บทที่ 10

การวิเคราะห์ความสัมพันธ์และการถดถอย

การวิเคราะห์ความสัมพันธ์หรือสหสัมพันธ์นั้น เป็นการศึกษาความสัมพันธ์ของตัวแปรตั้งแต่ 2 ตัวขึ้นไปว่ามีความสัมพันธ์กันหรือไม่ ลักษณะความสัมพันธ์นั้นมากน้อยเพียงใด การวิเคราะห์สหสัมพันธ์ แบ่งเป็น 3 ประเภท คือ การวิเคราะห์สหสัมพันธ์อย่างง่าย การวิเคราะห์สหสัมพันธ์เชิงซ้อน และสหสัมพันธ์เชิงส่วน ส่วนการวิเคราะห์การถดถอย เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรตามหนึ่งตัว กับตัวแปรอิสระตั้งแต่ 1 ตัวแปรขึ้นไป โดยมีวัตถุประสงค์เพื่อประมาณค่าของตัวแปรตาม เมื่อทราบค่าของตัวแปรอิสระนั้นแล้ว การวิเคราะห์การถดถอย แบ่งได้เป็น 2 ประเภท คือ การวิเคราะห์การถดถอยอย่างง่าย และการวิเคราะห์การถดถอยเชิงซ้อน ซึ่งเป็นสาระสำคัญที่ผู้ทำวิจัยควรทราบ ดังที่ได้นำเสนอไว้แล้วในบทนี้

การวิเคราะห์สหสัมพันธ์อย่างง่าย

การวิเคราะห์สหสัมพันธ์อย่างง่าย (simple correlation analysis) เป็นการศึกษาระดับความสัมพันธ์ของ X กับ Y ว่า มีความสัมพันธ์กันมากน้อยเพียงใด และมีทิศทางความสัมพันธ์อย่างไร เมื่อ X เป็นตัวแปรอิสระ และ Y เป็นตัวแปรตาม เช่น ความสัมพันธ์ระหว่างผลสัมฤทธิ์ทางการเรียนของวิชาคณิตศาสตร์และวิชาสถิติ ความสัมพันธ์ระหว่างส่วนสูง และน้ำหนัก เป็นต้น สิ่งที่จำเป็นต้องทราบในการวิเคราะห์สหสัมพันธ์อย่างง่ายมี 3 ประการ คือ ค่าสัมประสิทธิ์ความสัมพันธ์อย่างง่าย ค่าสัมประสิทธิ์การตัดสินใจ และการทดสอบสมมติฐานความสัมพันธ์ ดังรายละเอียดต่อไปนี้

1. ค่าสัมประสิทธิ์ความสัมพันธ์อย่างง่าย

ค่าสัมประสิทธิ์ความสัมพันธ์อย่างง่าย (simple correlation coefficent) เป็นค่าที่วัดความสัมพันธ์ของ X กับ Y ว่า มีขนาดและทิศทางของความสัมพันธ์อย่างไร กำหนดให้ ρ คือ สัมประสิทธิ์สหสัมพันธ์อย่างง่ายของประชากร เมื่อ $-1 \leq \rho \leq 1$ เนื่องจากในการวิเคราะห์

สหสัมพันธ์ไม่ได้เก็บข้อมูลจากประชากร แต่เป็นการเก็บข้อมูลจากกลุ่มตัวอย่าง ดังนั้นค่าสัมประสิทธิ์ที่ได้ จึงเป็นค่าสัมประสิทธิ์สหสัมพันธ์อย่างง่ายของกลุ่มตัวอย่าง ใช้สัญลักษณ์ r เมื่อ $-1 \leq r \leq 1$ การแปลความหมายของค่า r ใช้เกณฑ์ดังตารางที่ 10.1

ตารางที่ 10.1 *แสดงเกณฑ์ที่ใช้ในการแปลความหมายของสหสัมพันธ์อย่างง่าย*

ค่าสัมประสิทธิ์สหสัมพันธ์อย่างง่าย	ความหมาย
ค่า r เป็น +	x กับ y มีความสัมพันธ์ในทิศทางเดียวกัน
ค่า r เป็น –	x กับ y มีความสัมพันธ์ในทิศทางตรงกันข้าม
ค่า r เป็น 0	x กับ y ไม่มีความสัมพันธ์กันเลย
ค่า | r | มีค่าเข้าใกล้ 1	x กับ y มีความสัมพันธ์กันมาก
ค่า | r | มีค่าเข้าใกล้ 0	x กับ y มีความสัมพันธ์กันน้อย

2. ค่าสัมประสิทธิ์การตัดสินใจ

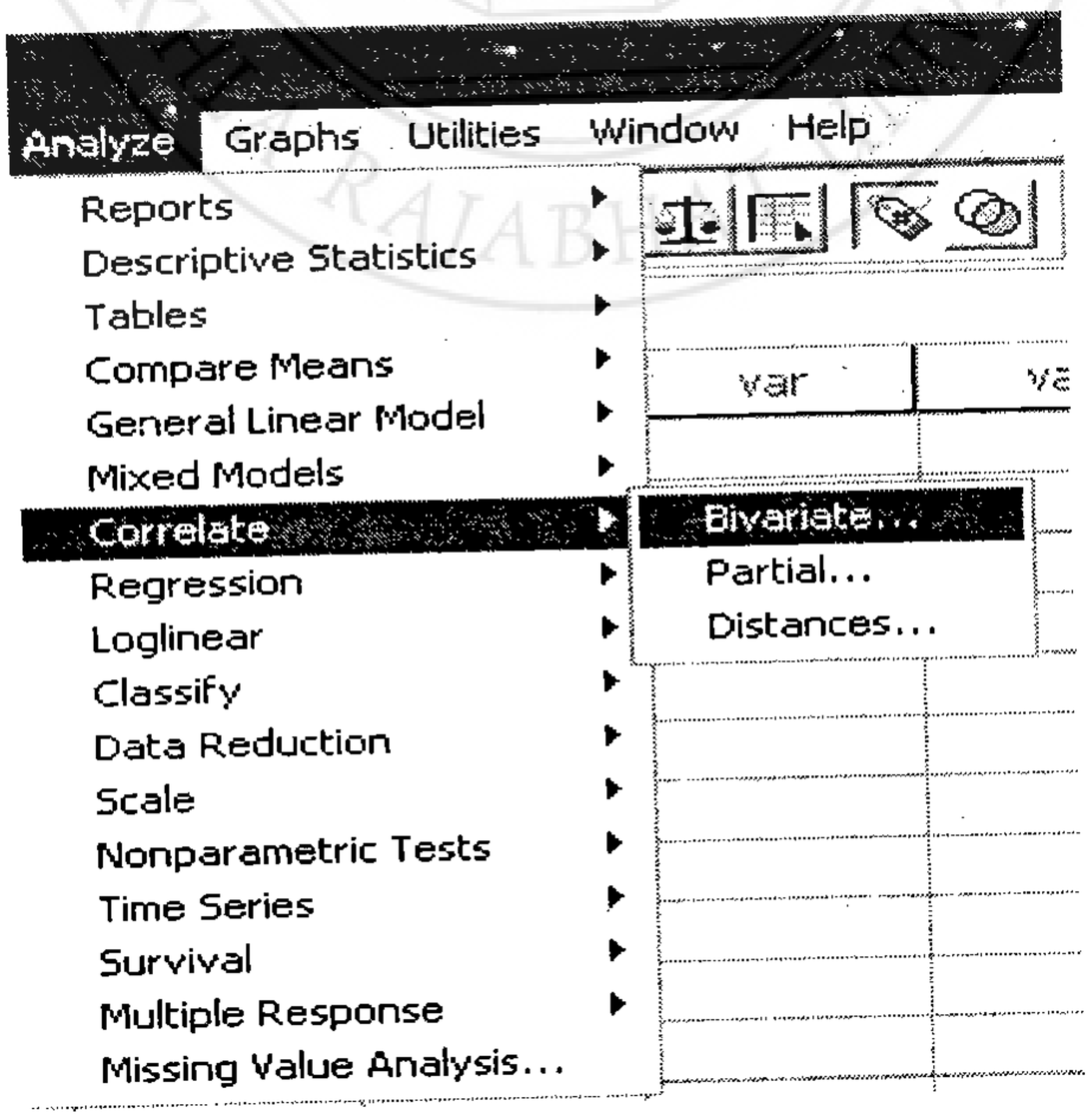
ค่าสัมประสิทธิ์การตัดสินใจ (coefficent of determination) เป็นค่าที่แสดงว่าตัวแปร X มีอิทธิพลต่อตัวแปร Y มากน้อยเพียงใด โดยที่ค่าสัมประสิทธิ์การตัดสินใจ มีค่าเท่ากับกำลังสองของสัมประสิทธิ์สหสัมพันธ์ หรือ "r^2" โดยที่ $0 \leq r^2 \leq 1$

ตัวอย่างที่ 10.1 แผนกวิจัยตลาดของบริษัท วิซาร์ทมิวสิค จำกัด ต้องการศึกษาความสัมพันธ์ระหว่างยอดขายและค่าใช้จ่ายในการโฆษณา ข้อมูลเกี่ยวกับค่าใช้จ่ายและยอดขายในรอบ 10 ปีที่ผ่านมา ดังในตาราง จงคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ (r) และสัมประสิทธิ์การตัดสินใจ (r^2) ของยอดขายสินค้าและค่าใช้จ่ายในการโฆษณาของบริษัทวิซาร์ทมิวสิค จำกัด

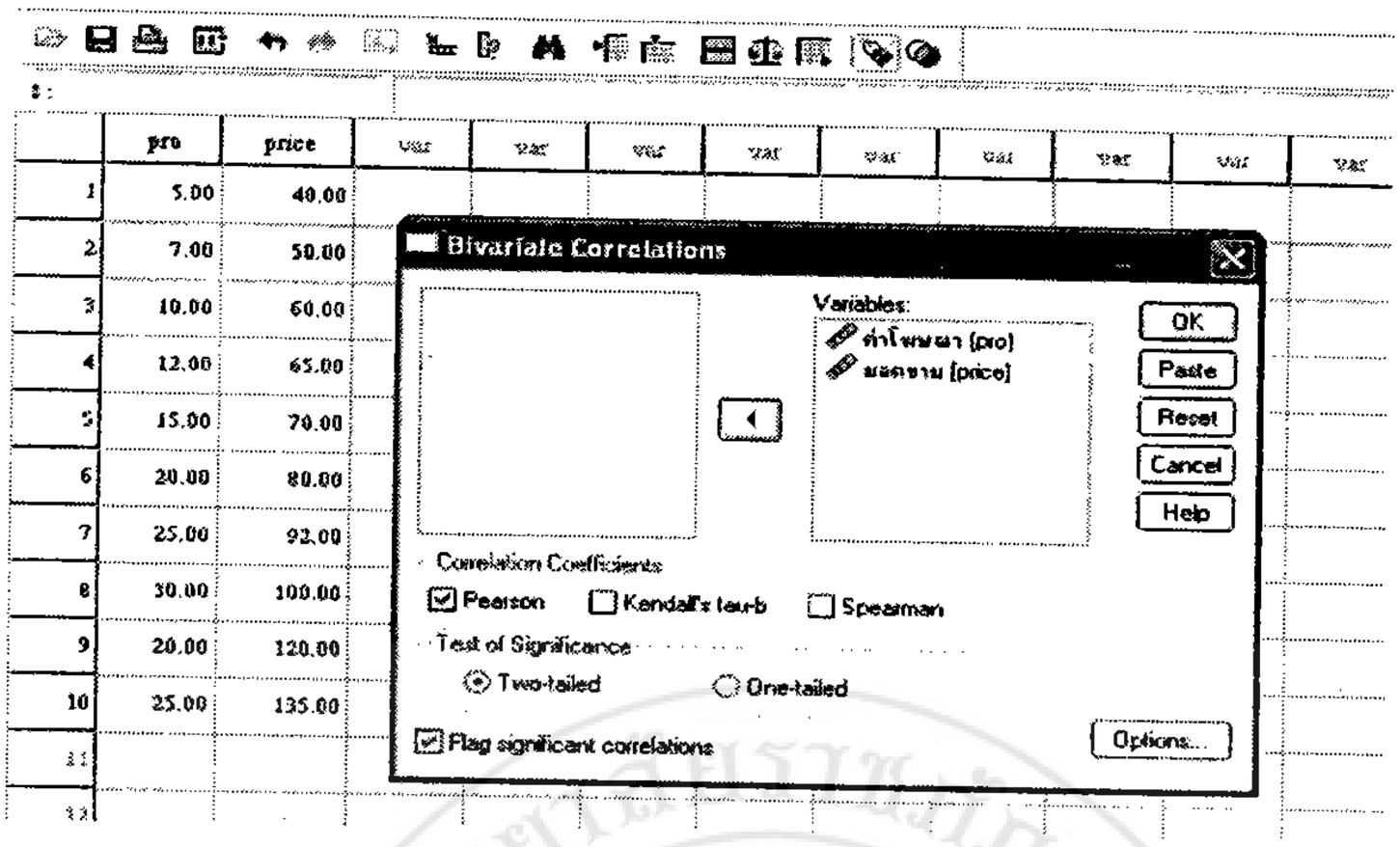
ปีที่	ค่าโฆษณา (แสนบาท)	ยอดขาย (แสนบาท)
1	5	40
2	7	50
3	10	60
4	12	65
5	15	70
6	20	80
7	25	92
8	30	100
9	20	120
10	25	135

วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ และสัมประสิทธิ์การตัดสินใจ

การหาค่าสัมประสิทธิ์สหสัมพันธ์และสัมประสิทธิ์การตัดสินใจของยอดขายสินค้าและค่าใช้จ่ายในการโฆษณาของ บริษัท วิซาร์ทมิวสิค จำกัด หาได้โดย การวิเคราะห์สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันและค่าสัมประสิทธิ์การตัดสินใจ เมื่อป้อนข้อมูลลง Data Editor แล้ว เลือกเมนู Analyze เลือกคำสั่ง Correlations และ Bivariate ดังแสดงในภาพที่ 10.1 เมื่อเลือกคำสั่งแล้วจะได้หน้าจอดังแสดงในภาพที่ 10.2

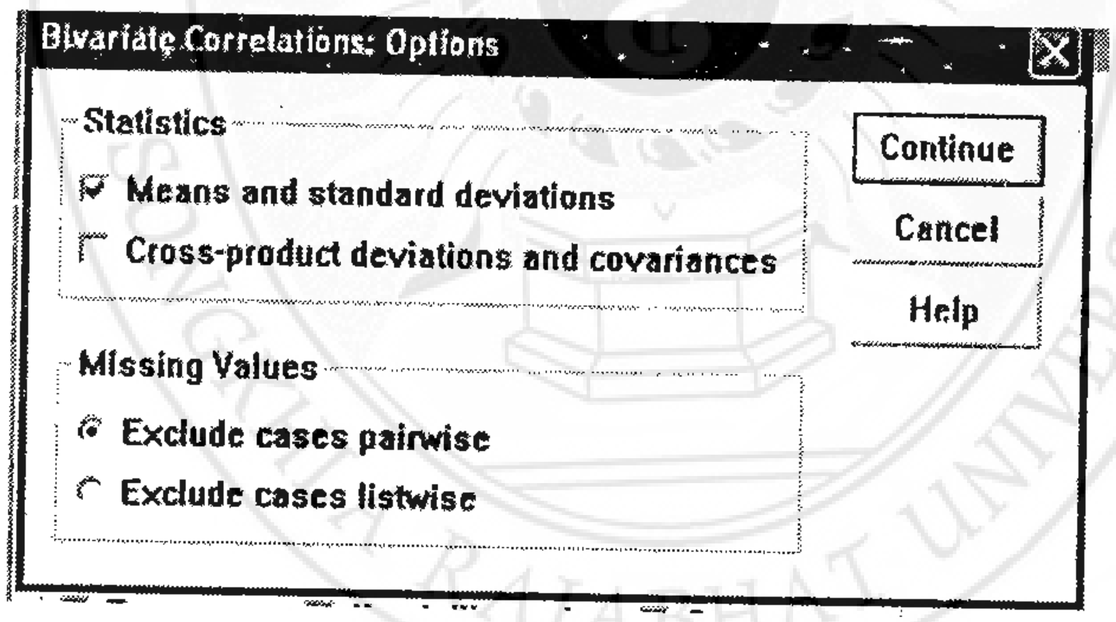


ภาพที่ 10.1 *แสดงการเลือกคำสั่ง Correlations และ Bivariate*



ภาพที่ 10.2 *แสดงการเลือกตัวแปรเพื่อทำการวิเคราะห์ความสัมพันธ์*

จากภาพที่ 10.2 แสดงการเลือกตัวแปรทั้งสองเพื่อทำการวิเคราะห์ค่าความสัมพันธ์โดยเลือกวิเคราะห์ค่าสัมประสิทธิสหสัมพันธ์แบบเพียร์สัน และเป็นการทดสอบแบบสองทาง และคลิก Options เพื่อเลือกคำนวณค่าสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน ดังแสดงในภาพที่ 10.3 จะได้ผลลัพธ์ดังแสดงในตารางที่ 10.2 และ 10.3



ภาพที่ 10.3 *แสดงการเลือกคำสั่งเพื่อคำนวณค่าสถิติพื้นฐาน*

ตารางที่ 10.2 *แสดงผลลัพธ์การคำนวณค่าสถิติพื้นฐาน*

Correlations

Descriptive Statistics

	Mean	Std. Deviation	N
ค่าโฆษณา	16.9000	8.41229	10
ยอดขาย	81.2000	30.51339	10

จากตารางที่ 10.2 ค่าโฆษณา มีค่าเฉลี่ยเท่ากับ 1,690,000 บาท ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 8.41229 ค่ายอดขายมีค่าเฉลี่ยเท่ากับ 8,120,000 บาท ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 30.51339 จำนวนตัวอย่าง เท่ากับ 10

ตารางที่ 10.3 *แสดงค่าสถิติที่ใช้ในการทดสอบสมมติฐาน และค่าสัมประสิทธิ์สหสัมพันธ์แบบ เพียร์สัน*

Correlations

		ค่าโฆษณา	ยอดขาย
ค่าโฆษณา	Pearson Correlation	1	.8364**
	Sig. (2-tailed)	.	.003
	N	10	10
ยอดขาย	Pearson Correlation	.8364**	1
	Sig. (2-tailed)	.003	.
	N	10	10

**. Correlation is significant at the 0.01 level

จากตารางที่ 10.3 พบว่า **ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน** (r) เท่ากับ **0.8364** หมายความว่า ค่าใช้จ่ายในการโฆษณามีความสัมพันธ์กับยอดขายสินค้าในทิศทางเดียวกันประมาณ 83.64%

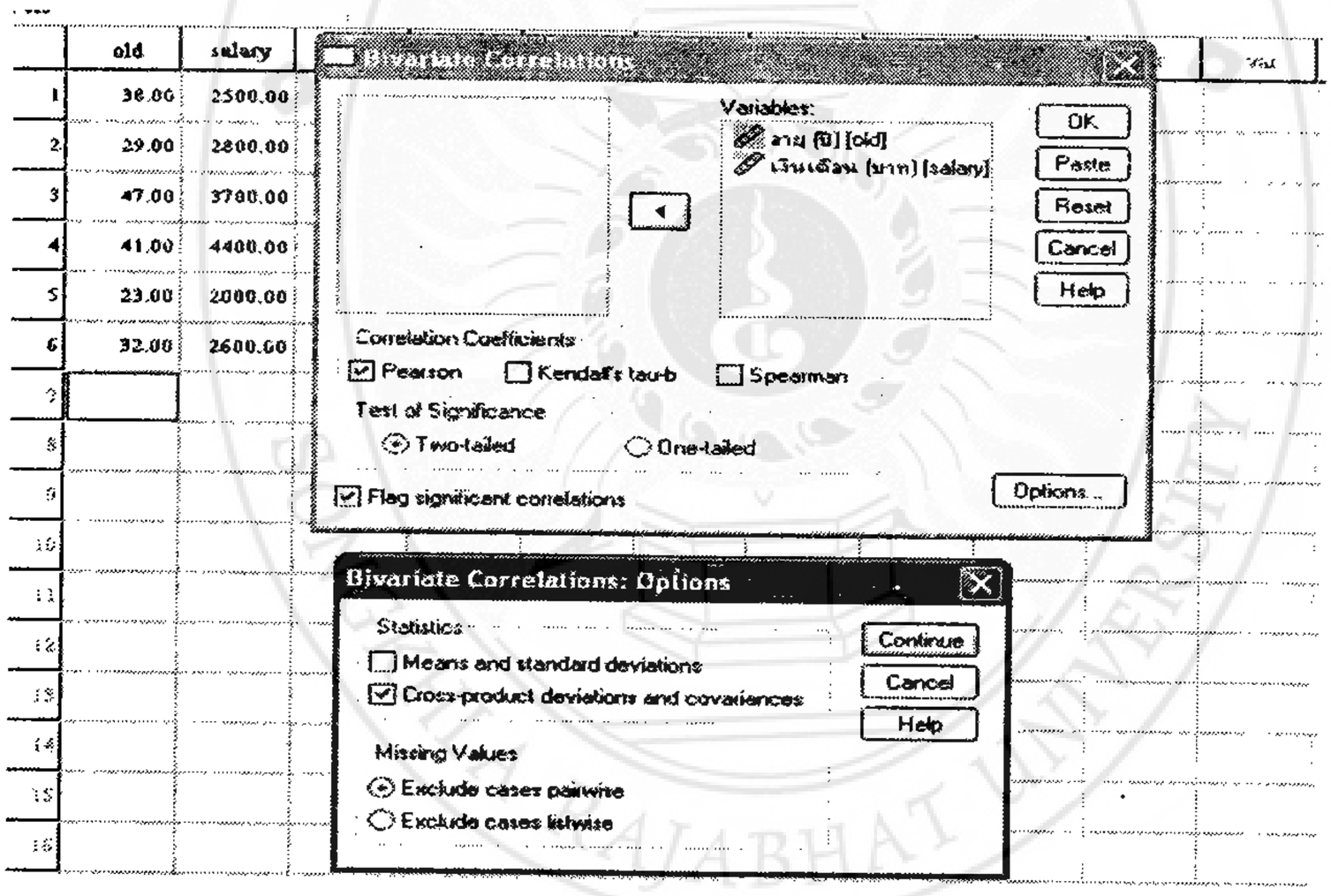
และ **เมื่อคำนวณค่าสัมประสิทธิ์การตัดสินใจ** (r^2) จะได้เท่ากับ **0.6996** หมายความว่า การเปลี่ยนแปลงของยอดขายสินค้ามีสาเหตุจากการเปลี่ยนแปลงค่าโฆษณาประมาณ 69.96% ส่วนที่เหลืออีก 30.04% เป็นผลมาจากสาเหตุอื่น

ตัวอย่างที่ 10.2 ในการวิเคราะห์ความสัมพันธ์ระหว่างเงินเดือนกับอายุของคนงานในโรงงานอุตสาหกรรมแห่งหนึ่ง ผู้จัดการของโรงงานจึงได้เลือกตัวอย่างคนงานมา 6 คน ปรากฏข้อมูลดังในตาราง จงคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ (r) และสัมประสิทธิ์การตัดสินใจ (r^2) ของเงินเดือนกับอายุของคนงาน

คนที่	อายุ (ปี)	เงินเดือน (บาท)
1	38	2,500
2	29	2,800
3	47	3,700
4	41	4,400
5	23	2,000
6	32	2,600

วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ และสัมประสิทธิ์การตัดสินใจ

การวิเคราะห์ค่าสัมประสิทธ์สหสัมพันธ์แบบเพียร์สัน สามารถทำได้โดยเมื่อป้อนข้อมูลลง Data Editor แล้ว เลือกเมนู Analyze เลือกคำสั่ง Correlations และ Bivariate จะได้หน้าจอดังแสดงในภาพที่ 10.4



***ภาพที่ 10.4** แสดงการเลือกตัวแปรเพื่อวิเคราะห์ความสัมพันธ์*

จากภาพที่ 10.4 แสดงการเลือกตัวแปรทั้งสองเพื่อทำการวิเคราะห์ค่าความสัมพันธ์โดยเลือกวิเคราะห์ค่าสัมประสิทธิสหสัมพันธ์แบบเพียร์สัน และเป็นการทดสอบแบบสองทาง และคลิก Options เพื่อเลือกคำนวณค่าสถิติ จะได้ผลลัพธ์ดังแสดงในตารางที่ 10.4

ตารางที่ 10.4 *แสดงค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันของเงินเดือนและอายุ*

Correlations

		เงินเดือน (บาท)	อายุ (ปี)
เงินเดือน (บาท)	Pearson Correlation	1	.774
	Sig. (2-tailed)	.	.071
	Sum of Squares and Cross-products	3900000.000	29700.00
	Covariance	780000.000	5940.000
	N	6	6
อายุ (ปี)	Pearson Correlation	.774	1
	Sig. (2-tailed)	.071	.
	Sum of Squares and Cross-products	29700.000	378.000
	Covariance	5940.000	75.600
	N	6	6

จากตารางที่ 10.6 พบว่า **ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน** (r) เท่ากับ **0.774** หมายความว่า อายุมีความสัมพันธ์กับเงินเดือนของคนในโรงงานอุตสาหกรรมในทิศทางเดียวกันประมาณ 77.4%

