Entered/Removed(b)

Model	Variables Entered	Variables Removed	Method
1	HOSC, GA, HOSB(a)	.	Enter

a All requested variables entered.

b Dependent Variable: BW

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.850	3	.950	43.069	.000(a)
	Residual	.794	36	.022		
	Total	3.644	39			

a Predictors: (Constant), HOSC, GA, HOSB

b Dependent Variable: BW

Enter ค่า Sig. มีอันเดียวเพราะนำข้อมูลเข้าครั้งเดียว แต่วิธีอื่นจะมากตามตัวแปรเพราะนำเข้าทีละคู่ แต่อ่านคู่ท้ายสุด

จากตาราง ANOVA ค่า Sig.= 0.000 ซึ่ง < 0.05 จะยอมรับ H_1 คือ x กับ y มีความสัมพันธ์เชิงเส้นตรง ที่ระดับนัยสำคัญ 0.05

1.2 ค่าคลาดเคลื่อนมีการแจกแจงปกติ

Residuals Statistics(a)

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	.749	1.813	1.270	.2703	40
Residual	-.295	.332	.000	.1427	40
Std. Predicted Value	-1.927	2.009	.000	1.000	40
Std. Residual	-1.989	2.235	.000	.961	40

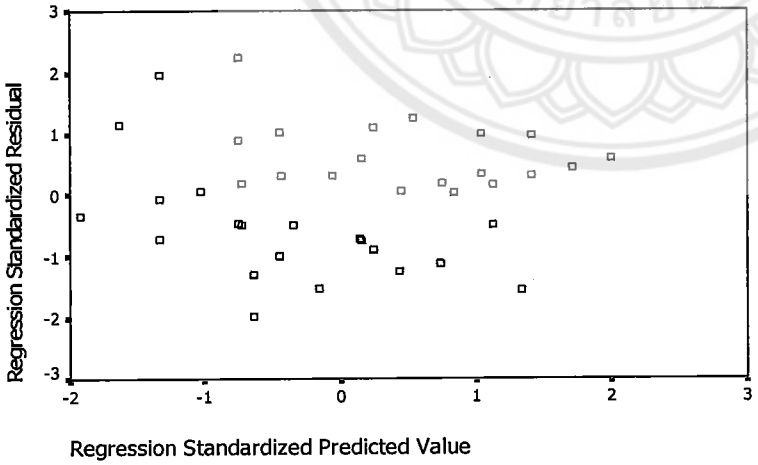
a Dependent Variable: BW

จากตาราง Residuals Statistics จะพบค่า mean = 0.000 ซึ่งเข้าใกล้ 0 และค่า SD = 0.961 ซึ่งเข้าใกล้ 1 แสดงว่าค่าคลาดเคลื่อนมีการแจกแจงแบบปกติ

1.3 ค่าแปรปรวนของค่าคลาดเคลื่อนมีค่าคงที่

Scatterplot

Dependent Variable: BW



จาก Scatterplot พบว่าข้อมูลมีการกระจายตัวไม่เป็นรูปแบบ แสดงว่า ค่าแปรปรวนของค่าคลาดเคลื่อนมีค่าคงที่

1.4 ค่าคลาดเคลื่อนต้องเป็นอิสระต่อกัน

Model Summary(b)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.884(a)	.782	.764	.1485	.782	43.069	3	36	.000	1.608

a Predictors: (Constant), HOSC, GA, HOSB

b Dependent Variable: BW

เปลี่ยน Durbin เป็น 1.608 จากตาราง Model Summary จะพบว่าค่า Durbin-Watson = 1.608 ซึ่งอยู่ในช่วง 1.5 – 2.5 แสดงว่าค่าคลาดเคลื่อนมีความเป็นอิสระต่อกัน

