



## บทที่ 8

# การวิเคราะห์ข้อมูลสถิติพื้นฐานด้วยโปรแกรม PSPP





ในบทนี้จะกล่าวถึงเกี่ยวกับการใช้โปรแกรมสำเร็จรูปทางสถิติด้วยโปรแกรม PSPP ซึ่งเป็นโปรแกรมโอเพนซอร์ส (open source) ด้วยการสนับสนุนจากสำนักงานส่งเสริมอุตสาหกรรมซอฟต์แวร์แห่งชาติ (องค์การมหาชน) ที่ส่งเสริมให้คนไทยใช้โปรแกรมโอเพนซอร์ส มาช่วยในการวิเคราะห์ข้อมูลโดยเฉพาะข้อมูลที่เป็นข้อมูลเชิงปริมาณ ซึ่งวิธีการวิเคราะห์ข้อมูล ด้วยวิธีการทางสถิติจะค่อนข้างยุ่งยากและซับซ้อน โดยจะแบ่งออกเป็นในส่วนของสถิติพรรณนา และสถิติอนุมาน ซึ่งจะอธิบายเฉพาะในส่วนของสถิติที่มีการนำไปใช้เสมอ ซึ่งการอธิบายจะเน้นที่การอ่านและการแปลความหมายผลลัพธ์ที่ได้จากการใช้โปรแกรม PSPP เป็นสำคัญ







## 8.1 สถิติพรรณนา

จะใช้ในการบรรยายลักษณะของข้อมูล หรือการวิเคราะห์ข้อมูลเบื้องต้น

มีสถิติสำคัญให้เลือกใช้ ดังนี้

- การหาค่าความถี่ และค่าร้อยละ
- การวัดแนวโน้มเข้าสู่ส่วนกลาง ได้แก่ ค่าเฉลี่ย มัธยฐาน และฐานนิยม
- การวัดการกระจาย ได้แก่ พิสัย ส่วนเบี่ยงเบนมาตรฐาน และค่าความแปรปรวน เป็นต้น
- Box Plot
- สถิติชีพและอัตราชีพ

แต่ในหัวข้อนี้จะขอกล่าวเพียง การหาค่าความถี่ และค่าร้อยละ การวัดแนวโน้มเข้าสู่ส่วนกลาง และการวัดการกระจาย โดยการวิเคราะห์ข้อมูลด้วยสถิติพรรณนามักจะวิเคราะห์จากแบบสอบถามในส่วนที่เป็นข้อมูลพื้นฐานหรือข้อมูลทั่วไป เช่น จากตัวอย่างแบบสอบถาม





1. เพศ

☐ ชาย

☐ หญิง

2. ภูมิลำเนา

☐ ภาคเหนือ

☐ ภาคกลาง

☐ ภาคใต้

☐ ภาคตะวันออกเฉียงเหนือ

☐ ภาคตะวันออก

3. ประสบการณ์การทำงาน.....ปี

4. อายุ.....ปี

5. รายได้ต่อเดือน.....บาท

6. อาชีพ

☐ รับราชการ

☐ ธุรกิจส่วนตัว

☐ พนักงานรัฐวิสาหกิจ

☐ พนักงานบริษัทเอกชน

☐ อื่นๆโปรดระบุ.....







# เก็บข้อมูลจากตัวอย่างจำนวน 25 ตัวอย่างได้ข้อมูลดังนี้

sex	path	experience	age	income	occupa
1	1	3	35	13,000	1
1	4	4	29	9,500	2
1	5	5	42	17,000	1
2	1	9	40	18,000	1
1	4	12	50	38,000	3
2	5	5	32	27,000	4
2	4	4	30	14,000	4
2	3	3	34	12,500	3
1	3	6	33	15,000	2
1	2	5	32	9,700	1
2	3	10	34	12,000	2
2	3	11	29	7,500	3
1	1	3	44	27,000	3
1	2	5	51	42,000	3
2	2	7	36	28,000	4
1	4	6	42	39,000	4
2	4	5	36	25,000	1
1	4	4	44	32,000	1
1	5	11	55	47,000	1
2	5	9	29	14,500	2
2	5	8	30	12,000	2
1	1	7	31	11,500	3
1	2	6	28	13,000	2
2	4	5	47	23,000	3
1	5	3	50	45,000	1





สร้างข้อมูลในตารางลงในโปรแกรมสำเร็จรูป PSPP ตามขั้นตอนดังนี้

ขั้นตอนที่ 1 สร้างตัวแปรลงในโปรแกรมสำเร็จรูป PSPP โดยคลิกไปที่

หน้าต่าง Variable view ดังรูป

ชื่อตัวแปร

ชนิดของตัวแปร

ความหมายของตัวแปร

ค่าของตัวแปร







ขั้นตอนที่ 2 คลิกไปที่หน้าต่าง Data view บันทึกข้อมูลในตารางลงในโปรแกรมสำเร็จรูป PSPP ดังรูป

\*[DataSet1] — PSPP/IRE Data Editor

File Edit View Data Transform Analyze Utilities Windows Help

Case	sex	path	experien	age	income	occupa	
1	1	1	3	35	1300	1	
2	1	4	4	29	9500	2	
3	1	5	5	42	17000	1	
4	2	1	9	40	18000	1	
5	1	4	12	50	38000	3	
6	2	5	5	32	27000	4	
7	2	4	4	30	14000	4	
8	2	3	3	34	12500	3	





9	1	3	6	33	15000	2	
10	1	2	5	32	97000	1	
11	2	3	10	34	12000	2	
12	2	3	11	29	75000	3	
13	1	1	3	44	27000	3	
14	1	2	5	51	42000	3	
15	2	2	7	36	28000	4	
16	1	4	6	42	39000	4	
17	2	4	5	36	25000	1	
18	1	4	4	44	32000	1	
19	1	5	11	55	47000	1	
20	2	5	9	29	14500	2	
21	2	5	8	30	12000	2	
22	1	1	7	31	11500	3	
23	1	2	6	28	13000	2	
24	2	4	5	47	23000	3	
25	1	5	3	50	45000	1	



