การวิเคราะห์ข้อมูลโดยใช้โปรแกรมสำเร็จรูปทางสถิติ

สถิติอนุมาน (Statistical Inference)

เป็นการนำข้อสรุปที่ได้จากการศึกษาข้อมูลตัวอย่างไปอ้างถึงลักษณะของประชากร ได้แก่ การประมาณค่า การทดสอบสมมติฐาน การวิเคราะห์ความแปรปรวน การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย การทดสอบภาวะสารูปสนิทดี เป็นต้น

1. การประมาณค่าและการทดสอบสมมติฐานของค่าเฉลี่ย 1 กลุ่ม

การทดสอบค่าเฉลี่ยของประชากร 1 กลุ่มใช้สำหรับทดสอบว่าค่าเฉลี่ยของประชากรเท่ากับค่าคงที่ที่ต้องการทดสอบหรือไม่ เช่น ต้องการทดสอบว่าน้ำหนักเฉลี่ยของผลไม้กระป๋องเท่ากับ 290 กรัมหรือไม่ หรือต้องการทดสอบว่าจำนวนสินค้าที่ขายได้โดยเฉลี่ยมากว่า 300 ชิ้น หรือไม่ เป็นต้น โดยในการวิเคราะห์จะใช้คำสั่ง One-Sample T Test

ตัวอย่าง 1 จากการสุ่มตรวจสอบอุณหภูมิในสถานที่แห่งหนึ่ง จำนวน 30 วัน (หน่วยเป็นองศาเซลเซียส) ได้ข้อมูลดังนี้

37.7	35.6	37.0	40.0	36.0	38.6	35.5	37.0	38.5	36.6
34.0	36.0	38.6	39.0	35.6	39.7	38.0	39.6	37.7	40.0
34.5	34.0	37.6	38.0	39.6	39.0	35.0	37.7	37.0	36.8

จงทดสอบว่าอุณหภูมิโดยเฉลี่ยของสถานที่แห่งนี้น้อยกว่า 37 องศาเซลเซียสหรือไม่ ที่ระดับนัยสำคัญ 0.05

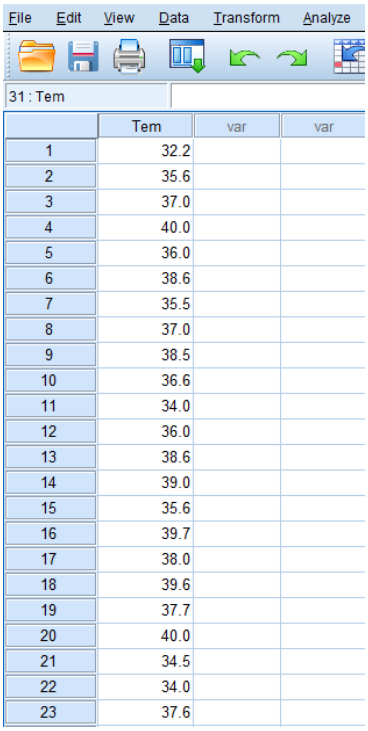
ขั้นตอนการทดสอบ มีดังนี้

1. กำหนดสมมติฐานในการทดสอบ

$H_0 : \mu = 37$

$H_1 : \mu < 37$

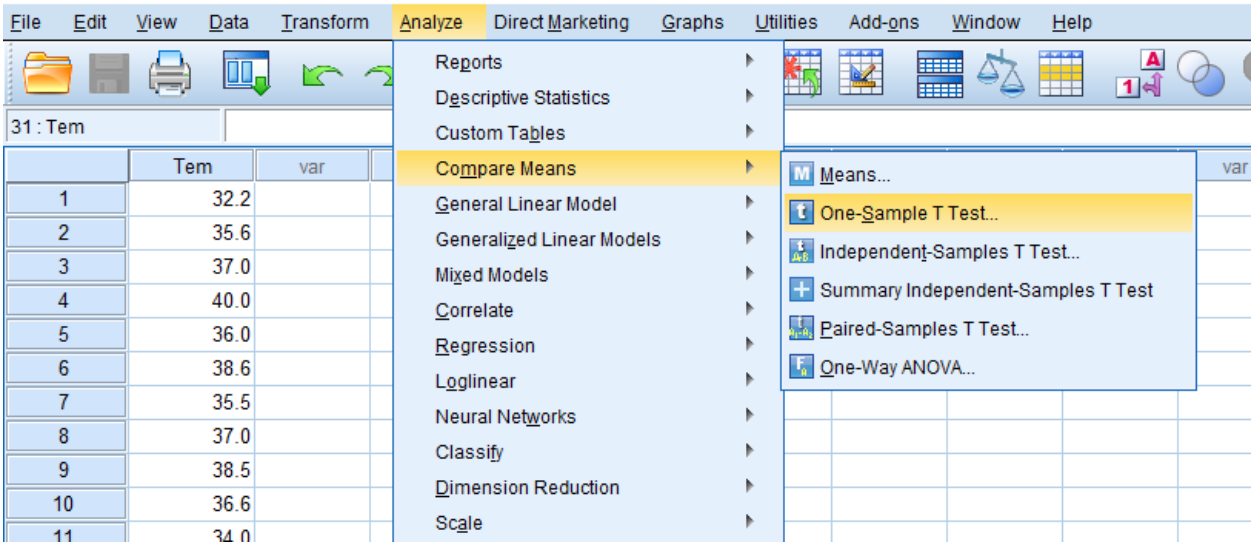
2. สร้างข้อมูลในตารางลงในโปรแกรมสำเร็จรูปทางสถิติ ดังภาพ