การคำนวณ **ค่าสัมประสิทธิการตัดสินใจ** (r^2) เท่ากับ **0.5991** หมายความว่า การเปลี่ยนแปลงเงินเดือนมีสาเหตุจากอายุประมาณ 59.91% ส่วนที่เหลือ 40.09% เป็นผลมาจากสาเหตุอื่น ๆ นั่นคือ อายุมีอิทธิพลต่อเงินเดือนประมาณ 59.91%

3. การทดสอบสมมติฐานค่าสัมประสิทธิ์สหสัมพันธ์

เนื่องจากค่า r ที่ได้ เป็นค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้มาจากการสุ่มตัวอย่างและนำไปอ้างอิงค่าสัมประสิทธิ์ของประชากร หรือ “ρ” ดังนั้น เพื่อความมั่นใจว่า ค่า r นั้นสามารถนำไปอ้างอิงค่า ρ ได้จริง จึงต้องทำการทดสอบมาตรฐาน ซึ่งแบ่งเป็น 2 วิธี คือ การทดสอบแบบ 2 ทาง และการทดสอบแบบทางเดียว ดังนี้

3.1 การทดสอบมาตรฐานแบบ 2 ทาง กรณีนี้เป็นการทดสอบว่าตัวแปร X กับตัวแปร Y มีความสัมพันธ์กันหรือไม่ การตั้งสมมติฐานกำหนดได้ดังนี้

$H_0 : \rho = 0$ (X กับ Y ไม่มีความสัมพันธ์กัน)

$H_1 : \rho \neq 0$ (X กับ Y มีความสัมพันธ์กัน)

3.2 การทดสอบมาตรฐานแบบทางเดียว กรณีนี้เป็นการทดสอบว่า ตัวแปร X กับ ตัวแปร Y มีความสัมพันธ์ และทิศทางตรงกับที่ได้จากค่า r หรือไม่ โดยที่การตั้งสมมติฐานแบบทางเดียวนั้น จะขึ้นอยู่กับค่า r ที่ได้ว่า มีค่าเป็น บวก หรือ ลบ

3.2.1 กรณีที่ค่า $r > 0$ ซึ่งกำหนดสมมติฐานได้ ดังนี้

$H_0 : \rho \leq 0$ (X กับ Y ไม่มีความสัมพันธ์ในทิศทางเดียวกัน)

$H_1 : \rho > 0$ (X กับ Y มีความสัมพันธ์ในทิศทางเดียวกัน)

3.2.2 กรณีที่ค่า $r < 0$ ซึ่งกำหนดสมมติฐานได้ ดังนี้

$H_0 : \rho \geq 0$ (X กับ Y ไม่มีความสัมพันธ์ในทิศทางตรงกันข้าม)

$H_1 : \rho < 0$ (X กับ Y มีความสัมพันธ์ในทิศทางตรงกันข้าม)

ตัวอย่างที่ 10.3 จากข้อมูลในตัวอย่างที่ 10.1 จงทดสอบสมมติฐานว่า ค่าใช้จ่ายในการโฆษณา กับยอดขายสินค้า มีความสัมพันธ์กันหรือไม่ ที่ระดับนัยสำคัญ 0.01

ขั้นตอนการทดสอบสมมติฐาน มีดังนี้

1. กำหนดสมมติฐานทางสถิติ

$H_0 : \rho = 0$ (ค่าใช้จ่ายในการโฆษณาไม่มีความสัมพันธ์กับยอดขาย)

$H_1 : \rho \neq 0$ (ค่าใช้จ่ายในการโฆษณามีความสัมพันธ์กับยอดขาย)

2. กำหนดระดับนัยสำคัญ .01

3. พิจารณา ค่า Sig. (2-tailed) เท่ากับ .003 ซึ่งน้อยกว่าค่านัยสำคัญที่กำหนด .01 นั่นคือ **ปฏิเสธสมมติฐานหลัก ยอมรับสมมติฐานทางเลือก**

4. สรุป **ค่าใช้จ่ายในการโฆษณามีความสัมพันธ์กับยอดขายที่ระดับนัยสำคัญ 0.01**

ตัวอย่างที่ 10.4 จากตัวอย่างที่ 10.2 จงทดสอบสมมติฐานว่า อายุ กับเงินเดือน มีความสัมพันธ์ในทิศทางเดียวกันหรือไม่ ที่ระดับความเชื่อมั่น 99%

ขั้นตอนการทดสอบสมมติฐาน มีดังนี้

1. เนื่องจากค่า r เท่ากับ 0.774 ซึ่งมากกว่า 0 ดังนั้นจึงกำหนดสมมติฐานได้ดังนี้

$H_0 : \rho \leq 0$ (อายุ กับ เงินเดือน ไม่มีความสัมพันธ์ในทิศทางเดียวกัน)

$H_1 : \rho > 0$ (อายุ กับ เงินเดือน มีความสัมพันธ์ในทิศทางเดียวกัน)

2. กำหนดระดับนัยสำคัญ .01

3. พิจารณา ค่า Sig. (2-tailed) เท่ากับ .071 ซึ่งมากกว่าค่านัยสำคัญที่กำหนด .01 นั่นคือ **ยอมรับสมมติฐานหลัก**

4. สรุป **อายุ กับ เงินเดือนไม่มีความสัมพันธ์ในทิศทางเดียวกันที่ระดับนัยสำคัญ 0.01**

การวิเคราะห์สหสัมพันธ์เชิงซ้อน

การวิเคราะห์สหสัมพันธ์เชิงซ้อน (multiple correlation analysis) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระตั้งแต่ 2 ตัวแปรขึ้นไป กับ ตัวแปรตาม 1 ตัวแปร

โดยที่ $r_{y.x1x2...xk}$ คือ ค่าสัมประสิทธิ์สหสัมพันธ์เชิงซ้อน

เมื่อ y แทน ตัวแปรตาม

x_i แทน ตัวแปรอิสระ เมื่อ $i = 1, 2, \ldots, k$

k แทน จำนวนตัวแปรอิสระ และ $k \geq 2$

คุณสมบัติของค่าสัมประสิทธิ์สหสัมพันธ์เชิงซ้อน $0 \leq r_{y.x1.x2 \ldots xk} \leq 1$ มีดังนี้

1. ถ้าค่า $r_{y.x1.x2 \ldots xk}$ มีค่าเข้าใกล้ 0 แสดงว่า ตัวแปรอิสระกับตัวแปรตามมีความสัมพันธ์กันน้อย

2. ถ้าค่า $r_{y.x1.x2 \ldots xk}$ มีค่าเข้าใกล้ 1 แสดงว่า ตัวแปรอิสระกับตัวแปรตามมีความสัมพันธ์กันมาก

ตัวอย่างที่ 10.5 จงหาค่าสัมประสิทธิ์สหสัมพันธ์ที่แสดงความสัมพันธ์ระหว่างยอดขายของพนักงานกับประสบการณ์ทำงานและระยะเวลาการฝึกอบรมของพนักงานที่ได้จากการเก็บข้อมูลพนักงาน 12 คน ปรากฏดังในตาราง

คนที่	ยอดขาย (พันบาท)	ประสบการณ์ (ปี)	ระยะเวลาการฝึกอบรม (สัปดาห์)
1	650	30	24
2	90	2	3
3	140	5	5
4	200	5	10
5	350	15	18
6	150	4	10
7	460	24	12
8	540	30	0
9	290	10	22
10	310	20	1
11	360	20	5
12	240	15	10

วิธีหาค่าสหสัมพันธ์เชิงซ้อน หาได้ดังนี้

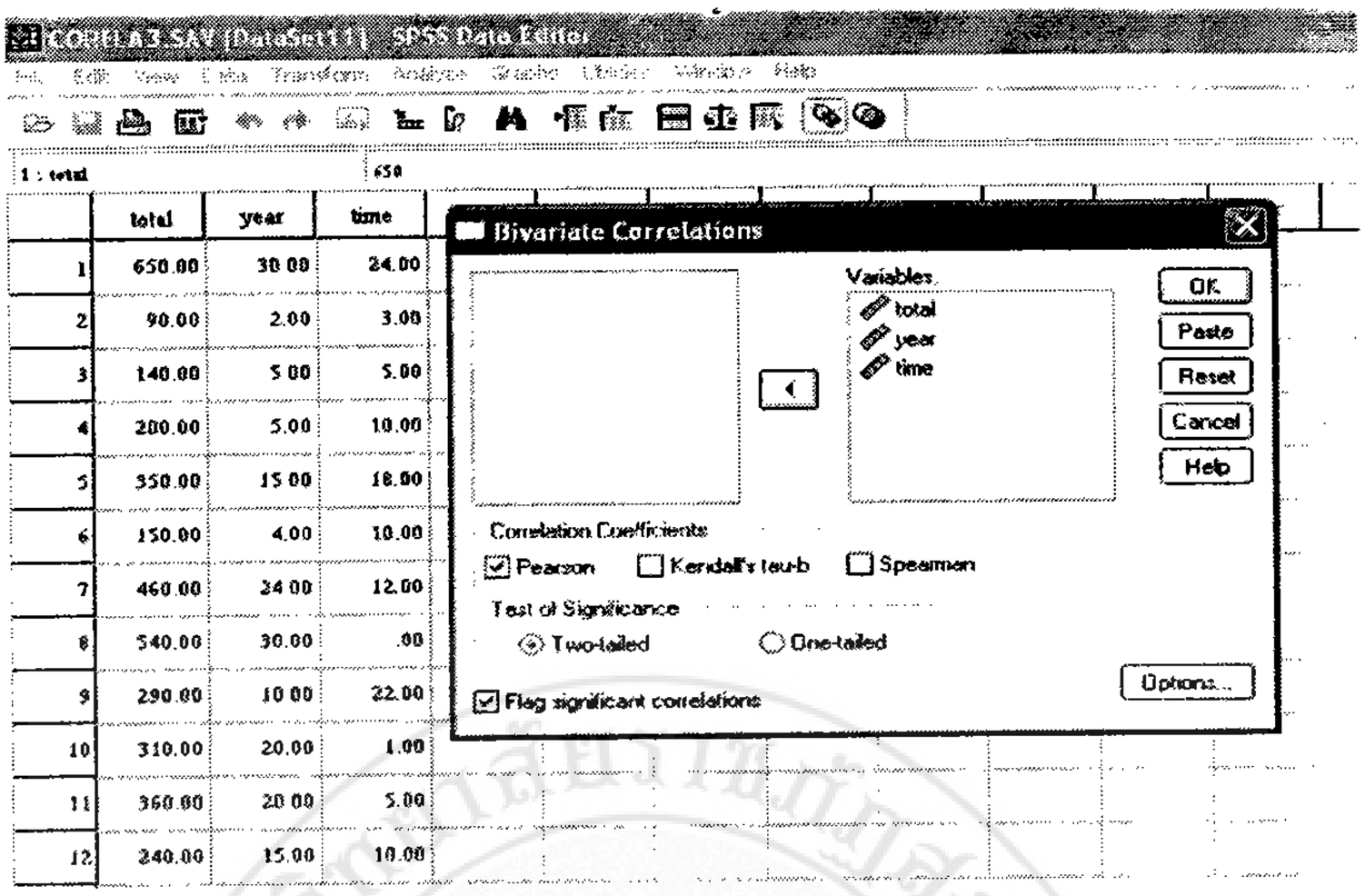
จากข้อมูลในตัวอย่างที่ 10.5

กำหนดให้ Y แทน ยอดขายของพนักงาน

X_1 แทน ประสบการณ์ของพนักงาน

X_2 แทน ระยะเวลาการฝึกอบรมของพนักงาน

เมื่อป้อนข้อมูลลงใน Data Editor แล้ว เลือกคำสั่ง Correlations และ Bivariable จากเมนู Analyze จะได้หน้าจอดังแสดงในภาพที่ 10.5



ภาพที่ 10.5 *แสดงการเลือกตัวแปรเพื่อวิเคราะห์ค่าสหสัมพันธ์เชิงซ้อน*

จากภาพที่ 10.5 แสดงการเลือกตัวแปรทั้ง 3 ตัวเพื่อทำการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์เชิงซ้อน จะได้ผลลัพธ์ดังแสดงในตารางที่ 10.5

ตารางที่ 10.5 *แสดงค่าสัมประสิทธิ์สหสัมพันธ์เชิงซ้อนของตัวแปร total year และ time*

Correlations

		TOTAL	YEAR	TIME
TOTAL	Pearson Correlation	1	.9508**	.338
	Sig. (2-tailed)	.	.000	.282
	N	12	12	12
YEAR	Pearson Correlation	.9508**	1	.102
	Sig. (2-tailed)	.000	.	.751
	N	12	12	12
TIME	Pearson Correlation	.338	.102	1
	Sig. (2-tailed)	.282	.751	.
	N	12	12	12