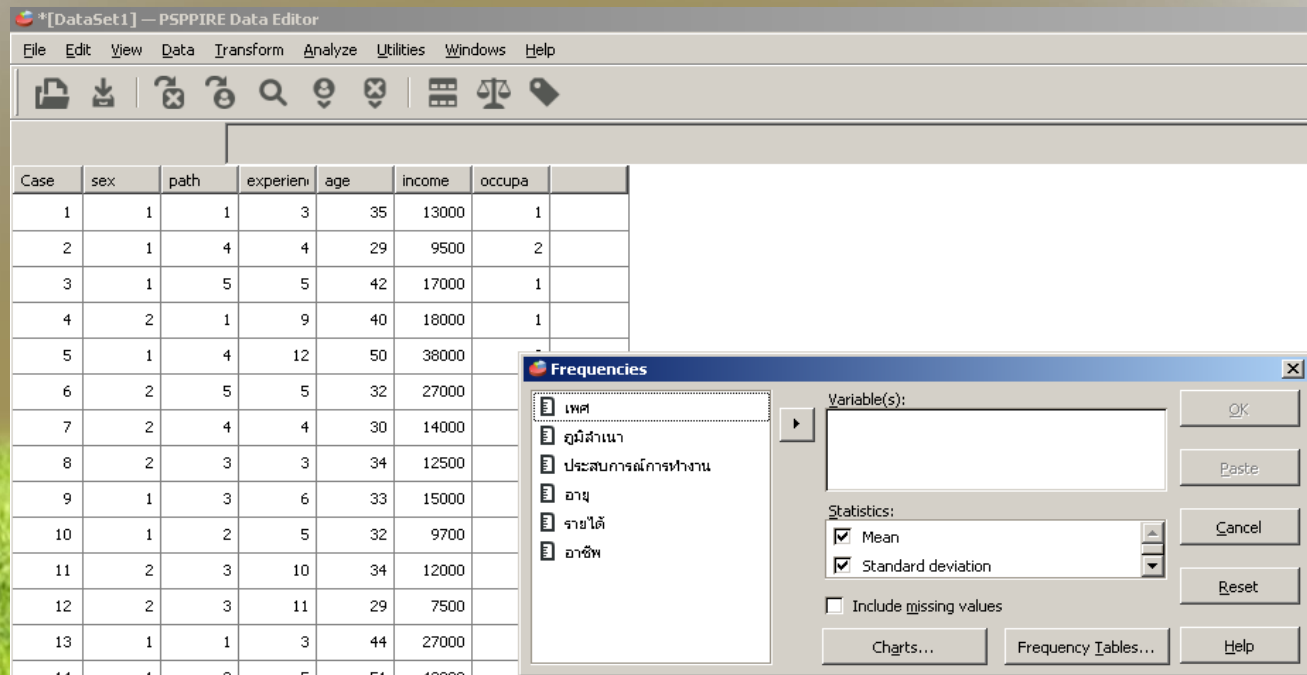




### ขั้นตอนที่ 3 บันทึกข้อมูลโดยใช้ชื่อ file ว่า Sample 1

สำหรับการพิจารณาว่าควรใช้สถิติพรรณนาตัวใดในการนำเสนอข้อมูลต้องพิจารณาที่มาตรวัดของข้อมูล ซึ่งในที่นี้จะพบว่า ตัวแปร เพศ ภูมิลำเนา และ อาชีพ อยู่ในมาตราการวัดแบบนามบัญญัติ (Nominal Scale) ดังนั้นควรเลือกใช้ค่าความถี่และร้อยละในการวิเคราะห์ข้อมูล โดยใช้คำสั่งดังนี้

1. Click Analyze —————> Frequencies จะได้หน้าจอดังรูป





2. นำตัวแปร เพศ ภูมิลำเนา และอาชีพใส่ไว้ในช่อง Variable (s) สำหรับในส่วนของ Statistics : ไม่ต้องเลือกค่าใดๆแต่ในโปรแกรมจะเลือกค่าที่โปรแกรมคิดว่าจำเป็นบางค่ามาให้โดยอัตโนมัติให้คลิกออก จากนั้นคลิก OK จะได้ผลลัพธ์ดังรูป

FREQUENCIES

FREQUENCIES

/VARIABLES= sex path occupa

/FORMAT=AVALUE TABLE

/STATISTICS=NONE.

เพศ

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
ชาย	1	14	56.00	56.00	56.00
หญิง	2	11	44.00	44.00	100.00
Total		25	100.0	100.0	

ภูมิลำเนา

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
ภาคเหนือ	1	4	16.00	16.00	16.00
ภาคกลาง	2	4	16.00	16.00	32.00
ภาคใต้	3	4	16.00	16.00	48.00
ภาคตะวันออกเฉียงเหนือ	4	7	28.00	28.00	76.00
ภาคตะวันออก	5	6	24.00	24.00	100.00
Total		25	100.0	100.0	

อาชีพ

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
รับราชการ	1	8	32.00	32.00	32.00
ธุรกิจส่วนตัว	2	6	24.00	24.00	56.00
พนักงานรัฐวิสาหกิจ	3	7	28.00	28.00	84.00
พนักงานบริษัทเอกชน	4	4	16.00	16.00	100.00
Total		25	100.0	100.0	

จากตาราง Output สามารถแปลผลได้ดังนี้





- Frequency หมายถึง ค่าความถี่หรือจำนวนข้อมูลในแต่ละกลุ่ม
- Percent หมายถึง ค่าที่แสดงความถี่ที่นับได้ในรูปร้อยละ
- Valid Percent หมายถึง ค่าที่แสดงความถี่ที่นับได้ในรูปร้อยละ  
ไม่รวมค่า Missing
- Cum Percent หมายถึง ค่าร้อยละสะสม

เมื่อมองโดยภาพรวมสามารถสรุปผลได้ว่า ผู้ตอบแบบสอบถามส่วนใหญ่เป็นเพศชาย จำนวน 14 คน คิดเป็นร้อยละ 56.0 มีภูมิลำเนาอยู่ในภาคตะวันออกเฉียงเหนือ จำนวน 7 คน คิดเป็นร้อยละ 28.0 โดยส่วนใหญ่จะประกอบอาชีพรับราชการ จำนวน 8 คน คิดเป็นร้อยละ 32.0 สำหรับตัวแปร ประสิทธิภาพการทำงาน อายุ และรายได้ต่อเดือน จะอยู่ในมาตราการวัดแบบอัตราส่วน (Ratio Scale) ดังนั้นควรเลือกใช้การวัดแนวโน้มเข้าสู่ส่วนกลางโดยใช้ค่าเฉลี่ย และวัดการกระจายโดยใช้ค่าส่วนเบี่ยงเบนมาตรฐาน โดยใช้คำสั่งดังนี้





1. Click Analyze → Frequencies จะได้หน้าจอดังรูป

SPSS Data Editor window showing a dataset with 22 cases and 7 variables (sex, path, experience, age, income, occupation, and an unnamed variable). The 'Descriptives' dialog box is open, showing the 'Statistics' section with 'Mean', 'Standard deviation', 'Variance', and 'Kurtosis' selected. The 'Options' section has 'Exclude entire case if any selected variable is missing' checked.

Case	sex	path	experience	age	income	occupation
1	1	1	3	35	13000	1
2	1	4	4	29	9500	2
3	1	5	5	42	17000	1
4	2	1	9	40	18000	1
5	1	4	12	50	38000	3
6	2	5	5	32	27000	4
7	2	4	4	30	14000	4
8	2	3	3	34	12500	3
9	1	3	6	33	15000	2
10	1	2	5	32	9700	1
11	2	3	10	34	12000	2
12	2	3	11	29	7500	3
13	1	1	3	44	27000	3
14	1	2	5	51	42000	3
15	2	2	7	36	28000	4
16	1	4	6	42	39000	4
17	2	4	5	36	25000	1
18	1	4	4	44	32000	1
19	1	5	11	55	47000	1
20	2	5	9	29	14500	2
21	2	5	8	30	12000	2
22	1	1	7	31	11500	3







2. นำตัวแปร ประสบการณ์การทำงาน อายุ และรายได้ต่อเดือน  
ใส่ไว้ในช่อง Variable (s) สำหรับในส่วนของ Statistics : ให้เลือกค่า  
Mean , Standard deviation และค่า Variance จากนั้นคลิก  
OK จะได้ผลลัพธ์ดังรูป

DESCRIPTIVES

DESCRIPTIVES

/VARIABLES= experience age income

/STATISTICS=MEAN STDDEV VARIANCE.

Valid cases = 25; cases with missing value(s) = 0.

Variable	N	Mean	Std Dev	Variance
ประสบการณ์การทำงาน	25	6.24	2.73	7.44
อายุ	25	37.72	8.16	66.63
รายได้	25	22088.00	12222.50	149389433.33





จากตาราง Output แปลผลได้ดังนี้

- N หมายถึง จำนวนข้อมูล
- Mean หมายถึง ค่าเฉลี่ย
- Std Dev หมายถึง ค่าเบี่ยงเบนมาตรฐาน
- Variance หมายถึง ค่าความแปรปรวน

สามารถสรุปได้ว่าผู้ตอบแบบสอบถามมีประสบการณ์การทำงานเฉลี่ย 6.24 ปี และมีค่าเบี่ยงเบนมาตรฐาน 2.73 ปี มีค่าความแปรปรวน 7.44 ปี<sup>2</sup> สำหรับอายุเฉลี่ยคือ 37.72 ปี มีค่าเบี่ยงเบนมาตรฐานของอายุคือ 8.163 ปี มีค่าความแปรปรวน 66.63 ปี<sup>2</sup> โดยรายได้เฉลี่ยจะอยู่ที่ 22,088.00 บาท มีค่าเบี่ยงเบนมาตรฐานของรายได้คือ 12,222.497 บาท และค่าความแปรปรวน 149389433.33 บาท<sup>2</sup>