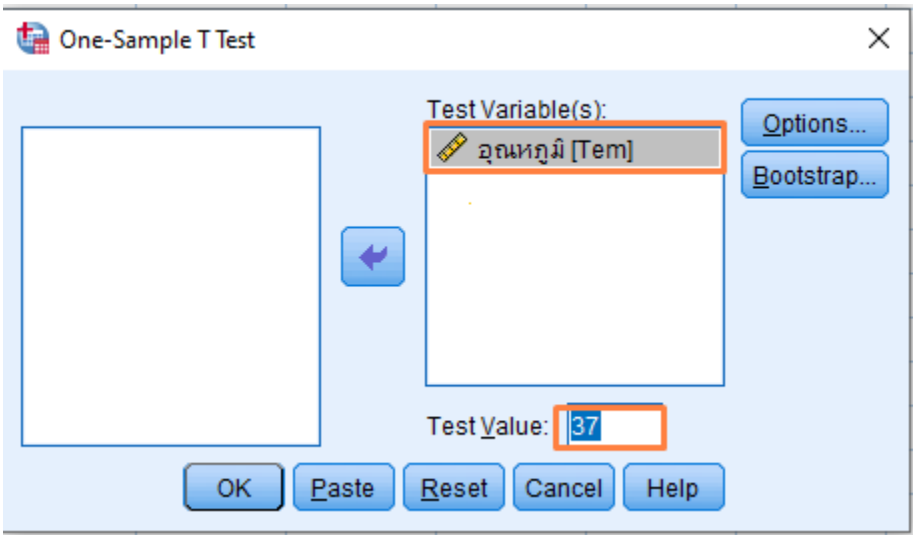
	Tem	var	var
1	32.2		
2	35.6		
3	37.0		
4	40.0		
5	36.0		
6	38.6		
7	35.5		
8	37.0		
9	38.5		
10	36.6		
11	34.0		
12	36.0		
13	38.6		
14	39.0		
15	35.6		
16	39.7		
17	38.0		
18	39.6		
19	37.7		
20	40.0		
21	34.5		
22	34.0		
23	37.6		

3. วิเคราะห์ข้อมูลโดยใช้คำสั่ง One-Sample T Test

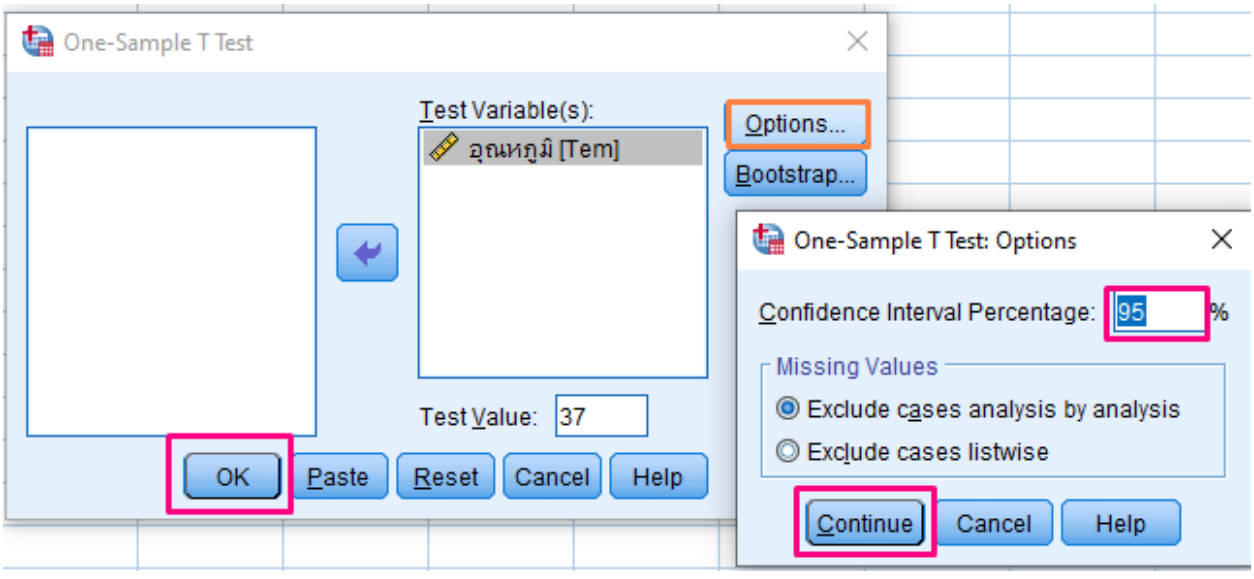
จากเมนูเลือก Analyze → Compare Means → One-Sample T Test...