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

จากตารางที่ 10.5 พบว่า **ค่าสหสัมพันธ์เชิงซ้อน** ($r_{y.x1x2}$) เท่ากับ **0.9508** หมายความว่า **ยอดขายของพนักงานมีความสัมพันธ์กับประสบการณ์ในการทำงานและระยะเวลาการฝึกอบรมของพนักงานประมาณ 95.08%**

สหสัมพันธ์เชิงส่วน

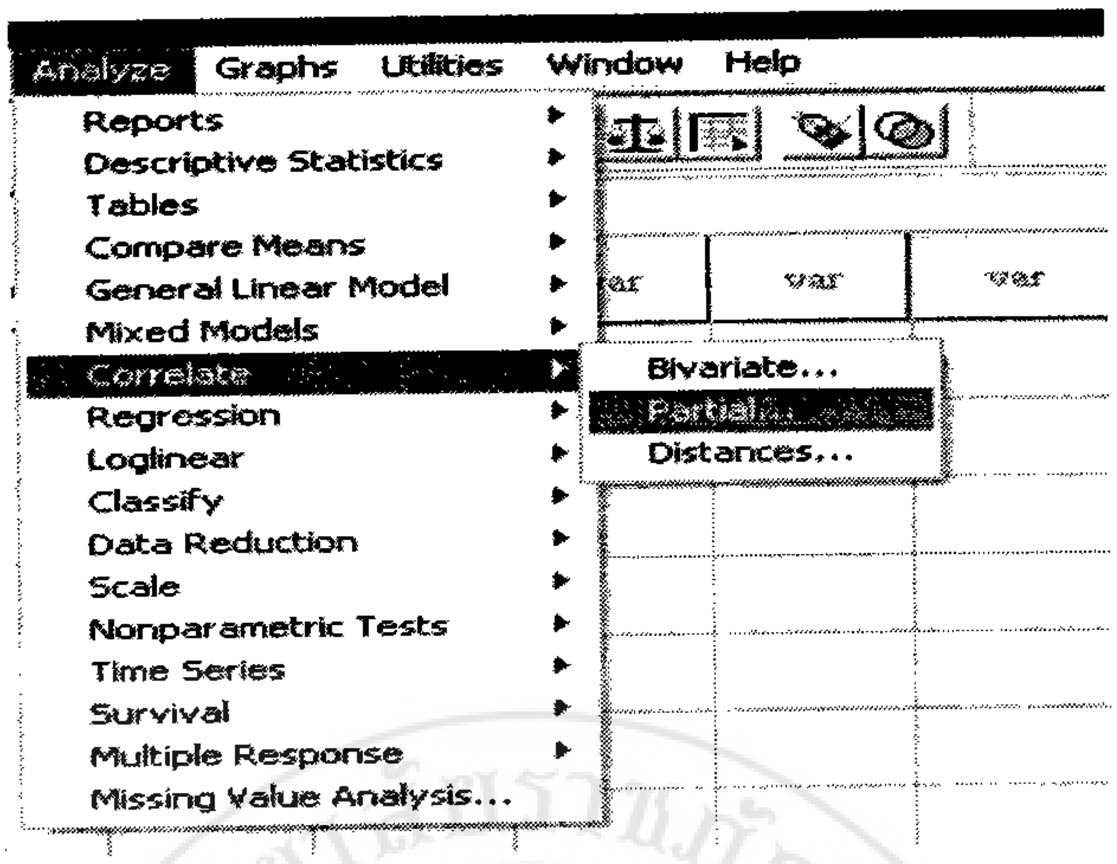
สหสัมพันธ์เชิงส่วน (partial correlation) เป็นการหาความสัมพันธ์ระหว่างตัวแปรตาม กับ ตัวแปรอิสระตัวใดตัวหนึ่ง โดยกำหนดให้ตัวแปรอิสระที่เหลือนั้นคงที่ ตัวอย่างเช่น ถ้ามีตัวแปรตาม 1 ตัว คือ Y และตัวแปรอิสระ 2 ตัว คือ X_1 กับ X_2 ดังนั้น ถ้าต้องการหาความสัมพันธ์ระหว่าง Y กับ X_1 โดยกำหนดให้ X_2 เป็นปัจจัยคงที่ ค่าที่ได้คือ $r_{yx1.x2}$ เรียกว่าค่าสัมประสิทธิ์สหสัมพันธ์เชิงส่วน ซึ่งเป็นค่าสัมประสิทธิ์สหสัมพันธ์เชิงส่วนที่แสดงความสัมพันธ์ระหว่าง Y กับ X_1 เมื่อ กำหนดให้ X_2 เป็นปัจจัยคงที่ โดย $-1 \leq r_{yx1.x2} \leq 1$ และ $-1 \leq r_{yx2.x1} \leq 1$

ตัวอย่างที่ 10.6 ในการศึกษาถึงความสัมพันธ์ระหว่างราคาตึกแถว กับ จำนวนห้องนอนและจำนวนห้องน้ำ ได้เลือกตึกแถวโดยการสุ่มมาจำนวน 8 ห้อง ผลปรากฏดังในตาราง เมื่อกำหนดให้ y แทน ราคาตึกแถว x_1 แทน จำนวนห้องนอน และx_2 แทน จำนวนห้องน้ำจงหาค่า $r_{y.x1x2}$ และ $r_{yx1.x2}$ พร้อมทั้งอธิบายค่าที่ได้

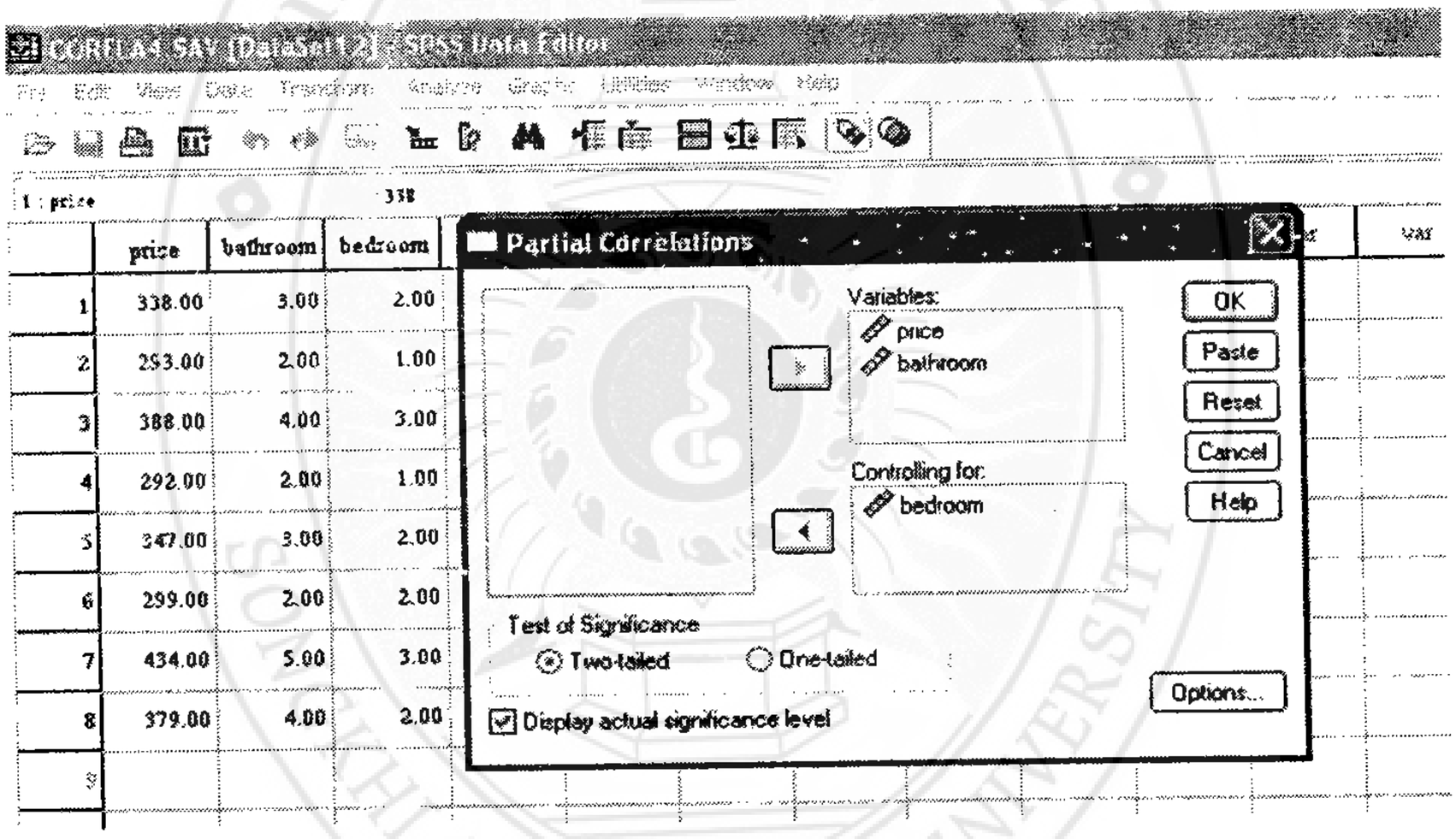
ราคาตึกแถว (พันบาท)	จำนวนห้องนอน	จำนวนห้องน้ำ
338	3	2
293	2	1
388	4	3
292	2	1
347	3	2
299	2	2
434	5	3
379	4	2

วิธีวิเคราะห์ค่าสหสัมพันธ์เชิงส่วน

ในโปรแกรม SPSS สามารถทำได้โดยเมื่อป้อนข้อมูลแล้วเลือกคำสั่ง Correlations และ Partial ดังภาพที่10.6 เมื่อเลือกคำสั่งนี้แล้ว จะได้หน้าจอดังแสดงในภาพที่ 10.7



ภาพที่10.6 *แสดงการเลือกคำสั่งเพื่อวิเคราะห์ค่าสหสัมพันธ์เชิงส่วน*



ภาพที่ 10.7 *แสดงการเลือกตัวแปรราคาตึกแถวและจำนวนห้องน้ำ เพื่อวิเคราะห์ค่าสหสัมพันธ์เชิงส่วน*

จากภาพที่ 10.7 เป็นการเลือกตัวแปรราคาตึกแถว (price) และจำนวนห้องน้ำ (bathroom) โดยควบคุมตัวแปร จำนวนห้องนอน (bedroom) ให้คงที่ จะได้ผลลัพธ์ดังแสดงใน ภาพที่ 10.8

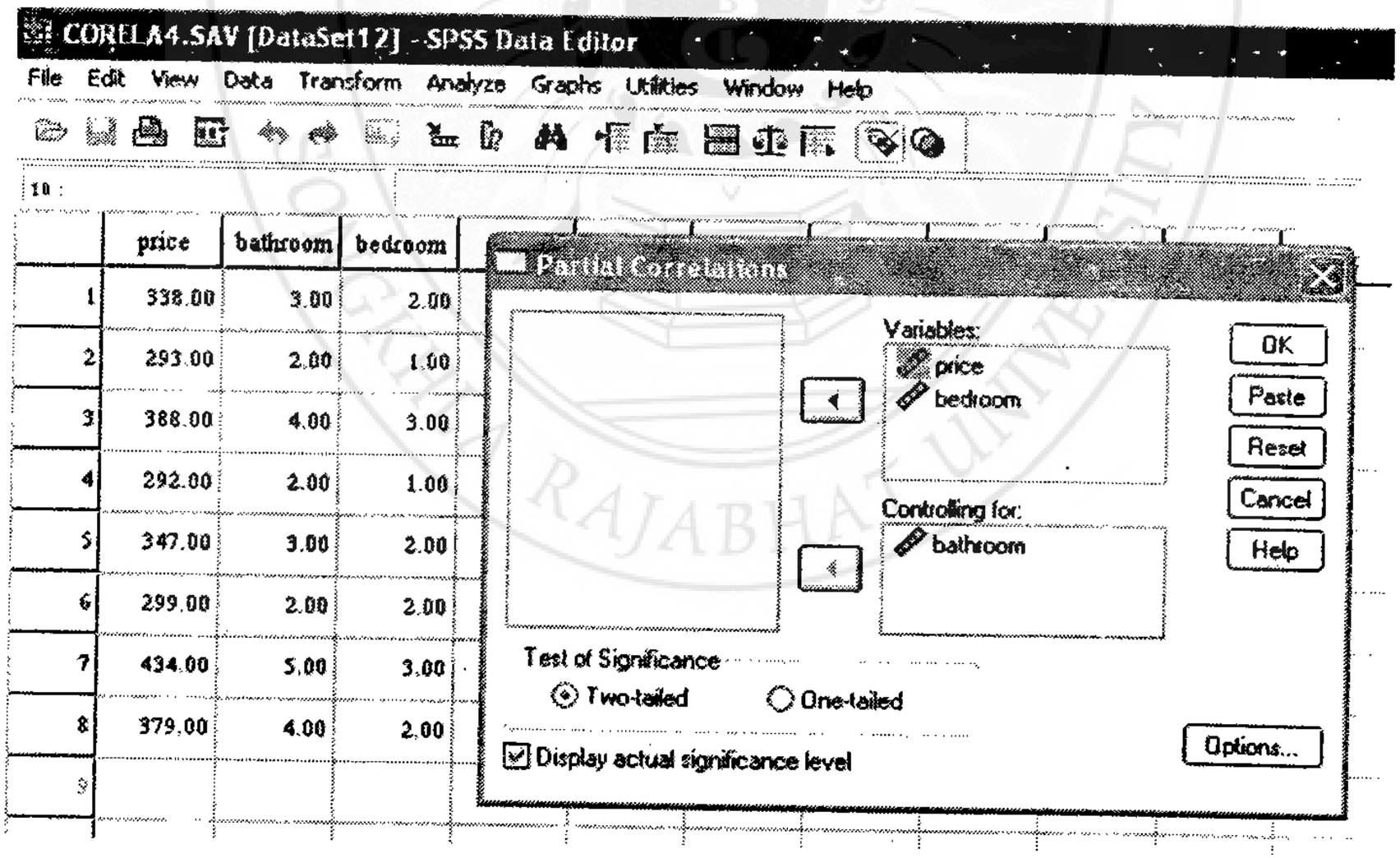
Correlations

Control Variables			price	bathroom
bedroom	price	Correlation	1.0000	.9924
		Significance (2-tailed)	.	.0000
		df	.0000	5.0000
	bathroom	Correlation	.9924	1.0000
		Significance (2-tailed)	.0000	.
		df	5.0000	.0000