## 8.2 สถิติอนุमान

เป็นการศึกษาสรุปลักษณะของประชากรโดยใช้ข้อมูลตัวอย่าง จัดเป็นการวิเคราะห์ข้อมูลชั้นสูง ได้แก่ การประมาณค่า การทดสอบสมมติฐาน การวิเคราะห์ถดถอยและสหสัมพันธ์ การวิเคราะห์ความแปรปรวน เป็นต้น โดยในบทนี้จะขอกกล่าวถึงเพียงแค่การประมาณค่าและการทดสอบสมมติฐาน ของค่าเฉลี่ย 1 กลุ่ม การประมาณค่าและการทดสอบสมมติฐานของค่าเฉลี่ย 2 กลุ่ม การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยมากกว่า 2 กลุ่ม การวิเคราะห์ถดถอยและสหสัมพันธ์เชิงเส้นอย่างง่าย และการวิเคราะห์ความแปรปรวน

จำแนกทางเดียว





## 8.2.1 การประมาณค่าและการทดสอบสมมติฐานของค่าเฉลี่ย 1 กลุ่ม

จากตัวอย่าง file Sample 1 หากต้องการทราบว่าประสิทธิภาพในการทำงานจะมากกว่า 5 ปีหรือไม่ โดยทดสอบที่ระดับนัยสำคัญ 0.01 สามารถทำตามขั้นตอนดังนี้

ขั้นตอนที่ 1 กำหนดสมมติฐานเพื่อการทดสอบ

$$H_0 : \mu = 5$$

$$H_1 : \mu > 5$$

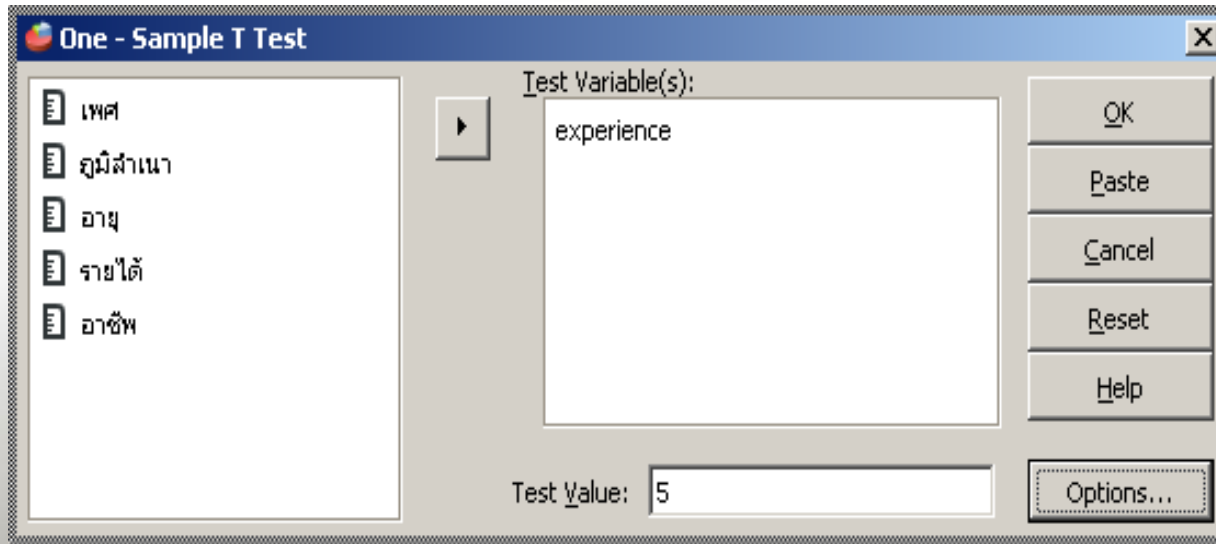
ขั้นตอนที่ 2 กำหนดระดับนัยสำคัญ ( $\alpha$ ) = 0.01

ขั้นตอนที่ 3 วิเคราะห์ข้อมูลโดย

1. Click Analyze → Compare Mean → One sample T test  
จะได้หน้าจอดังรูป



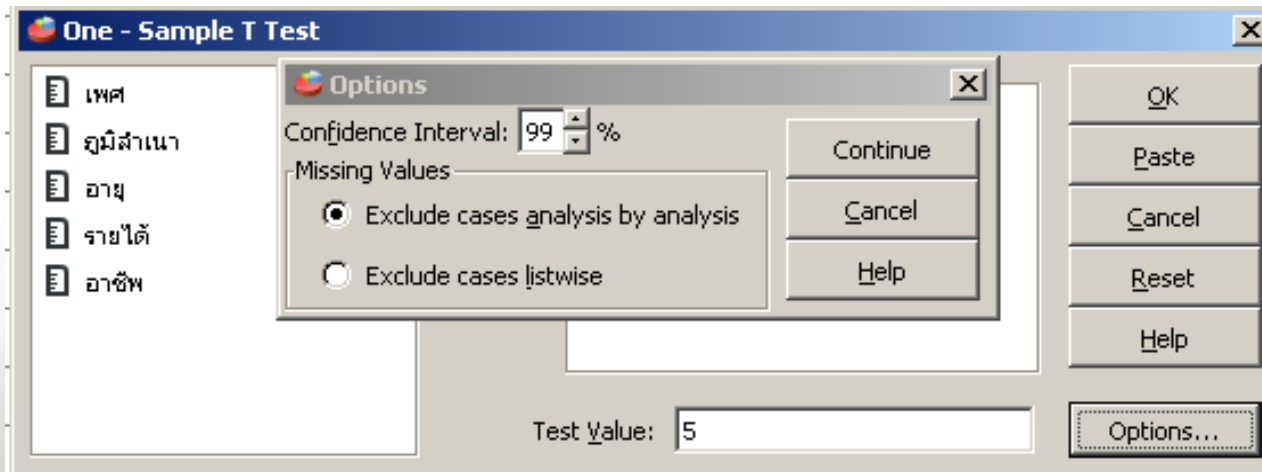




2. ในช่องของ Test Variable (s): ให้เลือกตัวแปร experience (ประสบการณ์การทำงาน) ซึ่งเป็นตัวแปรที่ต้องการทดสอบใส่ลงไป ส่วนในช่อง Test Value: ให้ใส่ค่าที่ต้องการทดสอบซึ่งในที่นี้คือ 5 ปี

3. Click Options จะได้หน้าจอดังรูป





ในช่อง Confidence Interval ให้ใส่ระดับนัยสำคัญที่ต้องการทดสอบ  
ลงไป ในที่นี้ให้ใส่ 99%

4. Click Continue จากนั้น OK จะได้ผลลัพธ์ดังรูป



T-TEST  
T-TEST /TESTVAL=5  
/VARIABLES= experience /MISSING=ANALYSIS  
/CRITERIA=CIN(0.99).

One-Sample Statistics

	N	Mean	Std. Deviation	S.E. Mean
ประสบการณ์การทำงาน	25	6.24	2.73	.55

One-Sample Test

	Test Value = 5.000000					
	t	df	Sig. (2-tailed)	Mean Difference	99% Confidence Interval of the Difference	
					Lower	Upper
ประสบการณ์การทำงาน	2.27	24	.032	1.24	-.29	2.77





จากตารางผลการวิเคราะห์ (Out put) สามารถแปลผลได้ดังนี้

- ตารางแรกบอกให้ทราบว่าข้อมูลเกี่ยวกับประสิทธิภาพการทำงานที่นำมาวิเคราะห์ทั้งหมด 25 case โดยมีประสิทธิภาพการทำงานเฉลี่ย 6.24 ปี มีค่าเบี่ยงเบนมาตรฐานเป็น 2.73 และมีค่าเบี่ยงเบนของความคลาดเคลื่อนเฉลี่ยเป็น 0.55
- ตารางที่สองเป็นการสรุปผลสมมติฐานที่กำหนดไว้ที่ระดับนัยสำคัญ 0.01 โดยที่สมมติฐานคือ

$$H_0 : \mu = 5$$

$$H_1 : \mu > 5$$

ซึ่งจากตารางจะได้ว่าค่า  $t = 2.27$  ซึ่งต้องไปเปิดตารางสถิติ  $t$  เปรียบเทียบ แต่เนื่องจากใน Output บอกค่า Sig มาด้วย ดังนั้นจะใช้ค่า Sig ในการสรุปผล โดยที่ถ้าค่า sig มากกว่าระดับนัยสำคัญจะยอมรับสมมติฐาน  $H_0$  ในทางกลับกันถ้าค่า sig น้อยกว่าระดับนัยสำคัญจะปฏิเสธสมมติฐาน  $H_0$  ซึ่งจาก Output มีค่า Sig (2 – tailed) เป็น 0.032 แต่เนื่องจากการทดสอบ