4. นำตัวแปรอุณหภูมิใส่ไว้ที่ช่อง Test Variable(s): ส่วนในช่อง Test Value: ให้ใส่ค่าที่ต้องการทดสอบซึ่งในที่นี้คือ 37 องศาเซลเซียส ดังภาพ



5. คลิกเลือก Option... โดยในช่อง Confidence Interval ใส่ 95% เนื่องจากโจทย์กำหนดระดับนัยสำคัญเป็น 0.05 จากนั้นคลิก Continue แล้วคลิก OK ดังภาพ



6. ผลลัพธ์ที่ได้จากการวิเคราะห์ เป็นดังนี้

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
อุณหภูมิ	30	37.147	2.0046	.3660

จากตารางผลการวิเคราะห์ สามารถแปลผลได้ดังนี้

มีข้อมูลเกี่ยวกับตรวจสอบอุณหภูมิที่นำมาวิเคราะห์ทั้งหมด 30 ตัวอย่าง โดยมีอุณหภูมิเฉลี่ย 37.147 องศาเซลเซียส มีค่าเบี่ยงเบนมาตรฐานเป็น 2.0046 และมีค่าเบี่ยงเบนของความคลาดเคลื่อนเฉลี่ยเป็น 0.3660

One-Sample Test

	Test Value = 37					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
อุณหภูมิ	.401	29	.692	.1467	-.602	.895

การสรุปผลการทดสอบ

จากตารางการวิเคราะห์จะได้ค่า p-value หรือค่า Sig $= \frac{0.692}{2} = 0.346 > 0.05$ ดังนั้น จะยอมรับ H_0 นั่นคือ อุณหภูมิโดยเฉลี่ยของสถานที่แห่งนี้ไม่น้อยกว่า 37 องศาเซลเซียส ที่ระดับนัยสำคัญ 0.05 และสามารถประมาณค่าเฉลี่ยของอุณหภูมิมีค่าอยู่ระหว่าง -0.602+37 = 36.398 ถึง 0.895+37 = 37.895 องศาเซลเซียส (36.398, 37.895 องศาเซลเซียส)

2. การประมาณค่าและการทดสอบสมมติฐานของค่าเฉลี่ย 2 กลุ่ม แบ่งออกเป็น 2 กรณี คือ

2.1 กรณีที่ประชากร 2 กลุ่มไม่เป็นอิสระต่อกัน (มีความสัมพันธ์กัน)

ในกรณีที่ประชากร 2 กลุ่มไม่เป็นอิสระต่อกัน เช่น การทดสอบความแตกต่างของคะแนนสอบก่อน - หลังเรียน การทดสอบการทำงานของคู่แฝด การทดสอบประสิทธิภาพการใช้งานก่อนและหลังการอบรม เป็นต้น โดยในการวิเคราะห์จะใช้คำสั่ง Paired-Samples T Test

ตัวสถิติทดสอบคือ $t = \frac{\overline{D} - \mu_D}{\frac{S_D}{\sqrt{n}}}$ จะมีการแจกแจงแบบที ที่มีองศาแห่งความเป็นอิสระ $\nu = n - 1$

ตัวอย่าง 2 ในการศึกษาประสิทธิภาพการลดน้ำหนักของโปรแกรมออกกำลังกายโปรแกรมหนึ่ง โดยการสุ่มผู้เข้าร่วมโปรแกรมลดน้ำหนักดังกล่าวมา 10 คน แล้วบันทึกน้ำหนักก่อนและหลังเข้าร่วมโปรแกรมปรากฏผลดังนี้

คนที่	น้ำหนักก่อน (กิโลกรัม)	น้ำหนักหลัง (กิโลกรัม)
1	82	80
2	73	71
3	86	84
4	85	84
5	95	93
6	91	90
7	75	77
8	78	76
9	83	81
10	82	79

จงทดสอบว่าโปรแกรมออกกำลังกายดังกล่าวมีประสิทธิภาพในการลดน้ำหนักหรือไม่ที่ระดับนัยสำคัญ 0.01