ภาพที่ 10.8 *แสดงค่าสหสัมพันธ์เชิงส่วนเมื่อกำหนดให้จำนวนห้องนอนคงที่*

จากภาพที่ 10.8 จะได้ค่า $r_{y.x1x2}$ เท่ากับ **0.9924** หมายความว่า **ราคาตึกแถวมีความสัมพันธ์กับจำนวนห้องน้ำ และจำนวนห้องนอน ประมาณ 99.24%** เมื่อกำหนดให้จำนวน ห้องนอนคงที่

วิธีการวิเคราะห์ค่าสหสัมพันธ์เชิงส่วนเมื่อกำหนดให้จำนวนห้องน้ำคงที่ สามารถทำได้โดยเลือกตัวแปรดังแสดงในภาพที่ 10.9 จะได้ผลลัพธ์ดังแสดงในภาพที่ 10.10



ภาพที่ 10.9 *แสดงการกำหนดให้ตัวแปรในการวิเคราะห์ค่าสหสัมพันธ์เชิงส่วน*

Correlations

Control Variables			price	bedroom
bathroom	price	Correlation	1.0000	.7057
		Significance (2-tailed)	.	.0764
		df	.0000	5.0000
	bedroom	Correlation	.7057	1.0000
		Significance (2-tailed)	.0764	.
		df	5.0000	.0000

ภาพที่ 10.10 *แสดงค่าสหสัมพันธ์เชิงส่วนเมื่อกำหนดให้ตัวแปรจำนวนห้องน้ำคงที่*

จากภาพที่ 10.10 จะได้ **ค่า** $r_{yx1.x2}$ เท่ากับ **0.7057** หมายความว่า **ราคาติดแถวกับจำนวนห้องนอนมีความสัมพันธ์กัน** ในทิศทางเดียวกันประมาณ 70.57% **เมื่อกำหนดให้จำนวนห้องน้ำคงที่**

การวิเคราะห์การถดถอย

การวิเคราะห์การถดถอยเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรตาม 1 ตัวแปร กับ ตัวแปรอิสระตั้งแต่ 1 ตัวแปรขึ้นไป โดยมีวัตถุประสงค์เพื่อประมาณการค่าของตัวแปรตาม เมื่อได้ทราบค่าของตัวแปรอิสระแล้ว สำหรับค่าสถิติที่ใช้ในการวิเคราะห์ความถดถอย โดยวัตถุประสงค์ที่สำคัญของการสร้างสมการเส้นถดถอยขึ้นมา ก็เพื่อประมาณหรือพยากรณ์ตัวแปรที่ต้องการศึกษา ดังนั้นเพื่อความมั่นใจว่าสมการเส้นถดถอยที่สร้างขึ้นมานั้น สามารถนำไปประมาณหรือพยากรณ์ตัวแปรที่เราต้องการศึกษาได้จริง จึงต้องมีการทดสอบโดยอาศัยค่าสถิติต่าง ๆ ดังต่อไปนี้

1. สัมประสิทธิ์การตัดสินใจ (coeficient of determination : R^2) คือ ค่าที่อธิบายว่าความผันแปรทั้งหมดของตัวแปรตามมีสาเหตุมาจากตัวแปรอิสระกี่ % โดยที่ $0 \leq R^2 \leq 1$ และ

e_i = ค่าความผิดพลาด นั่นคือ $e_i = y - \hat{y}$

ดังนั้น $\sum e_i^2$ คือ ผลรวมกำลังสองของค่าความผิดพลาด (error sum of square : ESS) หรือ ความแปรปรวนที่ไม่อาจอธิบายได้ (unexplained variation)

2. สัมประสิทธิ์การตัดสินใจที่ปรับค่าแล้ว (adjusted coefficient of determination : $\overline{R^2}$) ในกรณีที่ข้อมูลที่นำมาศึกษามีจำนวนน้อย อาจทำให้ค่า R^2 ที่คำนวณได้มีค่าสูงเกินไป ดังนั้นจึงนำค่า R^2 มาปรับเสียใหม่ เรียกว่า ค่าสัมประสิทธิ์การตัดสินใจที่ปรับค่าแล้ว หรือ $\overline{R^2}$

3. การทดสอบสมมติฐานสมการถดถอยเชิงเส้นอย่างง่าย จะมีการทดสอบสมมติฐานที่สำคัญอยู่ 2 ส่วน คือ การทดสอบสมมติฐานค่าคงที่ กับค่าสัมประสิทธิ์ของตัวแปรอิสระ ซึ่งมีขั้นตอนดังนี้

การทดสอบสมมติฐานค่าคงที่ ดังนี้

กำหนดสมมติฐาน

$H_0 : \beta_0 = 0$ (ค่าคงที่ = 0)

$H_1 : \beta_0 \neq 0$ (ค่าคงที่ ≠ 0)

การทดสอบสมมติฐานค่าสัมประสิทธิ์ของตัวแปรอิสระ ดังนี้

กำหนดสมมติฐาน

กรณีที่ค่า $b > 0$

$H_0 : \beta_1 \leq 0$ (x กับ y ไม่มีความสัมพันธ์ในทิศทางเดียวกัน)

$H_1 : \beta_1 > 0$ (x กับ y มีความสัมพันธ์ในทิศทางเดียวกัน)

กรณีที่ค่า $b < 0$

$H_0 : \beta_1 \geq 0$ (x กับ y ไม่มีความสัมพันธ์ในทิศทางตรงกันข้ามกัน)

$H_1 : \beta_1 < 0$ (x กับ y มีความสัมพันธ์ในทิศทางตรงข้ามกัน)

การวิเคราะห์การถดถอยแบ่งเป็น 2 ประเภท คือ การวิเคราะห์การถดถอยอย่างง่าย และการวิเคราะห์การถดถอยเชิงซ้อน ดังรายละเอียดต่อไปนี้

1. การวิเคราะห์การถดถอยอย่างง่าย

การวิเคราะห์การถดถอยอย่างง่าย (simple regression analysis) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรตาม (y) 1 ตัวแปร กับ ตัวแปรอิสระ (x) 1 ตัวแปร โดยที่ในช่วงแรกจะศึกษาความสัมพันธ์ระหว่าง Y กับ X ในรูปเชิงเส้นก่อน ดังนั้นสมการเส้นถดถอย คือ

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

เมื่อ β_0 แทน ค่าคงที่ หรือ จุดตัดแกนตั้ง

β_1 แทน ค่าความชัน

ε แทน ส่วนคลาดเคลื่อน

ปกติเราไม่สามารถเก็บข้อมูลจากประชากรทั้งหมดได้ จึงต้องใช้การสุ่มตัวอย่างมาใช้ในการสร้างสมการถดถอย ซึ่งเป็นสมการถดถอยที่ได้จากการประมาณ โดยใช้สูตร

สูตร $\hat{y}_i = a + bx_i$

เมื่อ $\hat{y}_i$ แทน ค่าประมาณการของ y

a แทน ค่าประมาณการของ β_0

b แทน ค่าประมาณการของ β_1

e_i แทน ค่าความผิดพลาด (error)

เมื่อ $e_i = y_i - \hat{y}_i$

ดังนั้น การประมาณสมการถดถอยที่ดี ควรเป็นสมการถดถอยที่เกิดค่าความผิดพลาดน้อยที่สุด ซึ่งวิธีที่นิยมใช้ คือ วิธีกำลังสองน้อยที่สุด (ordinary least square หรือ OLS)

ตัวอย่างที่ 10.7 จงใช้ข้อมูลในตาราง เพื่อประมาณสมการการบริโภค

ปี	รายได้ (ล้านบาท)	การบริโภค (ล้านบาท)
1990	20	7
1991	30	8
1992	33	9
1993	40	11
1994	15	5
1995	13	4
1996	26	8
1997	38	10
1998	35	9
1999	43	10