สมมติฐานเป็นแบบข้างเดียวทางขวา จึงต้องใช้ค่า Sig (2 – tailed)  
หาร 2 จะได้ค่า sig เท่ากับ  $\frac{0.032}{2} = 0.016$  ซึ่งมีค่ามากกว่าระดับนัยสำคัญ 0.01

จึงยอมรับสมมติฐาน  $H_0$  นั่นคือ ประสิทธิภาพในการทำงานไม่มากกว่า 5 ปี และ  
จากตารางจะสามารถประมาณค่าเฉลี่ยของประสิทธิภาพการทำงานได้ด้วยว่าอยู่ในช่วง -  
 $0.29 + 5 = 4.71$  ถึง  $2.77 + 5 = 7.77$  ปี (4.71, 7.77 ปี)

## 8.2.2 การประมาณค่าและการทดสอบสมมติฐานของค่าเฉลี่ย 2 กลุ่ม

แบ่งออกเป็น 2 กรณีคือ

### 8.2.2.1 กรณีที่ประชากร 2 กลุ่มไม่เป็นอิสระต่อกัน (มีความสัมพันธ์กัน)

ในกรณีที่ประชากร 2 กลุ่มไม่เป็นอิสระต่อกัน เช่น ทดสอบ  
การทำงานของคู่มือ การทดลองยาลดความอ้วน ประสิทธิภาพการใช้  
งานก่อน และหลังการอบรมการใช้ จะสังเกตเห็นว่าประชากรเป็น  
ประชากรไม่เป็นอิสระต่อกัน เราเรียกข้อมูลลักษณะนี้ว่าเป็นข้อมูลคู่  
(pair data)







ตัวสถิติทดสอบคือ 
$$t = \frac{\bar{D} - \mu_D}{\frac{S_D}{\sqrt{n}}}$$

จะมีการแจกแจงแบบ t ด้วยองศาความเป็นอิสระ  $n - 1$

ตัวอย่าง ทดสอบความสามารถของนักศึกษาคณะวิทยาศาสตร์ในการใช้อุปกรณ์ในห้องปฏิบัติการเคมีก่อนและหลังการอบรมได้เวลาในการทำ Lab ดังนี้

คนที่	เวลาก่อนอบรม	เวลาหลังอบรม
1	55	50
2	46	42
3	78	70
4	61	63
5	52	58
6	45	35
7	47	46





คนที่	เวลาก่อนอบรม	เวลาหลังอบรม
8	57	52
9	71	60
10	58	49

อยากทราบเมื่อนักศึกษาผ่านการอบรมจะใช้เวลาในการทำ Lab น้อยลงกว่าเดิมหรือไม่ ที่ระดับนัยสำคัญ 0.05

ขั้นตอนการทดสอบมีดังนี้

ขั้นตอนที่ 1 กำหนดสมมติฐานเพื่อการทดสอบ

$$H_0 : \mu_D = 0$$

$$H_1 : \mu_D > 0$$

ขั้นตอนที่ 2 กำหนดระดับนัยสำคัญ ( $\alpha$ ) = 0.05







### ขั้นตอนที่ 3 วิเคราะห์ข้อมูลโดยใช้คำสั่ง Pair T test โดยใช้โปรแกรมสำเร็จรูปทางสถิติ PSPP ดังนี้

1. สร้างข้อมูลในตารางลงในโปรแกรมสำเร็จรูป PSPP บันทึกข้อมูลจะได้ ดังรูป

\*[DataSet1] — PSPIRE Data Editor

File Edit View Data Transform Analyze Utilities Windows Help

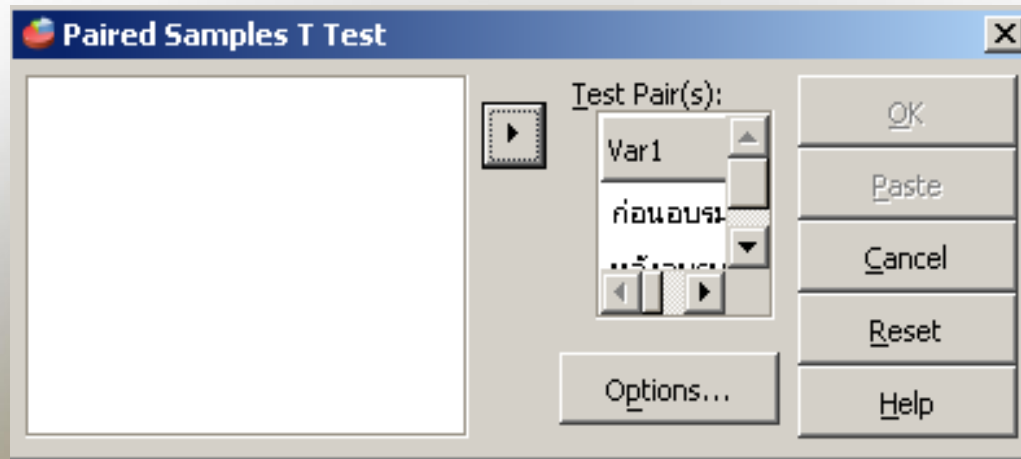
Icons: Save, Open, Print, Find, etc.

Case	ก่อนอบรม	หลังอบรม	
1	55	50	
2	46	42	
3	78	70	
4	61	63	
5	52	58	
6	45	35	
7	47	46	
8	57	52	
9	71	60	
10	58	49	
11			





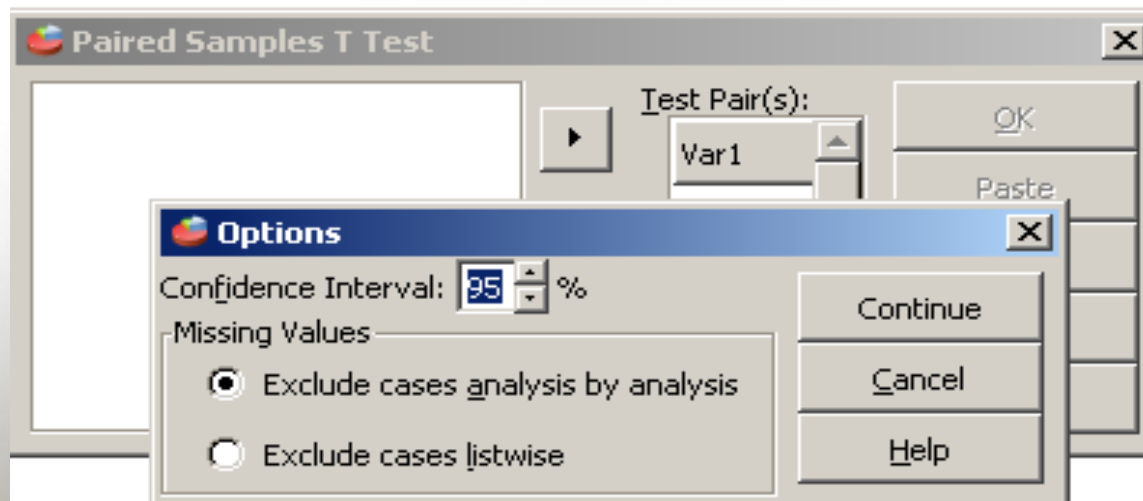
2. Click Analyze → Compare Mean → Paired sample T test จะได้หน้าจอดังรูป



3. นำตัวแปรก่อนการอบรมและหลังอบรมใส่ไว้ที่ช่อง Test Pair(s) จากนั้น Click options... จะได้ดังรูป







4. ในช่อง Confidence Interval ให้ใส่ 95% เนื่องจากโจทย์กำหนดระดับนัยสำคัญเป็น 0.05
5. Click Continue จากนั้น OK จะได้ผลลัพธ์ดังรูป





T-TEST

T-TEST

PAIRS = ก่อนอบรม WITH หลังอบรม (PAIRED)  
/MISSING=ANALYSIS  
(CRITERIA=CIN(0.95)).