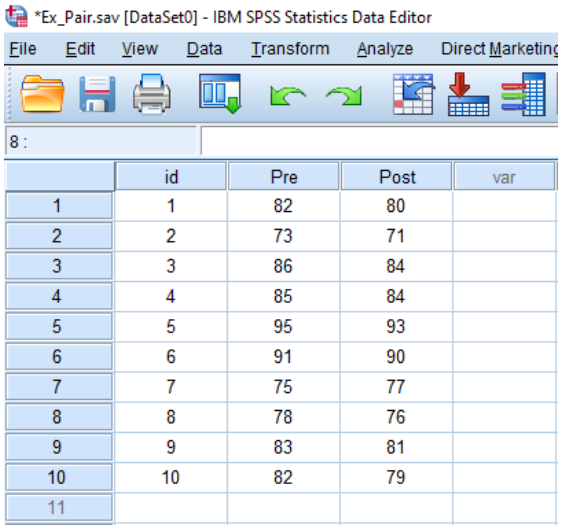
ขั้นตอนการทดสอบ มีดังนี้

1. กำหนดสมมติฐานในการทดสอบ

$$H_0 : \mu_D = 0$$

$$H_1 : \mu_D > 0$$

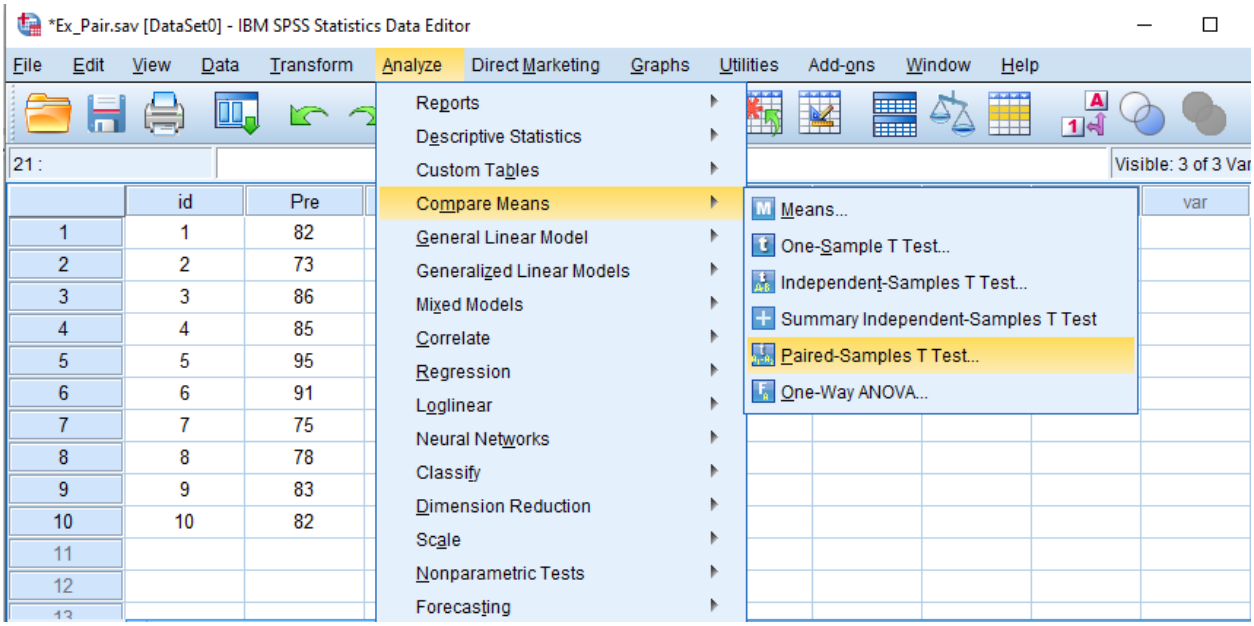
2. สร้างข้อมูลในตารางลงในโปรแกรมสำเร็จรูปทางสถิติ ดังภาพ



	id	Pre	Post	var
1	1	82	80	
2	2	73	71	
3	3	86	84	
4	4	85	84	
5	5	95	93	
6	6	91	90	
7	7	75	77	
8	8	78	76	
9	9	83	81	
10	10	82	79	
11				