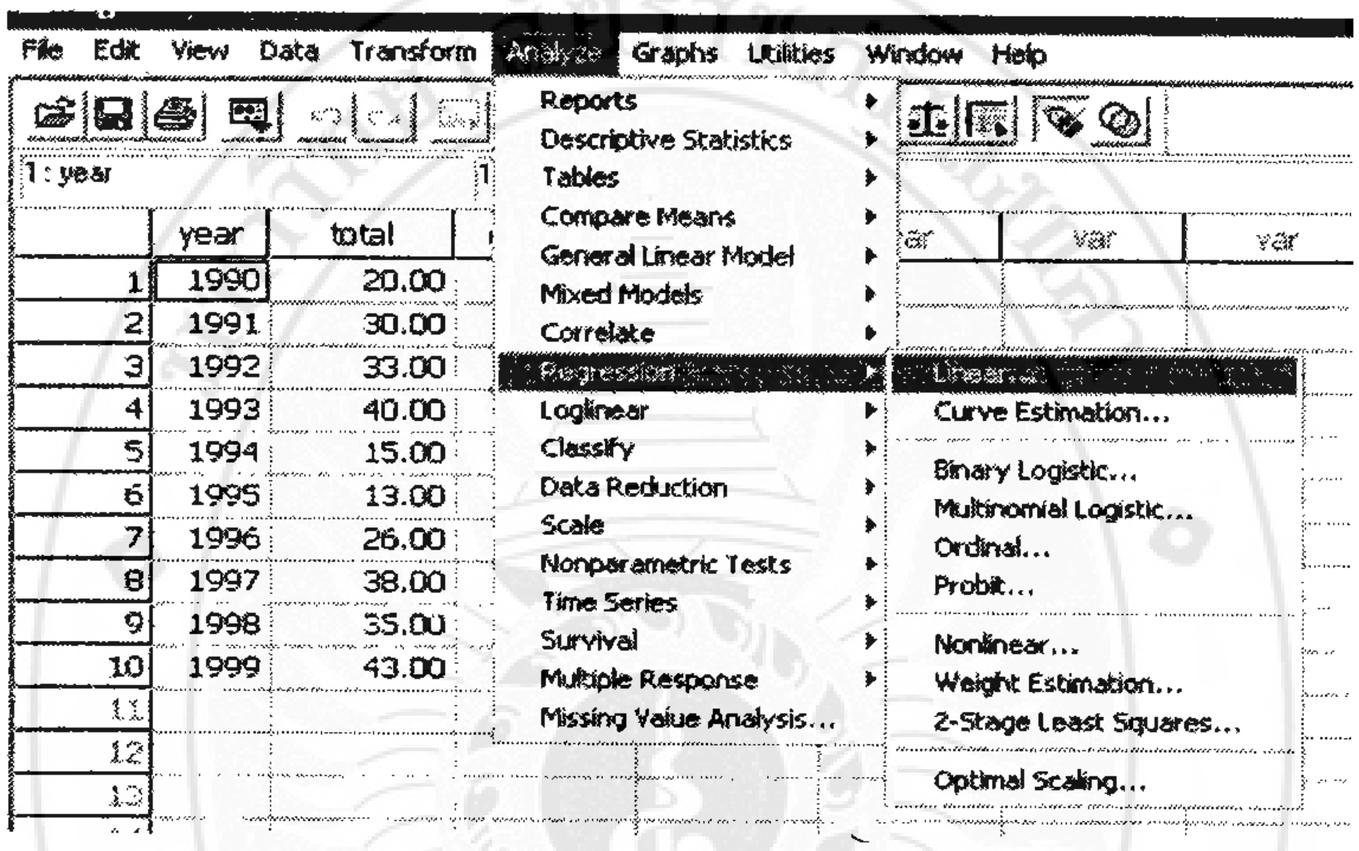
การหาค่าประมาณสมการการบริโภค

ใช้สูตร $\hat{y}_i = a + bx_i$

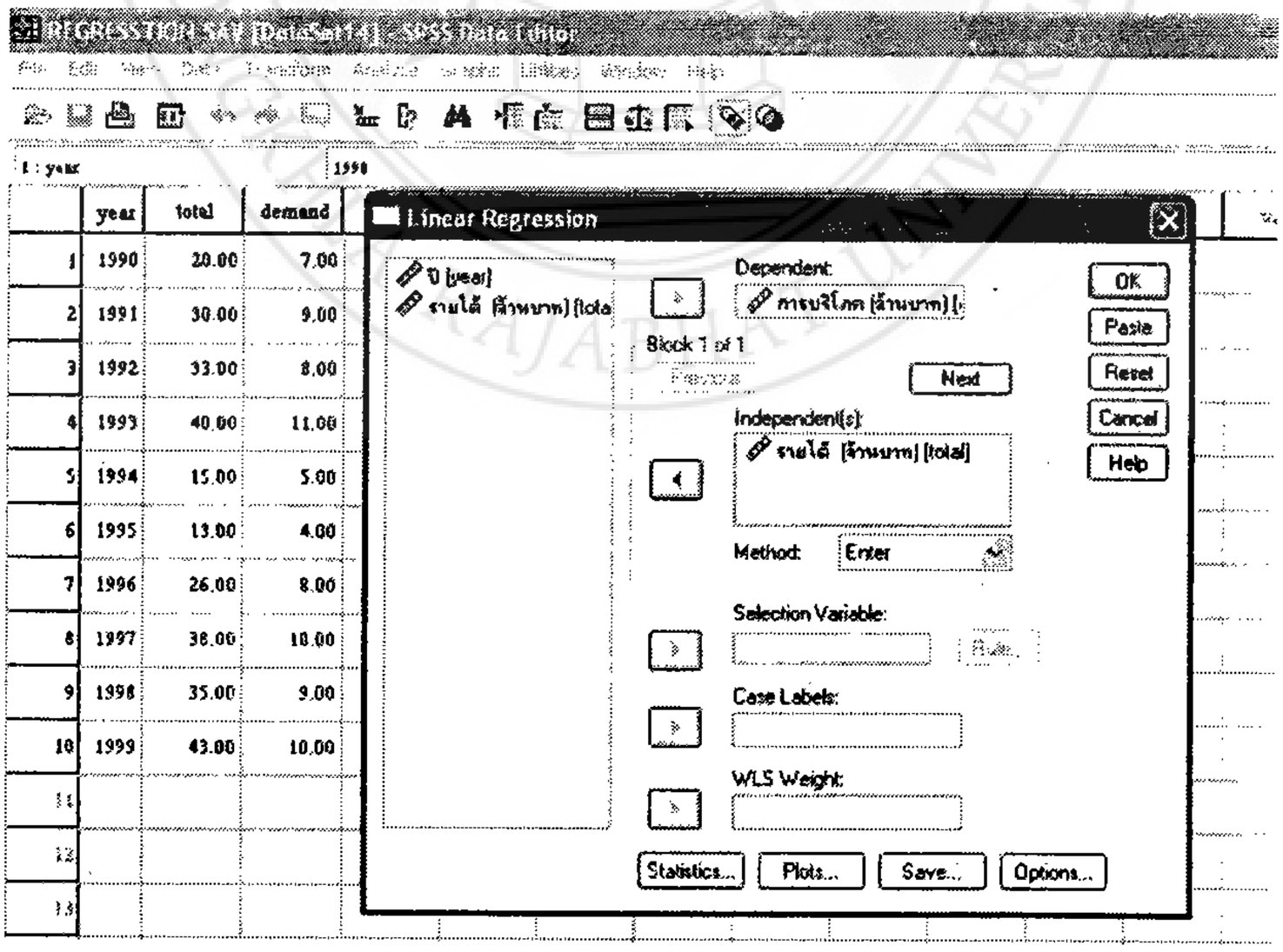
กำหนดให้ x แทน รายได้

y แทน การบริโภค

การวิเคราะห์การถดถอยอย่างง่ายด้วยโปรแกรม SPSS สามารถทำได้โดยเมื่อป้อนข้อมูลแล้วให้เลือกคำสั่ง Regression และ Linear จากเมนู Analyze ดังแสดงในภาพที่ 10.11 เมื่อเลือกคำสั่งนี้แล้วจะได้หน้าจอดังแสดงในภาพที่ 10.12



ภาพที่ 10.11 *แสดงการเลือกคำสั่ง Regression และ Linear*



ภาพที่ 10.12 *แสดงการกำหนดตัวแปรเพื่อวิเคราะห์การถดถอยอย่างง่าย*

จากภาพที่ 10.12 เป็นการกำหนดให้ตัวแปรการบริโภคเป็นตัวแปรตาม ตัวแปรรายได้ เป็นตัวแปรอิสระ จะได้ผลลัพธ์ดังแสดงในตารางที่ 10.6 - 10.9

ตารางที่ 10.6 *แสดงค่าสถิติความสัมพันธ์ของรายได้และการบริโภค*

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.951[a]	.904	.892	.73304

a. Predictors: (Constant), รายได้ (ล้านบาท)

จากตารางที่ 10.6 แสดงว่าตัวแปรทั้งสองตัว คือ การบริโภคและรายได้นั้น มีความสัมพันธ์กันในทิศทางเดียวกันเนื่องจากค่า r เท่ากับ .951

ตารางที่ 10.7 *แสดงค่าสถิติทดสอบความสัมพันธ์ระหว่างการบริโภคและรายได้*

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	40.601	1	40.601	75.559	.000[a]
	Residual	4.299	8	.537		
	Total	44.900	9			

a. Predictors: (Constant), รายได้ (ล้านบาท)

b. Dependent Variable: การบริโภค (ล้านบาท)

จากตารางที่ 10.7 พบว่า ข้อมูลของการบริโภคและรายได้ อยู่บนเส้นตรงเดียวกันและข้อมูลส่วนที่เบี่ยงเบนออกจากเส้นตรงมีความแตกต่างกันถึง 75.559 เท่า ถ้ากำหนดความเชื่อมั่นที่ 99% พบว่า รายได้สามารถนำไปใช้อธิบายการเปลี่ยนแปลงของการบริโภคได้ดีเนื่องจากค่า Sig. เท่ากับ .0

ตารางที่ 10.8 *แสดงค่าสถิติพื้นฐาน*

Descriptive Statistics

	Mean	Std. Deviation	N
การบริโภค (ล้านบาท)	8.1000	2.23358	10
รายได้ (ล้านบาท)	29.3000	10.49921	10

จากตารางที่ 10.8 ค่าเฉลี่ยของการบริโภคเท่ากับ 8.1000 ล้านบาท ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.23358 ค่าเฉลี่ยของรายได้เท่ากับ 29.30 ล้านบาท ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10.49921 กลุ่มตัวอย่างมีจำนวน 10 คน

ตารางที่ 10.9 *แสดงค่าสถิติทดสอบสมมติฐานการถดถอย*

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.1727	.7202		3.017	.017
	รายได้ (ล้านบาท)	.2023	.0233	.951	8.692	.000

a. Dependent Variable: การบริโภค (ล้านบาท)

จากตารางที่ 10.9 พบว่าค่า b เท่ากับ 0.2023 หมายความว่า ถ้าผู้บริโภคมีรายได้เปลี่ยนแปลงไป 1 ล้านบาทแล้ว เขาจะเปลี่ยนแปลงรายจ่ายเพื่อการบริโภคไป 0.2023 ล้านบาทในทิศทางเดียวกัน

หาค่า $a = \overline{y} - b\overline{x}$

$= 8.1 - (0.2023)(29.3) = 2.1727$

$a = 2.1727$ นั่นคือ รายจ่ายขั้นต่ำเพื่อการบริโภคของผู้บริโภคเท่ากับ 2.1727 ล้านบาท

ดังนั้น **สมการการบริโภค** คือ $\hat{y} = 2.1727 + 0.2023\, x_i$

ตัวอย่างที่ 10.8 ประธานหลักสูตร MBA ของมหาวิทยาลัยแห่งหนึ่งต้องการทราบว่าเกรดเฉลี่ยสะสม (GPA) ของผู้จบการศึกษามีผลต่อเงินเดือนเมื่อเริ่มเข้าทำงานหรือไม่ จึงได้เก็บข้อมูลจากกลุ่มตัวอย่างของผู้ที่สำเร็จการศึกษา จำนวน 7 คน ปรากฏข้อมูล ดังในตาราง

คนที่	GPA	เงินเดือนเมื่อเริ่มทำงาน (พันบาท)
1	3.26	28.20
2	2.60	24.80
3	3.35	27.90
4	2.86	25.30
5	3.82	30.30
6	2.21	23.00
7	3.47	29.40

จงวิเคราะห์ข้อมูลเพื่อตอบคำถามต่อไปนี้

1. จงหาสมการถดถอยเชิงเส้นที่แสดงความสัมพันธ์ระหว่าง GPA กับ เงินเดือนเมื่อเริ่มทำงาน

2. จงหาค่า R^2 , $\overline{R}^2$ พร้อมทั้งอธิบายค่าที่ได้

3. จงทดสอบสมมติฐานค่าคงที่ของสมการถดถอยที่ได้ โดยกำหนดระดับนัยสำคัญ = 0.05

4. จงทดสอบสมมติฐานค่าสัมประสิทธิ์ของตัวแปร x ในสมการถดถอยที่ได้ โดยกำหนดระดับนัยสำคัญ = 0.05

วิธีวิเคราะห์ข้อมูล

ในโปรแกรม SPSS เมื่อป้อนข้อมูลลง Data Editor เสร็จแล้ว ให้เลือกคำสั่ง Regresstion และ Linear ในเมนู Analyze แล้วกำหนดตัวแปรที่ต้องการวิเคราะห์ค่าสถิติที่ต้องการ จากนั้นจะได้ผลลัพธ์ดังแสดงในตารางที่ 10.10 - 10.13

ตารางที่ 10.10 *แสดงการเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธี Enter*

Regression

Variables Entered/Removed[b]

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

a. All requested variables entered.

b. Dependent Variable: MONEY

จากตารางที่ 10.10 แสดงการเลือกตัวแปรอิสระ คือ GPA โดยวิธี Enter เนื่องจากเป็นตัวแปรอิสระจะเลือกใช้วิธีใดก็ได้

ตารางที่ 10.11 *แสดงค่าสัมประสิทธิ์สหสัมพันธ์ และค่าสัมประสิทธิ์การตัดสินใจ*

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.9875[a]	.9751	.9701	460.69025

a. Predictors: (Constant), GPA

จากตารางที่ 10.11 แสดงค่าสถิติสรุปของสมการถดถอย พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ (r) เท่ากับ .9875 ค่าสัมประสิทธิ์การตัดสินใจ (r^2) เท่ากับ .9751

ตารางที่ 10.12 *แสดงการวิเคราะห์ความแปรปรวนของสมการ*

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.2E+07	1	41567393.9	195.855	.000[a]
	Residual	1061178	5	212235.502		
	Total	4.3E+07	6			

a. Predictors: (Constant), GPA

b. Dependent Variable: MONEY

จากตารางที่ 10.12 ค่า F เท่ากับ 195.855 ค่า Sig. เท่ากับ .0 จึงสรุปได้ว่า **ตัวแปร GPA และ MONEY สัมพันธ์กันรูปเชิงเส้น**

ตารางที่ 10.13 *แสดงค่าประมาณสัมประสิทธิ์การถดถอย*

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	12342.42	1060.726		11.636	.000
	GPA	4752.111	339.562	.987	13.995	.000

a. Dependent Variable: MONEY

จากตารางที่ 10.13 พบว่า a เท่ากับ 12342.42 ค่า SE(a) เท่ากับ 1060.726 และ b เท่ากับ 4752.111 ค่า SE(b) เท่ากับ 339.562

ผลจากการวิเคราะห์ข้อมูล สามารถนำไปตอบคำถามต่าง ๆ ได้ดังต่อไปนี้

1. การหาสมการถดถอยเชิงเส้นที่แสดงความสัมพันธ์ระหว่าง GPA กับ เงินเดือนเมื่อเริ่มทำงาน