Paired Sample Statistics

		Mean	N	Std. Deviation	S.E. Mean
Pair 1	ก่อนอบรม	57.00	10	10.79	3.41
	หลังอบรม	52.50	10	10.44	3.30

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	ก่อนอบรม & หลังอบรม	10	.87	.001

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	ก่อนอบรม - หลังอบรม	4.50	5.48	1.73	.58	8.42	2.60	9	.029

จากตารางการวิเคราะห์จะได้ค่า p-value หรือค่า  $\text{Sig} = \frac{0.029}{2} = 0.0145$

$< 0.05$  ดังนั้นจะ ปฏิเสธ  $H_0$  สรุปว่าเมื่อนักศึกษาผ่านการอบรมจะใช้เวลา

ในการทำ Lab น้อยลงกว่าเดิมที่ระดับนัยสำคัญ 0.05 และ ผลต่างของเวลาที่ใช้ในการทำ Lab ก่อนและหลังการอบรมมีค่าอยู่ระหว่าง 0.58 และ 8.42 ด้วยความเชื่อมั่น 95 %







### 8.2.2.2 กรณีที่ประชากร 2 กลุ่มเป็นอิสระต่อกัน (ไม่มีความสัมพันธ์กัน)

เป็นการทดสอบผลต่างระหว่างค่าเฉลี่ยของลักษณะที่สนใจของ 2 ประชากรว่าแตกต่างกันหรือไม่ หรือเป็นการทดสอบว่าค่าเฉลี่ยของประชากรที่ 1 มากกว่าประชากรที่ 2 หรือไม่ เช่น ต้องการทดสอบรายได้ของเพศชายกับเพศหญิงว่าแตกต่างกันหรือไม่ สมมติฐานการวิจัยจะกำหนดได้ดังนี้

$H_0 : \mu_1 = \mu_2$  หมายถึง รายได้เฉลี่ยเท่ากันหรือไม่แตกต่างกัน

$H_1 : \mu_1 \neq \mu_2$  หมายถึง รายได้เฉลี่ยแตกต่างกัน

หรือ  $H_1 : \mu_1 > \mu_2$  หรือ  $H_1 : \mu_1 < \mu_2$





ตัวอย่าง ฝ่ายวิจัยต้องการทดสอบว่า เวลานอนของนักศึกษาชายและหญิงคณะ  
วิทยาศาสตร์ แตกต่างกันหรือไม่ โดยการแจกแจงของเวลานอนมีการแจกแจงปกติ  
กลุ่มตัวอย่างนักศึกษาคณะวิทยาศาสตร์ บันทึกเวลานอนเป็นชั่วโมง/วัน ดังนี้

ชาย				หญิง				
7.0	8.0	4.0	8.0	8.0	8.0	7.0	6.0	7.5
5.5	7.0	7.0	5.0	7.0	9.0	8.0	7.5	7.0
9.0	6.0	6.0	6.5	8.5	6.5	7.0	6.0	8.5
6.5	10.0	7.0	9.0	6.0	8.0	6.0	7.5	9.5
6.0	7.0	7.5	7.0	6.0	7.0	7.5	8.0	
7.0	6.0	7.0	5.5	8.0	7.0	8.0	9.0	
7.5	6.5	8.0	6.5	6.5	8.0	8.0	5.0	
8.0	7.0	8.0	6.0	7.0	5.5	6.0	8.0	
7.0	7.0	4.5	8.5	7.0	8.0	8.5	7.0	
6.0	5.0	7.0	8.0	6.0	9.0	8.0	7.0	
6.0	6.0	6.0		7.5	6.0	7.5	8.0	

จงทดสอบสมมติฐานว่าเวลานอนของนักศึกษาชายและหญิงคณะวิทยาศาสตร์  
แตกต่างกันหรือไม่ พร้อมทั้งประมาณค่าผลต่างของเวลานอนของนักศึกษาชาย  
และหญิงคณะวิทยาศาสตร์ โดยใช้ระดับนัยสำคัญ 0.05





กำหนดให้  $\mu_1$  แทนเวลานอนเฉลี่ยของนักศึกษาชายคณะวิทยาศาสตร์  
 $\mu_2$  แทนเวลานอนเฉลี่ยของนักศึกษาหญิงคณะวิทยาศาสตร์

ขั้นตอนการทดสอบสมมติฐาน

ขั้นตอนที่ 1 กำหนดสมมติฐานเพื่อการทดสอบ

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

ขั้นตอนที่ 2 กำหนดระดับนัยสำคัญ ( $\alpha$ ) = 0.05

ขั้นตอนที่ 3 วิเคราะห์ข้อมูลโดยใช้คำสั่ง Independent - Sample T test  
โดยใช้โปรแกรมสำเร็จรูปทางสถิติ PSPP ดังนี้





1. สร้างข้อมูลในตารางลงในโปรแกรมสำเร็จรูป PSPP บันทึกข้อมูลโดยใช้ชื่อ file ว่า Sample 3 ดังรูป

sample3.sav [DataSet1] — PSPPIRE Data Editor

File Edit View Data Transform Analyze Utilities Windows Help

22 : เวลาที่ใช้ในการนอน 7.0

Case	เพศ	เวลาที่ใช้ในการนอน
1	1	7.0
2	1	5.5
3	1	9.0
4	1	6.5
5	1	6.0
6	1	7.0
7	1	7.5
8	1	7.0
9	1	6.0
10	1	6.0
11	1	10.0
12	1	7.0
13	1	6.0
14	1	6.5
15	1	7.0
16	1	7.0
17	1	5.0
18	1	6.0
19	1	4.0
20	1	7.0







sample3.sav [DataSet2] — PSPPIRE Data Editor

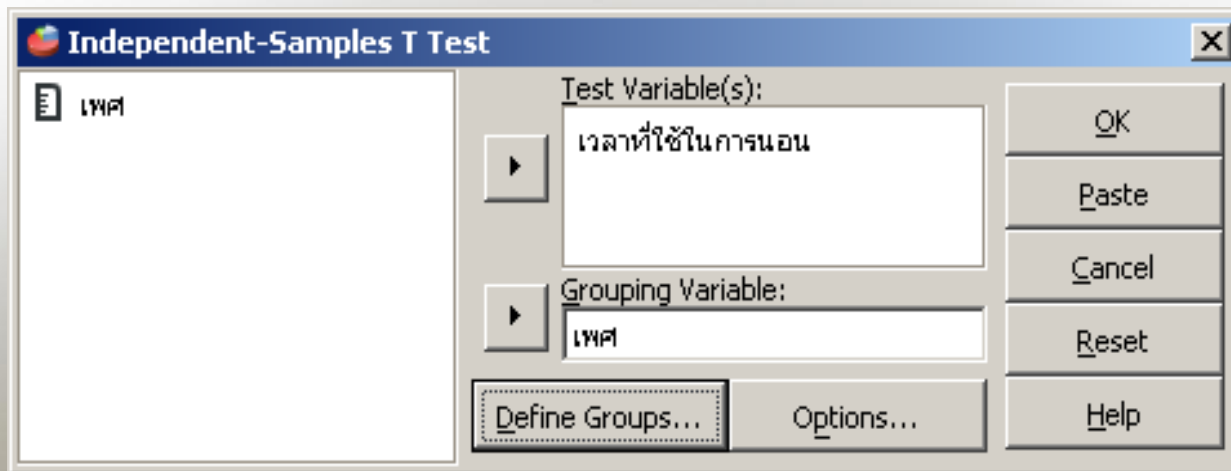
File Edit View Data Transform Analyze Utilities Windows Help

Case	เพศ	เวลาที่ใช้
71	2	6.0
72	2	8.0
73	2	6.0
74	2	8.5
75	2	8.0
76	2	7.5
77	2	6.0
78	2	7.5
79	2	6.0
80	2	7.5
81	2	8.0
82	2	9.0
83	2	5.0
84	2	8.0
85	2	7.0
86	2	7.0
87	2	8.0
88	2	7.5
89	2	7.0
90	2	8.5
91	2	9.5





2. Click Analyze → Compare Mean → Independent - Sample T test  
จะได้หน้าจอดังรูป



3. นำตัวแปรเวลาที่ใช้ในการนอนใส่ไว้ในช่อง Test Variable(s): และตัวแปรเพศใส่ไว้ในช่อง Grouping Variable: จากนั้น Click Define Groups... จะได้ดังรูป







The image shows two overlapping SPSS dialog boxes. The top box is the 'Independent-Samples T Test' dialog, with 'เพศ' (Gender) in the list on the left. The 'Test Variable(s):' field contains 'เวลาที่ใช้ในการนอน' (Time spent sleeping). The 'Grouping Variable:' field contains 'เพศ'. The bottom box is the 'Define Groups' dialog, with the 'Use specified values:' radio button selected. 'Group1 value:' is set to 'ชาย' (Male) and 'Group2 value:' is set to 'หญิง' (Female). The 'Continue' button is visible in the bottom right of the 'Define Groups' dialog.