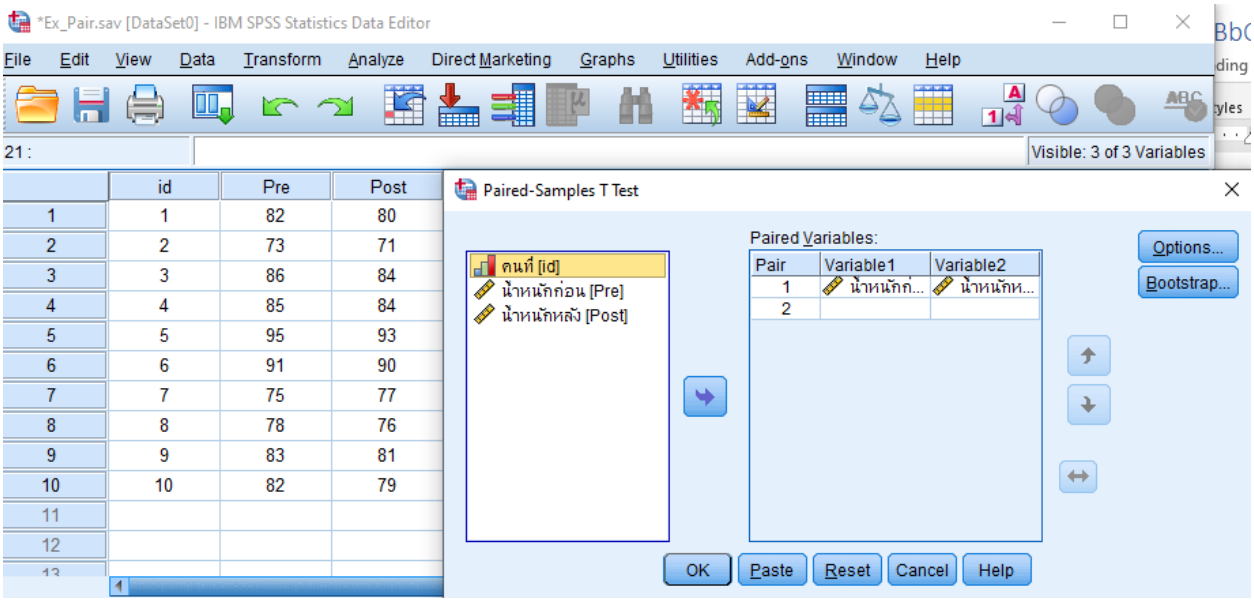
3. วิเคราะห์ข้อมูลโดยใช้คำสั่ง Paired-Samples T Test

จากเมนูเลือก Analyze → Compare Means → Paired-Samples T Test...



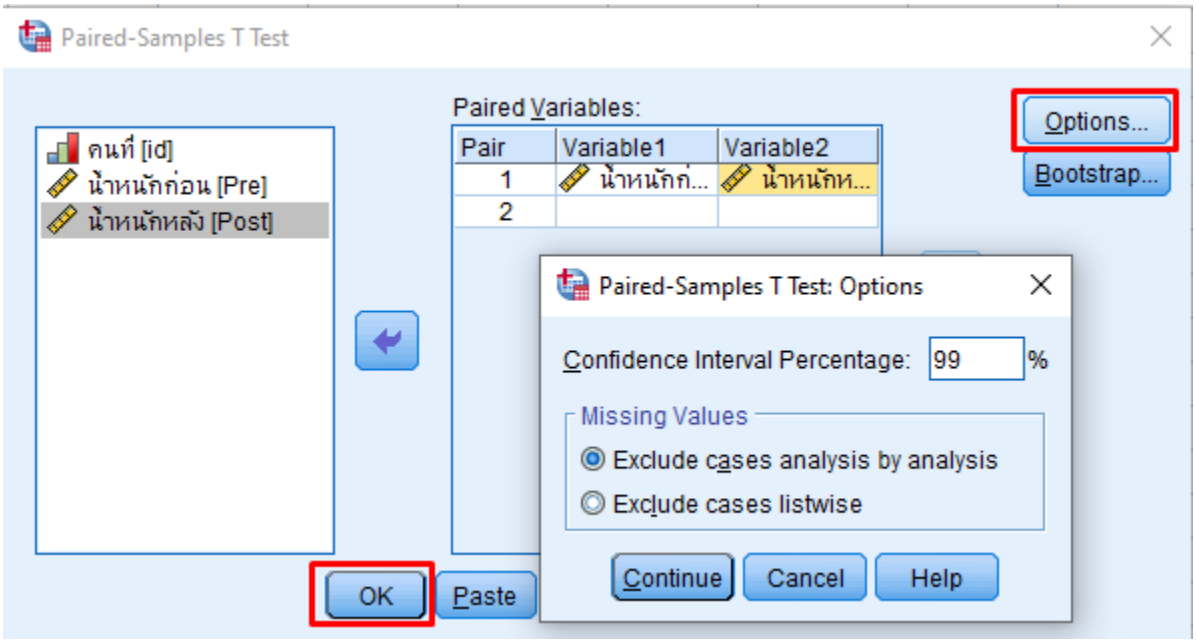
	id	Pre	var
1	1	82	
2	2	73	
3	3	86	
4	4	85	
5	5	95	
6	6	91	
7	7	75	
8	8	78	
9	9	83	
10	10	82	
11			
12			