กำหนดให้ x แทน GPA

y แทน เงินเดือนเมื่อเริ่มทำงาน

b = 4752.111

a = 1234.24

ดังนั้น **สมการถดถอยที่ได้** คือ $\hat{y} = 1234.24 - 4752.111\, x_i$

2. การหาค่า R^2 , $\overline{R}^2$ พร้อมทั้งอธิบายค่าที่ได้ ดังนี้

R^2 = .9751 หมายความว่า **ความผันแปรทั้งหมดของเงินเดือนมีสาเหตุมาจาก GPA 97.01%**

$\overline{R}^2$ = 0.9701 หมายความว่า **เมื่อได้มีการปรับค่าสัมประสิทธิ์การตัดสินใจแล้ว ความผันแปรทั้งหมดของเงินเดือนมีสาเหตุมาจาก GPA 97.01%**

3. การทดสอบสมมติฐานค่าคงที่ของสมการถดถอยที่ได้ โดยกำหนดระดับนัยสำคัญ = 0.05

เป็นการทดสอบสมมติฐานค่าคงที่ กำหนดสมมติฐานทางสถิติได้ดังนี้

$H_0 : \beta_0 = 0$

$H_1 : \beta_0 \neq 0$

พบว่า ค่า t ที่ได้เท่ากับ 11.636 ค่า Sig. เท่ากับ .0 ดังนั้นจึง **ปฏิเสธ H_0** สรุป ได้ว่า **ค่าคงที่ไม่เท่ากับ 0 ที่ระดับนัยสำคัญ 0.05**

4. การทดสอบสมมติฐานค่าสัมประสิทธิ์ของตัวแปร x ในสมการถดถอยที่ได้ โดยกำหนดระดับนัยสำคัญ = 0.05

กำหนดสมมติฐานทางสถิติได้ดังนี้

$H_0 : \beta_1 \leq 0$

$H_1 : \beta_1 > 0$

พบว่า ค่า t ที่ได้มีค่าเท่ากับ 13.995 ค่า Sig. เท่ากับ .0 ดังนั้น **ปฏิเสธ H_0** สรุปได้ว่า **GPA กับเงินเดือนเมื่อเริ่มทำงานมีความสัมพันธ์ในทิศทางเดียวกัน ที่ระดับนัยสำคัญ 0.05**

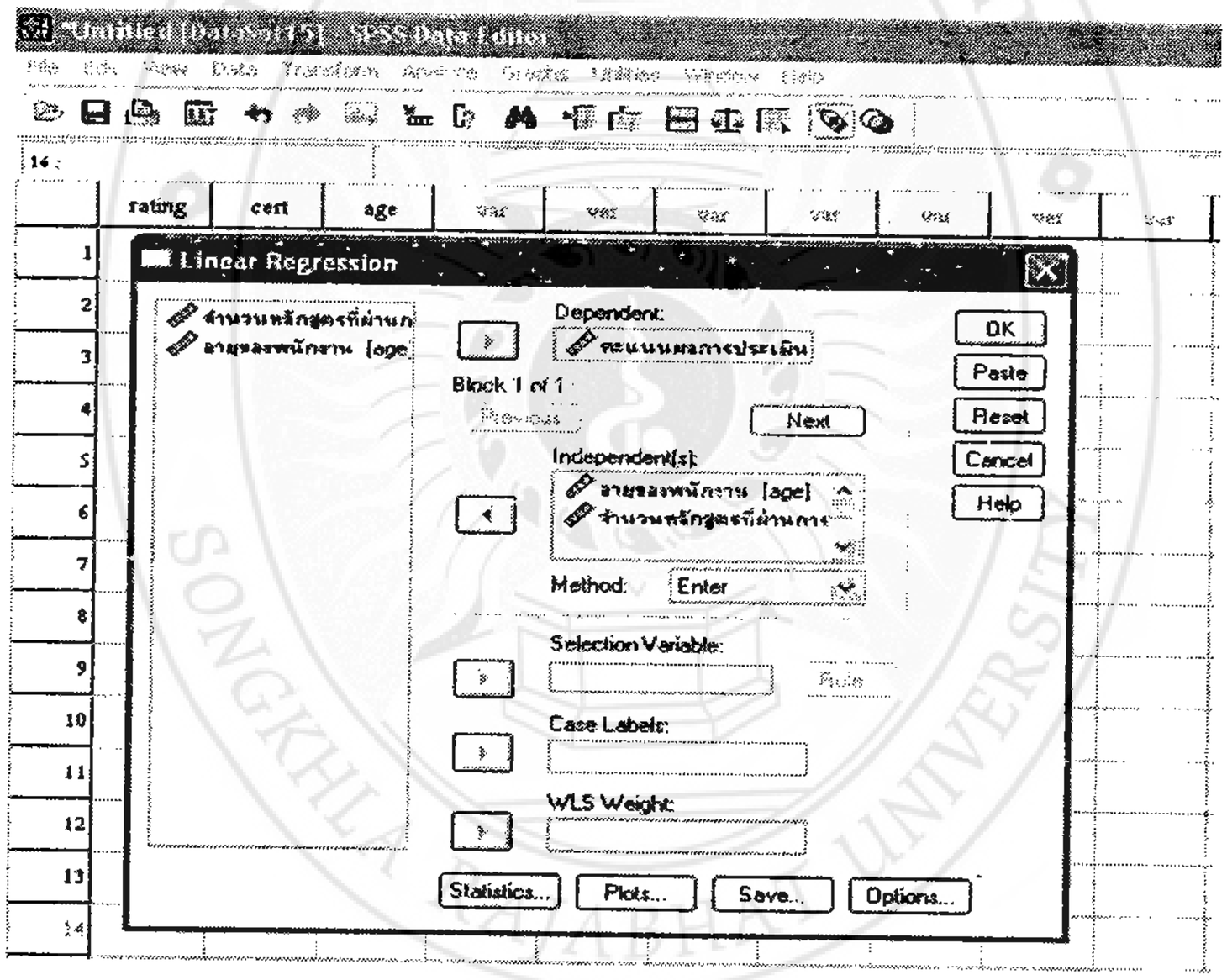
2. การวิเคราะห์การถดถอยเชิงซ้อน

การวิเคราะห์การถดถอยเชิงซ้อน (multiple regression analysis) หรือ การวิเคราะห์การถดถอยพหุคูณ เป็นการพยากรณ์ตัวแปรตาม 1 ตัว แต่มีตัวแปรอิสระตั้งแต่ 2 ตัวขึ้นไป การวิเคราะห์การถดถอยเชิงซ้อนช่วยให้ทราบอิทธิพลของตัวแปรอิสระแต่ละตัว และทั้งกลุ่มที่มีต่อตัวแปรตาม ดังตัวอย่างที่ 10.9

ตัวอย่างที่ 10.9 จงวิเคราะห์ประสิทธิภาพของพนักงานขาย จำนวน 30 คน โดยมีข้อมูลประกอบด้วย คะแนนผลการประเมิน (rating) จำนวนหลักสูตรที่ผ่านการอบรม (cert) และอายุของพนักงาน (age) เพื่อพิจารณาว่าจำนวนหลักสูตรที่พนักงานขายผ่านการอบรม และอายุของพนักงานขายมีความสัมพันธ์กันอย่างไร

วิธีการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต่าง ๆ

ในโปรแกรม SPSS ทำตามขั้นตอน คือ เลือกเมนู Analyze คำสั่ง Regression และ Linear ตามลำดับ ดังแสดงในภาพที่ 10.13 แล้วเลือกตัวแปร rating ไปไว้ในช่อง Dependent เลือกตัวแปร cert และตัวแปร age ไปไว้ในช่อง Independent(s) จะได้ผลลัพธ์ดังตารางที่ 10.14 - 10.16



ภาพที่ 10.13 *แสดงการวิเคราะห์การถดถอยเชิงซ้อน*

ตารางที่ 10.14 *แสดงค่าสัมประสิทธิ์ความสัมพันธ์*

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.823[a]	.677	.639	3.44

a. Predictors: (Constant), AGE, CERT

จากตารางที่ 10.14 แสดงผลลัพธ์ในตาราง Model Summary แสดงให้เห็นว่า ตัวแปรทั้ง 3 ตัว คือ rating, cert และ age มีความสัมพันธ์ในทิศทางเดียวกัน 82.3% โดยพิจารณาจากค่า R

ตารางที่ 10.15 *แสดงการวิเคราะห์ความแปรปรวน*

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	422.273	2	211.137	17.811	.000[a]
	Residual	201.527	17	11.855		
	Total	623.800	19			

a. Predictors: (Constant), AGE, CERT

b. Dependent Variable: RATING

จากตารางที่ 10.15 พบว่า ข้อมูลของตัวแปรทั้ง 3 ตัวอยู่บนเส้นตรงเดียวกัน และข้อมลส่วนที่เบี่ยงเบนออกจากเส้นตรงมีความแตกต่างกันถึง 17.811 เท่า ถ้ากำหนดระดับความเชื่อมั่นที่ระดับ 95% พบว่า ตัวแปร age และ cert สามารถนำมาใช้อธิบายการเปลี่ยนแปลงของ rating ได้ดี พิจารณาจากค่า Sig. เท่ากับ .000

ตารางที่ 10.16 *แสดงค่าประมาณสัมประสิทธิ์การถดถอย*

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2.970	2.686		-1.106	.284
	CERT	.754	.328	.351	2.303	.034
	AGE	.283	.071	.609	3.993	.001

a. Dependent Variable: RATING

จากตารางที่ 10.16 พบว่า b_1 มีค่าเท่ากับ .754 และ b_2 มีค่าเท่ากับ .283 ค่าคงที่ (constant หรือ a) เท่ากับศูนย์ เนื่องจากค่า Sig. มีค่ามากกว่า.05

ดังนั้น **สมการที่ใช้ในการพยากรณ์** RATING คือ

$$\text{rating} = a + b_{CERT}\,\text{cert} + b_{AGE}\,\text{age}$$

แทนค่า a ค่า b_{CERT} และค่า b_{AGE} ในสมการจะได้

$$\textbf{rating} = \textbf{0.754 cert} + \textbf{0.283 age}$$

หมายความว่า ถ้าเราทราบจำนวนหลักสูตรที่พนักงานขายผ่านการอบรมและอายุของพนักงานขาย สามารถคาดเดาได้ว่า พนักงานขายคนนั้นจะได้รับคะแนนการประเมินผลเท่าไรจากสมการข้างต้น

ตัวอย่างที่ 10.10 จากข้อมูลในตาราง จงวิเคราะห์ และตอบคำถามต่อไปนี้

1. จงหาสมการถดถอย
2. จงหาสมการถดถอยสำหรับพนักงานชาย
3. จงหาสมการถดถอยสำหรับพนักงานหญิง
4. จงหาเงินเดือนของพนักงานชายที่มีอายุการทำงาน 3 ปี
5. จงทดสอบว่าเพศและอายุการทำงานมีผลต่อเงินเดือนหรือไม่
6. เพศ และอายุการทำงาน สามารถอธิบายความผันแปรของเงินเดือนได้กี่เปอร์เซ็นต์

เงินเดือน (Y)	อายุการทำงาน (ปี) (X_1)	เพศ	เพศ (X_2)
5,300	0.3	ช	1
6,000	1.6	ญ	0
5,700	1.2	ญ	0
6,500	3.6	ช	1
6,200	2.1	ช	1
6,100	2.0	ช	1
6,200	0.9	ญ	0
7,200	4.1	ญ	0
6,400	1.7	ญ	0
6,800	3.3	ช	1
5,500	0.6	ช	1
6,200	1.89	ช	1
6,700	4.3	ช	1
6,000	2.0	ช	1
6,900	3.1	ญ	0

วิธีการวิเคราะห์การถดถอยเชิงซ้อน

ในโปรแกรม SPSS โดยการป้อนข้อมูลและวิเคราะห์ เลือกเมนู Analyze คำสั่ง Regression และ Linear ตามลำดับ จะได้ผลลัพธ์ดังแสดงในตารางที่ 10.17 - 10.20

ตารางที่ 10.17 *แสดงการเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธี Enter*

Regression

Variables Entered/Removed [b]

Model	Variables Entered	Variables Removed	Method
1	X2, X1 [a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Y

จากตารางที่ 10.17 แสดงการเลือกตัวแปรอิสระคือ X1 และ X2 ส่วน Y เป็นตัวแปรตาม โดยวิธี Enter

ตารางที่ 10.18 *แสดงค่าสัมประสิทธิ์สหสัมพันธ์และค่าสัมประสิทธิ์การตัดสินใจ*

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.935[a]	.874	.853	200.56271

a. Predictors: (Constant), X2, X1

จากตารางที่ 10.18 แสดงค่าสถิติสรุปของสมการถดถอย พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ (r) เท่ากับ .935 ค่าสัมประสิทธิ์การตัดสินใจ (r^2) เท่ากับ .874