4. ใน Group1 value : ใส่ชาย และใน Group2 value : ใส่หญิง  
จากนั้น Click Continue แล้ว OK จะได้ผลลัพธ์ดังรูป





#### T-TEST

T-TEST /VARIABLES= เวลาที่ใช้ในการนอน  
/GROUPS=เพศ(1,2) /MISSING=ANALYSIS  
/CRITERIA=CIN(0.95).

#### Group Statistics

	เพศ	N	Mean	Std. Deviation	S.E. Mean
เวลาที่ใช้ในการนอน	ชาย	43	6.81	1.22	.19
	หญิง	48	7.34	1.02	.15

#### Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
เวลาที่ใช้ในการนอน	Equal variances assumed	.43	.512	-2.26	89.00	.026	-.53	.23	-.99	-.06
	Equal variances not assumed			-2.24	82.28	.028	-.53	.24	-1.00	-.06

ความหมายของผลลัพธ์ในตาราง

อธิบายได้เป็น 2 ขั้นตอน ดังนี้

ขั้นที่ 1 ต้องตรวจสอบว่าค่าความแปรปรวนของประชากรทั้ง 2 เท่ากันหรือไม่ โดยกำหนดสมมติฐานการทดสอบดังนี้







$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

- สถิติที่ใช้ทดสอบใช้ F โดยดูได้จาก Column ของ Levene's Test for Equality of Variances เนื่องจากในที่นี้เป็นการทดสอบ 2 ด้านจึงเปรียบเทียบ Sig. กับค่า  $\alpha$  ที่กำหนด ในตัวอย่างนี้ Sig. = 0.512 ซึ่งมากกว่า 0.05 จึงยอมรับ  $H_0$  นั่นคือ เวลานอนของนักศึกษาชายและนักศึกษาหญิงมีความแปรปรวนไม่แตกต่างกัน

ขั้นที่ 2 การทดสอบสมมติฐานเปรียบเทียบเวลานอนของนักศึกษาชายและนักศึกษาหญิง

- จากขั้นที่ 1 พบว่า ค่าความแปรปรวนของ 2 กลุ่มตัวอย่างไม่แตกต่างกัน ดังนั้นจะใช้ค่าสถิติทดสอบ t ในส่วน Equal Variances Assumed ซึ่งจะได้อ่านค่า t จากตารางผลลัพธ์เท่ากับ -2.26 นำไปเปรียบเทียบกับค่า t ที่เปิดจากตาราง





- ค่า df หมายถึง องศาความเป็นอิสระของ

$$t = \frac{\left[ \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right]}{\frac{\left[ \frac{s_1^2}{n_1} \right]^2}{n_1 - 1} + \frac{\left[ \frac{s_2^2}{n_2} \right]^2}{n_2 - 1}} = 89$$

- ค่า Sig (2-tailed) เท่ากับ 0.026 สรุปผลการทดสอบเปรียบเทียบกับค่า  $\alpha = 0.05$  โดยมีค่าน้อยกว่า  $\alpha = 0.05$  ดังนั้นจะปฏิเสธสมมติฐาน  $H_0$  การทดสอบมีนัยสำคัญ นั่นคือเวลาอนเจเลียของนักศึกษาชายและนักศึกษาหญิงคณะวิทยาศาสตร์แตกต่างกัน

- Mean Difference หมายถึง ผลต่างของรายได้เฉลี่ยของผู้จัดการชายและหญิง

$$(\bar{X}_1 - \bar{X}_2) = (6.814 - 7.344) = -0.53$$

Std. Error Difference หมายถึง  $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 0.23$







- 95% Confidence Interval of the Mean หมายถึง ค่าประมาณแบบช่วงของ  $\mu_1 - \mu_2$  ที่ระดับความเชื่อมั่น 95% โดยมีค่าเท่ากับ  $0.99 < \mu_1 - \mu_2 < -0.06$  นั่นคือ ผลต่างระหว่างเวลานอนเฉลี่ยของชายและนักศึกษาหญิงจะอยู่ในช่วง 0.99 ชั่วโมง ถึง -0.06 ชั่วโมง



### 8.2.3 การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยมากกว่า 2 กลุ่ม (การวิเคราะห์ความแปรปรวน)

ในบทนี้จะขอกล่าวถึงเพียงแค่การวิเคราะห์ความแปรปรวนจำแนกทางเดียว หรือ One-way ANOVA ซึ่งเป็นวิธีการทดสอบเพื่อวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระหรือตัวแปรต้นตัวเดียวกับตัวแปรตามเพียงตัวเดียวโดยที่ตัวแปรอิสระหรือตัวแปรต้นอาจมีลักษณะเป็นตัวแปรเชิงคุณภาพ (Qualitative Variable) ที่จำแนกออกเป็นระดับหรือประเภทต่าง ๆ เช่น เก่ง-ปานกลาง-อ่อน ดีมาก-ดี-พอใช้-แย่ เป็นต้น ส่วนตัวแปรตามอาจมีลักษณะเป็นตัวแปรเชิงปริมาณ (Quantitative Variable) เพื่อศึกษาความสัมพันธ์ของตัวแปรอิสระหรือตัวแปรต้นว่าจะส่งผลอย่างไรกับตัวแปรตาม ตามสมมติฐานการวิจัยที่กำหนดไว้





ตัวอย่าง . จากการสอบถามร้านที่จำหน่ายเครื่องคอมพิวเตอร์  
4 ยี่ห้อ โดยสอบถามมายี่ห้อละ 6 ร้าน เพื่อเปรียบเทียบจำนวนที่ขายได้ของ  
เครื่องคอมพิวเตอร์ทั้ง 4 ยี่ห้อว่ามีความแตกต่างกันหรือไม่ได้ข้อมูลดังตาราง

ยี่ห้อคอมพิวเตอร์			
I	II	III	IV
78	64	55	75
91	72	66	93
97	68	49	78
82	77	64	71
85	56	70	63
77	95	68	76

จงทดสอบว่าจำนวนที่ขายได้ของเครื่องคอมพิวเตอร์ทั้ง 4 ยี่ห้อแตกต่างกันหรือไม่ โดยทดสอบที่ระดับนัยสำคัญ .10







## ขั้นตอนการทดสอบสมมติฐาน

ขั้นตอนที่ 1 กำหนดสมมติฐานเพื่อการทดสอบ

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

$H_1$ : มีอย่างน้อยหนึ่งคู่ที่ไม่เท่ากัน

ขั้นตอนที่ 2 กำหนดระดับนัยสำคัญ ( $\alpha$ ) = 0.10

ขั้นตอนที่ 3 วิเคราะห์ข้อมูลโดยใช้คำสั่ง One – Way ANOVA โดยใช้โปรแกรมสำเร็จรูปทางสถิติ PSPP ดังนี้

1. สร้างข้อมูลในตารางลงในโปรแกรมสำเร็จรูป PSPP บันทึกข้อมูลโดยใช้ชื่อ file ว่า Sample 4 ดังรูป