4. นำตัวแปรน้ำหนักก่อนและน้ำหนักหลังใส่ไว้ที่ช่อง Paired Variables: โดยเลือกน้ำหนักก่อนใส่ในช่อง Variable 1 และน้ำหนักหลังใส่ในช่อง Variable 2 ดังภาพ



	id	Pre	Post
1	1	82	80
2	2	73	71
3	3	86	84
4	4	85	84
5	5	95	93
6	6	91	90
7	7	75	77
8	8	78	76
9	9	83	81
10	10	82	79
11			
12			

5. คลิกเลือก Option... โดยในช่อง Confidence Interval ใส่ 99% เนื่องจากโจทย์กำหนดระดับนัยสำคัญเป็น 0.01 จากนั้นคลิก Continue แล้วคลิก OK ดังภาพ



6. ผลลัพธ์ที่ได้จากการวิเคราะห์ เป็นดังนี้

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	น้ำหนักก่อน	83.00	10	6.766	2.140
	น้ำหนักหลัง	81.50	10	6.553	2.072

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	น้ำหนักก่อน & น้ำหนักหลัง	10	.980	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	99% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	น้ำหนักก่อน - น้ำหนักหลัง	1.500	1.354	.428	.109	2.891	3.503	9	.007

ในการสรุปผลการทดสอบสมมติฐานจะแตกต่างจากการคำนวณมือ โดยให้พิจารณาจาก Output ที่ค่า Sig. ดังนี้

- หากค่า Sig. มีค่าน้อยกว่าระดับนัยสำคัญที่กำหนด แสดงว่า ปฏิเสธสมมติฐานหลัก (H_0) แล้วยอมรับสมมติฐานรอง (H_1)

- หากค่า Sig. มีค่ามากกว่าหรือเท่ากับระดับนัยสำคัญที่กำหนด แสดงว่า ยอมรับสมมติฐานหลัก (H_0)

การสรุปผลการทดสอบ

จากตารางการวิเคราะห์จะได้ค่า p-value หรือค่า Sig $= \frac{0.007}{2} = 0.0035 < 0.01$ ดังนั้น ปฏิเสธ H_0 แล้วยอมรับ H_1 นั่นคือ น้ำหนักหลังเข้าร่วมโปรแกรมออกกำลังกายมีค่าน้อยกว่าน้ำหนักก่อนเข้าร่วมโปรแกรมออกกำลังกาย ทำให้ประสิทธิภาพในการลดน้ำหนักลดลง ที่ระดับนัยสำคัญ 0.01 และผลต่างของน้ำหนักก่อนและหลังการเข้าร่วมโปรแกรมออกกำลังกายมีค่าอยู่ระหว่าง 0.109 ถึง 2.891 ด้วยความเชื่อมั่น 99%

2.2 กรณีที่ประชากร 2 กลุ่มเป็นอิสระต่อกัน (ไม่มีความสัมพันธ์กัน)

เป็นการทดสอบผลต่างระหว่างค่าเฉลี่ยของลักษณะที่สนใจของ 2 ประชากรว่าแตกต่างกันหรือไม่ หรือเป็นการทดสอบว่าค่าเฉลี่ยของประชากรที่ 1 มากกว่าประชากรที่ 2 หรือไม่ เช่น ต้องการทดสอบรายได้ของเพศชายกับเพศหญิงว่าแตกต่างกันหรือไม่ เป็นต้น โดยในการวิเคราะห์จะใช้คำสั่ง Independent-Samples T Test ซึ่งแบ่งการทดสอบออกเป็น 2 ขั้นตอนดังนี้

ขั้นตอนที่ 1 ทดสอบความแปรปรวนของประชากรสองกลุ่ม

ขั้นตอนที่ 2 ทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากรทั้งสองกลุ่ม

ตัวอย่าง 3 จากการสุ่มสอบถามรายได้ของผู้ชายจำนวน 12 คน และรายได้ของผู้หญิงจำนวน 9 คน (หน่วยเป็นบาท) ได้ข้อมูลดังนี้

ผู้ชาย			ผู้หญิง		
27,500	35,000	28,890	24,500	25,000	22,000
20,500	33,000	23,000	29,000	31,000	31,050
25,500	31,000	32,800	32,500	24,050	21,000
22,000	29,000	29,990			