ตารางที่ 10.19 *แสดงการวิเคราะห์ความแปรปรวน*

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3354629	2	1677314.256	41.698	.000[a]
	Residual	482704.8	12	40225.402		
	Total	3837333	14			

a. Predictors: (Constant), X2, X1

b. Dependent Variable: Y

จากตารางที่ 10.19 พบว่า ข้อมูลของตัวแปรทั้ง 3 ตัว อยู่บนเส้นตรงเดียวกัน และข้อมูลส่วนที่เบี่ยงเบนออกจากเส้นตรงมีความแตกต่างกันถึง 41.698 เท่า ถ้ากำหนดระดับความเชื่อมั่นที่ระดับ 95% พบว่า ตัวแปร X1 และ X2 สามารถนำมาใช้อธิบายการเปลี่ยนแปลงของ Y ได้ดี พิจารณาจากค่า Sig. เท่ากับ .000

ตารางที่ 10.20 *แสดงค่าประมาณสัมประสิทธิ์การถดถอย*

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5601.694	122.155		45.857	.000
	X1	380.146	43.167	.903	8.806	.000
	X2	-305.819	105.860	-.296	-2.889	.014

a. Dependent Variable: Y

จากตารางที่ 10.20 พบว่า ค่าคงที่ หรือ a เท่ากับ 5601.694 ตัวแปร X1 มีค่าเท่ากับ 380.146 และตัวแปร X2 มีค่าเท่ากับ -305.819

ผลจากการวิเคราะห์ข้อมูล สามารถนำไปตอบคำถามต่าง ๆ ได้ดังต่อไปนี้

1. สมการถดถอย คือ $\hat{Y} = \mathbf{5601.694 + 380.146X_1 - 305.819X_2}$

2. สมการถดถอยสำหรับพนักงานชาย

ชาย : $X_2 = 1$ ได้สมการถดถอยสำหรับพนักงานชาย

$$\hat{Y} = 5601.694 + 380.146X_1 - 305.819$$

$$\hat{Y} = \mathbf{5295.875 + 380.146X_1}$$

3. สมการถดถอยสำหรับพนักงานหญิง

หญิง : $X_2 = 0$ ได้สมการถดถอยสำหรับพนักงานหญิง

$$\hat{Y} = \mathbf{5601.694 + 380.146X_1}$$

4. พนักงานชายที่มีอายุการทำงาน 3 ปี ควรได้รับเงิน

พนักงานชายที่มีอายุการทำงาน 3 ปี แสดงว่า $X_1 = 3;\ X_2 = 1$

$$\hat{Y} = 5295.875 + 380.146X_3$$

$$= 5676.021$$

พนักงานชายที่มีอายุการทำงาน 3 ปี จะได้รับเงิน 5676.021 บาท

5. ทดสอบว่าเพศและอายุการทำงานมีผลต่อเงินเดือนหรือไม่

ได้ค่า t เท่ากับ -2.889 ค่า Sig. เท่ากับ .014 สรุปได้ว่า **เพศและอายุการทำงานมีผลต่อเงินเดือนของพนักงาน**

6. เพศ และอายุการทำงาน สามารถอธิบายความผันแปรของเงินเดือนได้ ดังนี้

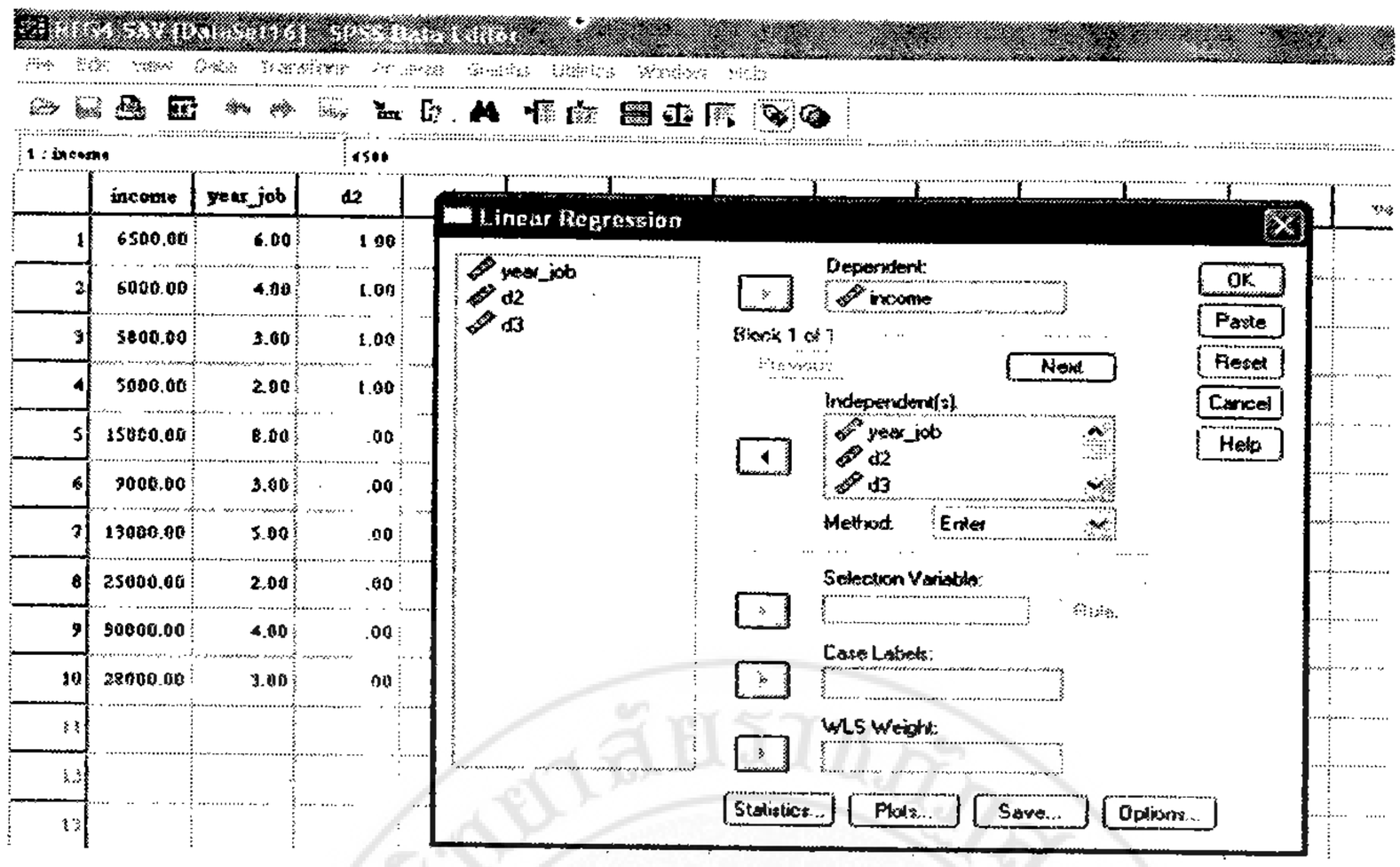
ค่า $r^2_{y.12}$ = 0.874 หมายถึง เพศ และอายุการทำงานอธิบายการผันแปรของเงินเดือนได้ 87.4%

ตัวอย่างที่ 10.11 จงใช้ข้อมูลในตาราง เพื่อสร้างสมการถดถอย

คนที่	รายได้ (Y)	อายุการทำงาน (X_1)	D_2	D_3	
1	6,500	6	1	0	อนุปริญญา
2	6,000	4	1	0	
3	5,800	3	1	0	
4	5,000	2	1	0	
5	15,000	8	0	1	ปริญญาตรี
6	9,000	3	0	1	
7	13,000	5	0	1	
8	25,000	2	0	0	สูงกว่าปริญญาตรี
9	30,000	4	0	0	
10	28,000	3	0	0	

วิธีการสร้างสมการถดถอย

การวิเคราะห์ข้อมูลด้วยโปรแกรม SPSS เมื่อป้อนข้อมูลเสร็จแล้ว เลือกคำสั่ง Regresstion Linear จะได้หน้าจอดังแสดงในภาพที่ 10.14 และกำหนดตัวแปรที่ใช้ในการวิเคราะห์การถดถอยเชิงซ้อน จะได้ผลลัพธ์ดังแสดงในตารางที่ 10.21 - 10.24



ภาพที่ 10.14 *แสดงการเลือกตัวแปรที่ใช้ในการวิเคราะห์การถดถอยเชิงซ้อน*

ตารางที่ 10.21 *แสดงการเลือกตัวแปรอิสระเข้าสมการถดถอยด้วยวิธี Enter*

Regression

Variables Entered/Removed [b]

Model	Variables Entered	Variables Removed	Method
1	D3, YEAR_JOB, D2	.	Enter

a. All requested variables entered.

b. Dependent Variable: INCOME

จากตารางที่ 10.21 แสดงการเลือกตัวแปรอิสระคือ D3 YEAR_JOB และ D2 ส่วน INCOME เป็นตัวแปรตาม โดยวิธี Enter

ตารางที่ 10.22 *แสดงค่าสัมประสิทธิ์สหสัมพันธ์ และค่าสัมประสิทธิ์การตัดสินใจ*

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.994[a]	.988	.982	1323.61374

a. Predictors: (Constant), D3, YEAR_JOB, D2

จากตารางที่ 10.22 แสดงค่าสถิติสรุปของสมการถดถอย พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ (r) เท่ากับ .994 ค่าสัมประสิทธิ์การตัดสินใจ (r^2) เท่ากับ .988

ตารางที่ 10.23 *แสดงค่าความแปรปรวน*

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8.6E+08	3	285629760	163.035	.000[a]
	Residual	1.1E+07	6	1751953.341		
	Total	8.7E+08	9			

a. Predictors: (Constant), D3, YEAR_JOB, D2

b. Dependent Variable: INCOME

จากตารางที่ 10.23 พบว่า ข้อมูลของตัวแปรทั้ง 4 ตัว อยู่บนเส้นตรงเดียวกัน และข้อมลส่วนที่เบี่ยงเบนออกจากเส้นตรงมีความแตกต่างกันถึง 163.035 เท่า ถ้ากำหนดระดับความเชื่อมั่นที่ระดับ 95% พบว่า ตัวแปร D3 YEAR_JOB และ D2 สามารถนำมาใช้อธิบายการเปลี่ยนแปลงของ INCOME ได้ดี พิจารณาจากค่า Sig. เท่ากับ .000

ตารางที่ 10.24 *แสดงค่าประมาณสัมประสิทธิ์การถดถอย*

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	24759.55	1121.309		22.081	.000
	YEAR_JOB	969.039	273.526	.186	3.543	.012
	D2	-22568.4	1031.531	-1.187	-21.879	.000
	D3	-17594.4	1255.111	-.866	-14.018	.000

a. Dependent Variable: INCOME

จากตารางที่ 10.24 พบว่า ค่าคงที่ หรือ a เท่ากับ 24759.55 YEAR_JOB มีค่าเท่ากับ 969.039 D2 มีค่าเท่ากับ -22568.4 และ D3 มีค่าเท่ากับ -17594.4 ดังนั้น จะได้สมการถดถอย คือ

$$\hat{Y} = 24759.5492 + 969.0391X_1 - 22568.4D_2 - 17594.4D_3$$

บทสรุป

การวิเคราะห์สหสัมพันธ์ เป็นการวิเคราะห์เพื่อวัดขนาดความสัมพันธ์ระหว่างตัวแปรว่า มีความสัมพันธ์กันมากน้อยเพียงใด โดยไม่สนใจตัวแปรใดว่าเป็นตัวแปรอิสระและตัวแปรใดเป็นตัวแปรตาม สามารถแบ่งเป็น 3 ประเภท คือ การวิเคราะห์สหสัมพันธ์อย่างง่าย การวิเคราะห์สหสัมพันธ์เชิงซ้อน และสหสัมพันธ์เชิงส่วน ส่วนการวิเคราะห์การถดถอย เป็นการวิเคราะห์เพื่อหารูปแบบสมการที่จะใช้อธิบายความสัมพันธ์ระหว่างตัวแปร โดยสมการที่ได้มานั้น จะใช้สำหรับประมาณค่าตัวแปรตัวหนึ่งที่เป็นตัวแปรตามจากตัวแปรอื่นที่เป็นตัวแปรอิสระที่ทราบค่าแล้วว่า มีผลหรือมีอิทธิพลต่อตัวแปรตามอย่างไร การวิเคราะห์การถดถอยแบ่งได้เป็น 2 ประเภท คือ การวิเคราะห์การถดถอยอย่างง่าย และการวิเคราะห์การถดถอยเชิงซ้อน

คำถามทบทวน

1. บอกความหมายของสหสัมพันธ์
2. สัมประสิทธ์สหสัมพันธ์คืออะไร
3. บอกความหมายของค่าสัมประสิทธิ์การตัดสินใจ
4. บอกขั้นตอนการทดสอบสหสัมพันธ์
5. บอกขั้นตอนการวิเคราะห์การถดถอย