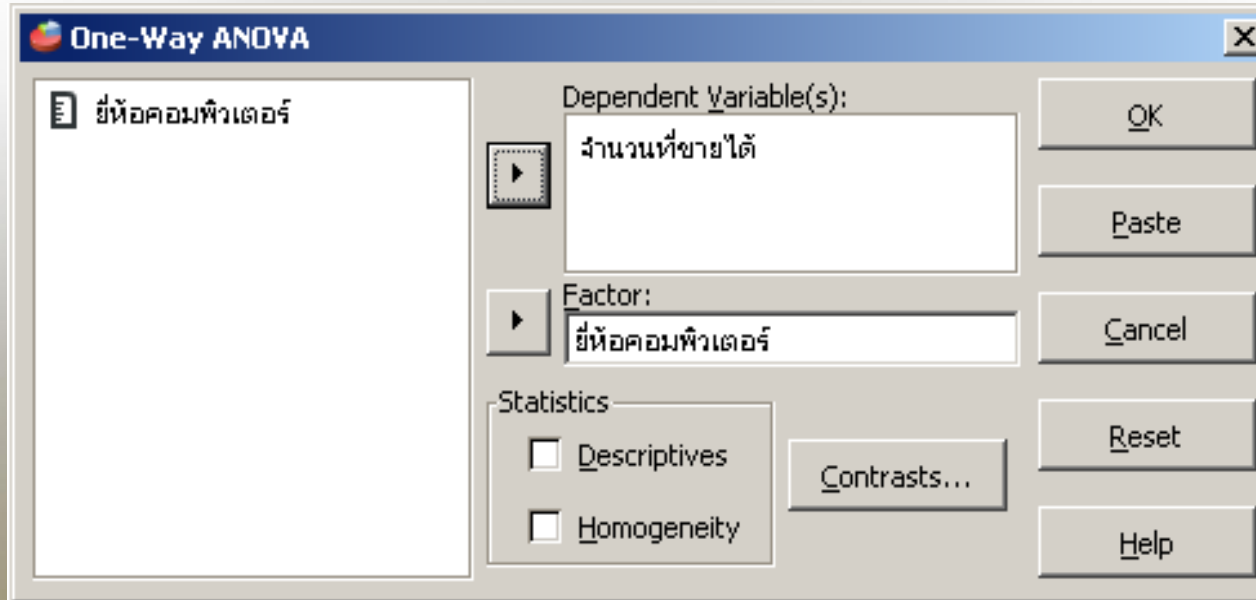


Case	ยี่ห้อคอมพิวเตอร์	จำนวนที่ขายได้	
1	1	78.00	
2	1	91.00	
3	1	97.00	
4	1	82.00	
5	1	85.00	
6	1	77.00	
7	2	64.00	
8	2	72.00	
9	2	68.00	
10	2	77.00	
11	2	56.00	
12	2	95.00	
13	3	55.00	
14	3	66.00	
15	3	49.00	
16	3	64.00	
17	3	70.00	
18	3	68.00	
19	4	75.00	
20	4	93.00	
21	4	78.00	





2. Click Analyze → Compare Mean → One - Way ANOVA  
จะได้หน้าจอ ดังรูป



3. ตัวแปรจำนวนที่ขายได้ไปไว้ในช่อง Dependent Variable (s) ส่วนตัวแปรยี่ห้อคอมพิวเตอร์ใส่ไว้ในช่อง Factor : จากนั้น Click OK จะได้ผลลัพธ์ ดังรูป





ONEWAY

ONEWAY /VARIABLES= จำนวนที่ขายได้ BY ยี่ห้อคอมพิวเตอร์.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
จำนวนที่ขายได้	Between Groups	1636.50	3	545.50	5.41	.007
	Within Groups	2018.00	20	100.90		
	Total	3654.50	23			

## การสรุปผล

ให้พิจารณาค่า Sig ถ้าค่า  $Sig < \alpha$  จะปฏิเสธสมมติฐาน  $H_0$

แต่ ถ้าค่า  $Sig > \alpha$  จะยอมรับสมมติฐาน  $H_0$

ในที่นี้ค่า  $Sig = 0.007$  ซึ่งน้อยกว่าค่า  $\alpha = 0.05$  ดังนั้นจะปฏิเสธสมมติฐาน  $H_0$  นั่นคือ จำนวนที่ขายได้ของเครื่องคอมพิวเตอร์ทั้ง 4 ยี่ห้อแตกต่างกัน







## 8.2.4 การวิเคราะห์ถดถอยและสหสัมพันธ์เชิงเส้นอย่างง่าย

กรณีที่มีตัวแปร 2 ตัวแปร จากหน่วยตัวอย่างเดียวกัน ซึ่งเรียกว่า ข้อมูลตัวแปรคู่ จะศึกษาความสัมพันธ์ของตัวแปรทั้งสองตัว เช่น ศึกษาความสัมพันธ์ระหว่างปริมาณการดื่มแอลกอฮอล์(ลิตร/วัน)กับอัตราการตายด้วยโรคมะเร็งตับ ความดันโลหิตกับอายุ ตัวแปรที่ต้องการทราบค่าหรือต้องการพยากรณ์เรียกว่าตัวแปรตาม ในที่นี้แทนด้วย  $Y$  ส่วนตัวแปรที่ทราบค่าซึ่งเป็นตัวแปรที่มีผลกระทบต่อตัวแปรตาม เรียกว่า ตัวแปรอิสระ ในที่นี้แทนด้วย  $X$  จากตัวอย่าง ที่กล่าวมาข้างต้นอาจสรุปได้ว่า อัตราการตายด้วยโรคมะเร็งตับ และความดันโลหิต ต่างเป็นตัวแปรตามแทนด้วย  $Y$  และปริมาณการดื่มแอลกอฮอล์ (ลิตร/วัน) กับอายุ เป็นตัวแปรอิสระ แทนด้วย  $X$  โดยถ้าศึกษาความสัมพันธ์ของตัวแปร  $X$  และ  $Y$  เพื่อพยากรณ์ค่าของ  $Y$  เมื่อทราบค่า  $X$  จะใช้สมการถดถอย (regression equation) ซึ่งรูปแบบความสัมพันธ์ของตัวแปร  $X$  กับ  $Y$  จะมีอยู่หลายรูปแบบทั้งแบบเส้นตรง พาราโบลา และเอ็กซ์โพเนนเชียล แต่ถ้าต้องการศึกษาเฉพาะความสัมพันธ์ระหว่างตัวแปรต้นกับตัวแปรตามเท่านั้นจะดูได้จากค่าสัมประสิทธิ์สหสัมพันธ์ โดยในบทนี้จะกล่าวถึงเฉพาะรูปแบบความสัมพันธ์ที่เป็นเชิงเส้นตรงอย่างง่ายเท่านั้น





ตัวอย่าง จากข้อมูลแสดงส่วนสูงและน้ำหนักของนักเรียน  
โรงเรียนแห่งหนึ่ง จำนวน 20 คน ดังนี้

ส่วนสูง	น้ำหนัก	ส่วนสูง	น้ำหนัก
90	12	140	27
100	15	150	50
100	20	150	55
110	20	155	57
130	23	135	43
115	16	140	40
95	13	130	23
120	22	120	23
140	25	90	13
140	26	155	56







เมื่อทำการพล็อตกราฟจะได้ว่าส่วนสูงและน้ำหนักของนักเรียนโรงเรียนแห่งนี้มีความสัมพันธ์กันเป็นเชิงเส้นตรง ดังนั้นสามารถหารูปแบบของสมการถดถอย โดยใช้สมการเชิงเส้นตรงและหาค่าสัมประสิทธิ์สหสัมพันธ์ได้ดังนี้