จงทดสอบว่าผู้ชายมีรายได้แตกต่างจากผู้หญิงหรือไม่ ที่ระดับนัยสำคัญ 0.05

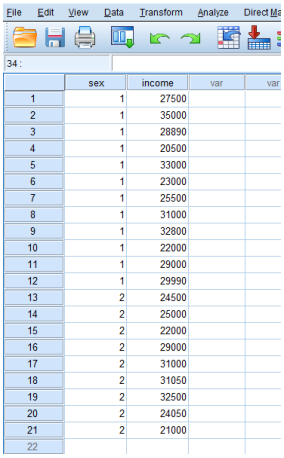
ขั้นตอนการทดสอบ มีดังนี้

1. กำหนดสมมติฐานในการทดสอบ

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

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

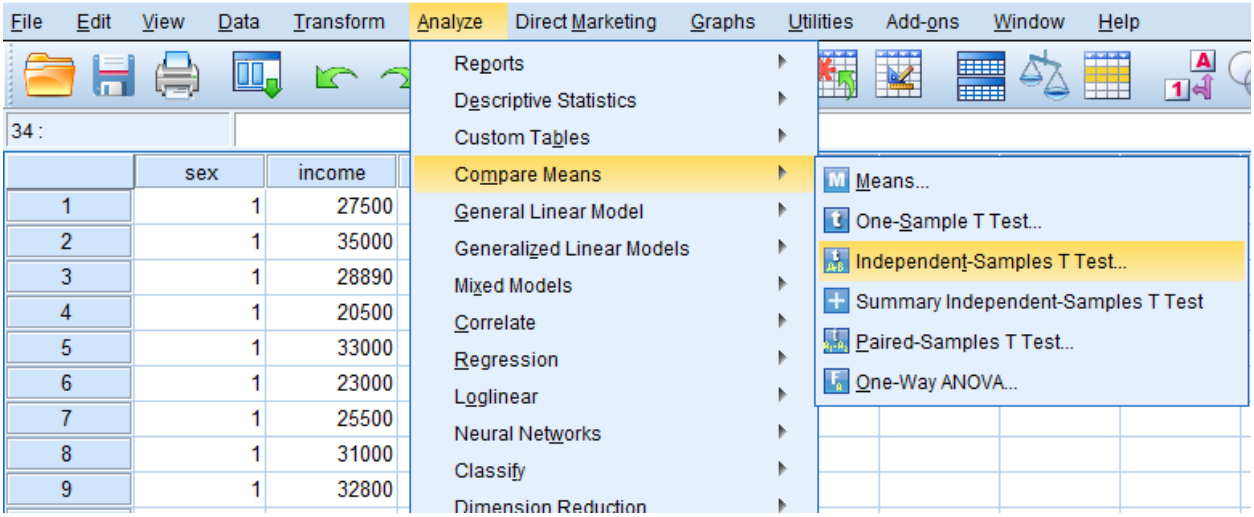
2. สร้างข้อมูลในตารางลงในโปรแกรมสำเร็จรูปทางสถิติ ดังภาพ



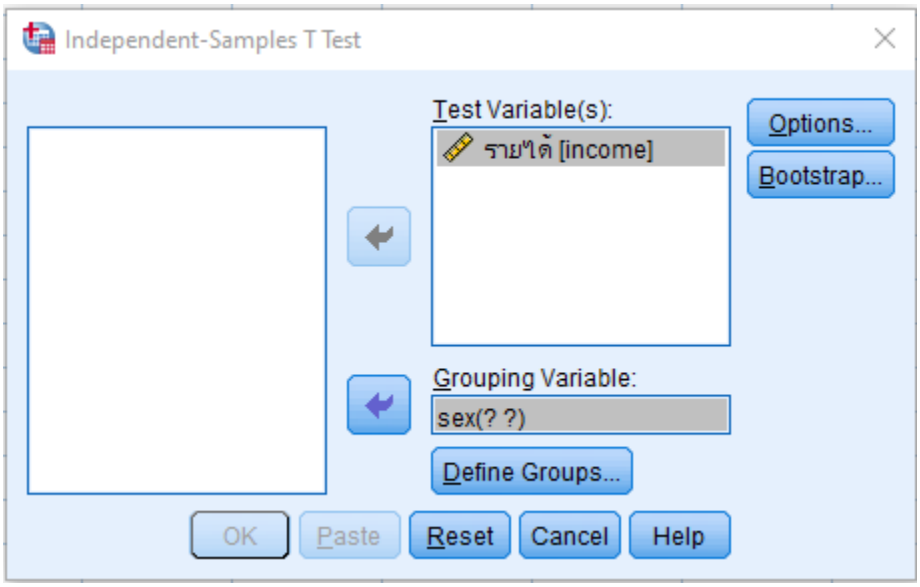
	sex	income	var	var
1	1	27500		
2	1	35000		
3	1	28890		
4	1	20500		
5	1	33000		
6	1	23000		
7	1	25500		
8	1	31000		
9	1	32800		
10	1	22000		
11	1	29000		
12	1	29990		
13	2	24500		
14	2	25000		
15	2	22000		
16	2	29000		
17	2	31000		
18	2	31050		
19	2	32500		
20	2	24050		
21	2	21000		
22				