1.5 ตัวแปรอิสระต้องไม่มีความสัมพันธ์กันเอง

H_0 : x ไม่มีความสัมพันธ์กันเอง

H_1 : x มีความสัมพันธ์กันเอง ที่ $\alpha < 0.05$

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-1.325	.266		-4.974	.000		
	GA	.080	.009	.741	9.098	.000	.912	1.097
	HOSB	.267	.056	.421	4.732	.000	.764	1.309
	HOSC	.083	.062	.123	1.355	.184	.730	1.371

a Dependent Variable: BW

เข้าใกล้ 1

จากตาราง Coefficients จะอ่านค่า Tolerance ของ x_1 = 0.764, x_2 = 0.730, x_3 = 0.912 ซึ่งไม่เกิน และค่า VIF ของ x_1 = 1.309, x_2 = 1.371, x_3 = 1.097 ซึ่งไม่เกิน 10 แสดงว่า ยอมรับ H_0 คือ x ไม่มีความสัมพันธ์กันเอง ที่ระดับนัยสำคัญ 0.05

สรุป คือ จากการทดสอบข้อตกลงเบื้องต้น จะทดสอบผ่านทั้ง 5 ข้อ จะสามารถใช้การวิเคราะห์แบบ MRA ได้

(ดูว่าตัวไหนพยากรณ์ได้ มีค่า Sig. แล้วจะดูว่าตัวไหนพยากรณ์ได้มากกว่ากันดูค่า Beta ที่ Std. Coefficient โดยเครื่องหมาย +/- ไม่มีความหมายอะไรบอกแค่ทิศทาง อันนี้ GA พยากรณ์ได้มากกว่า HOSB ส่วนค่า B ในช่อง Unstd. จะใช้เขียนสมการพยากรณ์)

2. **ทดสอบสมมติฐาน** (Overall test)

H_0 : x ไม่มีส่วนร่วมในการพยากรณ์ y

H_1 : x มีส่วนร่วมในการพยากรณ์ y อย่างน้อย 1 ตัว ที่ $\alpha < 0.05$

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.850	3	.950	43.069	.000(a)
	Residual	.794	36	.022		
	Total	3.644	39			

a Predictors: (Constant), HOSC, GA, HOSB

b Dependent Variable: BW

Model Summary(b)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.884(a)	.782	.764	.1485	.782	43.069	3	36	.000	1.608

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-1.325	.266		-4.974	.000		
	GA	.080	.009	.741	9.098	.000	.912	1.097
	HOSB	.267	.056	.421	4.732	.000	.764	1.309
	HOSC	.083	.062	.123	1.355	.184	.730	1.371

a Dependent Variable: BW

จากตาราง ANOVA มีค่า Sig. = 0.000 ซึ่ง Sig. < 0.05 จะยอมรับ H_1 คือ x มีส่วนร่วมในการพยากรณ์ y อย่างน้อย 1 ตัว จากตาราง Model Summary จะพบว่า x สามารถพยากรณ์ y ได้ 78.2 % (r^2 = 0.782) และจากตาราง Coefficients จะพบว่าตัวแปร x ที่สามารถพยากรณ์ y ได้มี 2 ตัว คือ GA และ HOSB โดยดูจากค่า Sig. < 0.05 และ GA สามารถพยากรณ์ y ได้ดีกว่า HOSB โดยดูจากค่า Beta ในช่อง Standardized Coefficients

3. **สมการพยากรณ์** (Partial test)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-1.325	.266		-4.974	.000		
	GA	.080	.009	.741	9.098	.000	.912	1.097
	HOSB	.267	.056	.421	4.732	.000	.764	1.309
	HOSC	.083	.062	.123	1.355	.184	.730	1.371

a Dependent Variable: BW

จากตาราง Coefficients จะสามารถเขียนสมการพยากรณ์ได้ดังนี้คือ

Birth Weight = -1.325 + 0.08 (GA) + 0.267 (HOSB) + 0.083 (HOSC)

แต่จะไม่แทนค่า HOSC เพราะไม่ Sig. ? ขณะที่ใช้สมการพยากรณ์ ต้องแทนค่าทุกตัวแปร แต่การใช้วิธี Enter ไม่ได้มีวัตถุประสงค์เพื่อสร้างสมการ หากแต่ใช้เพื่อหาความสัมพันธ์ระหว่าง x กับ y โดยที่ adjust x ตัวอื่นๆ