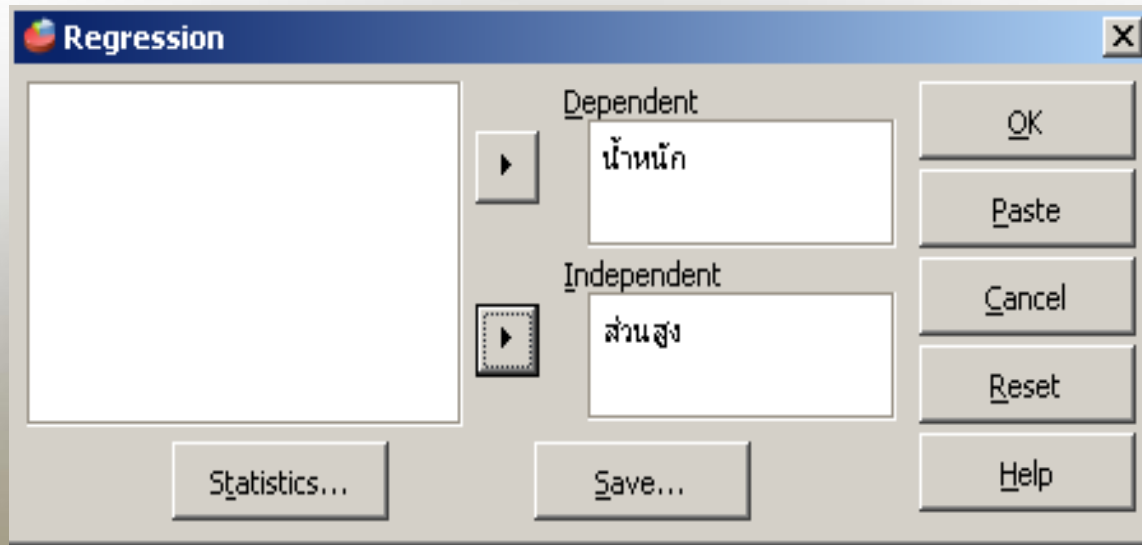
1. สร้างข้อมูลในตารางลงในโปรแกรมสำเร็จรูป PSPP บันทึกข้อมูลโดยใช้ชื่อ file ว่า Sample 5 ดังรูป

Case	ส่วนสูง	น้ำหนัก
1	90	12
2	100	15
3	100	20
4	110	20
5	130	23
6	115	16
7	95	13
8	120	22
9	140	25
10	140	26
11	140	27
12	150	50
13	150	55
14	155	57
15	135	43
16	140	40
17	130	23
18	120	23
19	90	13
20	155	56





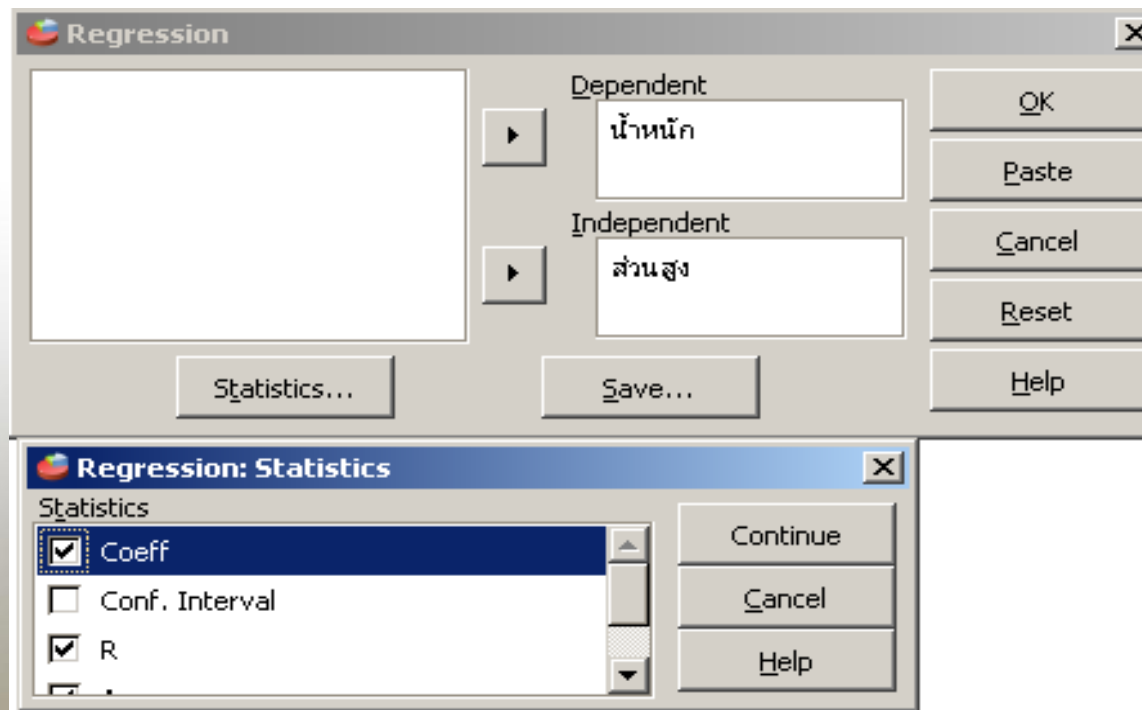
2. Click Analyze → Regression → Linear  
จะได้หน้าจอ ดังรูป



3. นำตัวแปรน้ำหนักใส่ไว้ในช่อง Dependent ส่วนตัวแปรส่วนสูงใส่ไว้ในช่อง Independent จากนั้น Click Statistics...จะได้ดังรูป







4. Click เลือกค่าที่ต้องการคำนวณแต่ในโปรแกรมจะ Click เลือกค่าที่คิดว่าจำเป็นมาให้แล้วคือ Coeff , ค่า R และ ANOVA หากต้องการค่าใดเพิ่มให้คลิกเลือกแต่ในที่นี้ต้องการเพียงเท่านั้น จากนั้น Click Continue แล้ว OK จะได้ผลลัพธ์ดังรูป





# REGRESSION

## REGRESSION

/VARIABLES= ส่วนสูง  
/DEPENDENT= น้ำหนัก  
/STATISTICS=COEFF R ANOVA.

### Model Summary (น้ำหนัก)

R	R Square	Adjusted R Square	Std. Error of the Estimate
.86	.74	.73	8.02

### ANOVA (น้ำหนัก)

	Sum of Squares	df	Mean Square	F	Sig.
Regression	3303.86	1	3303.86	51.40	.000
Residual	1157.09	18	64.28		
Total	4460.95	19			

### Coefficients (น้ำหนัก)

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-46.63	10.69	.00	-4.36	.000
ส่วนสูง	.60	.08	.86	7.17	.000

จากตารางผลลัพธ์สามารถสรุปได้ว่าค่าสัมประสิทธิ์แสดงความสัมพันธ์มีค่าเท่ากับ 0.86 แสดงว่าน้ำหนักและส่วนสูงของนักเรียนตัวอย่างกลุ่มนี้มีความสัมพันธ์กันค่อนข้างมาก และสามารถเขียนสมการถดถอยเป็นดังนี้คือ น้ำหนัก(Y) = -46.63+ 0.60 (ส่วนสูง)







## 8.3 บทสรุป

ในการวิเคราะห์ข้อมูลโดยเฉพาะข้อมูลที่เป็นข้อมูลเชิงปริมาณวิธีการวิเคราะห์ข้อมูลด้วยวิธีการทางสถิติจะค่อนข้างยุ่งยากและซับซ้อน ดังนั้นในการวิเคราะห์ข้อมูลเชิงปริมาณด้วยวิธีการทางสถิติจึงมักนิยมใช้โปรแกรมคอมพิวเตอร์เข้ามาช่วย โดยโปรแกรมสำเร็จรูปที่ใช้ในการวิเคราะห์ข้อมูลทางสถิติจะมีค่อนข้างหลากหลายขึ้นอยู่กับผู้ทำการวิเคราะห์จะเลือกใช้ โดยผลที่ได้จากการวิเคราะห์จะมีลักษณะที่คล้ายกัน ซึ่งโปรแกรมสำเร็จรูปในปัจจุบันส่วนใหญ่มักจะมีลิขสิทธิ์ทำให้มีข้อจำกัดในการเลือกใช้ โปรแกรม PSPP (a program of statistical analysis of sample data) เป็นโปรแกรมโอเพนซอร์ส ที่มีฟังก์ชันและลักษณะการทำงานเหมือนกับโปรแกรม SPSS ดังนั้นผู้ที่เคยใช้โปรแกรม SPSS มาก่อนจะสามารถที่จะใช้โปรแกรมสำเร็จรูปตัวนี้ได้ไม่ยาก แต่อย่างไรก็ตามสิ่งสำคัญที่สุดสำหรับการวิเคราะห์ข้อมูลโดยใช้โปรแกรมสำเร็จรูปทางสถิติไม่ว่าจะเลือกใช้โปรแกรมใดก็คือการแปลผลและสรุปผลที่ได้จากการวิเคราะห์ข้อมูลให้ถูกต้องที่สุด