3. วิเคราะห์ข้อมูลโดยใช้คำสั่ง Independent-Samples T Test

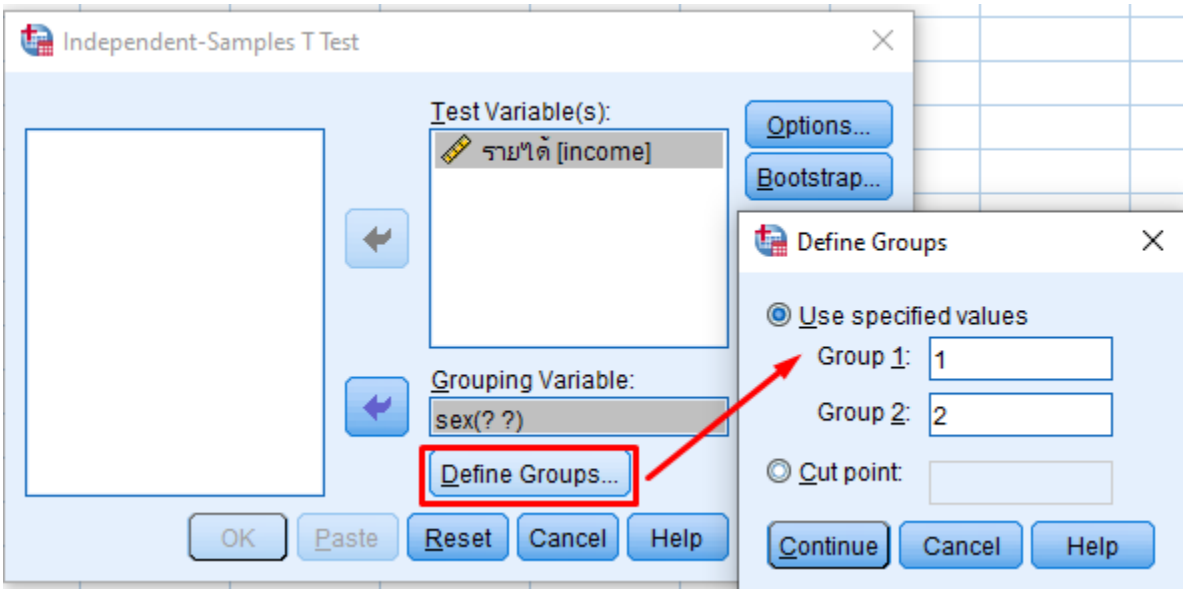
จากเมนูเลือก Analyze → Compare Means → Independent-Samples T Test...



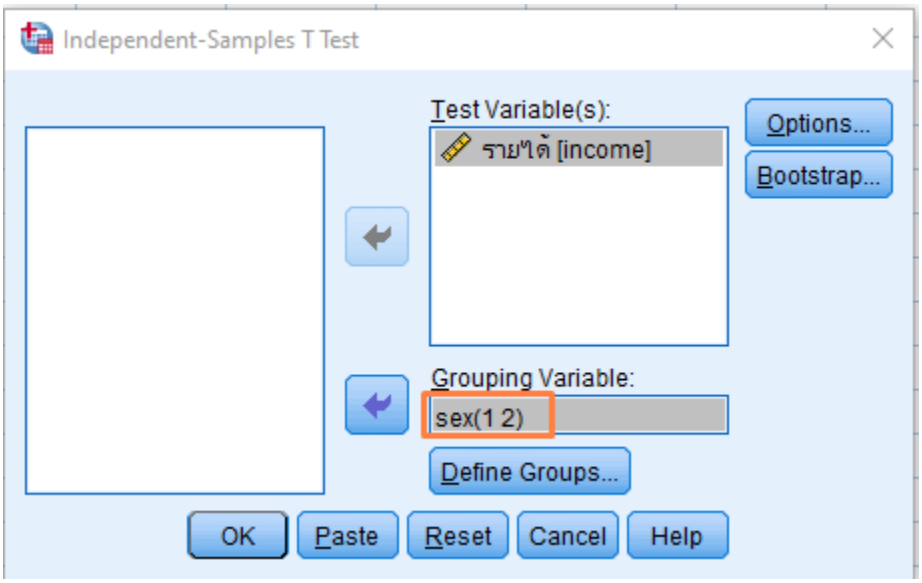
4. นำตัวแปรรายได้ใส่ไว้ที่ช่อง Test Variable(s): และตัวแปรเพศใส่ไว้ที่ช่อง Grouping Variable: ดังภาพ



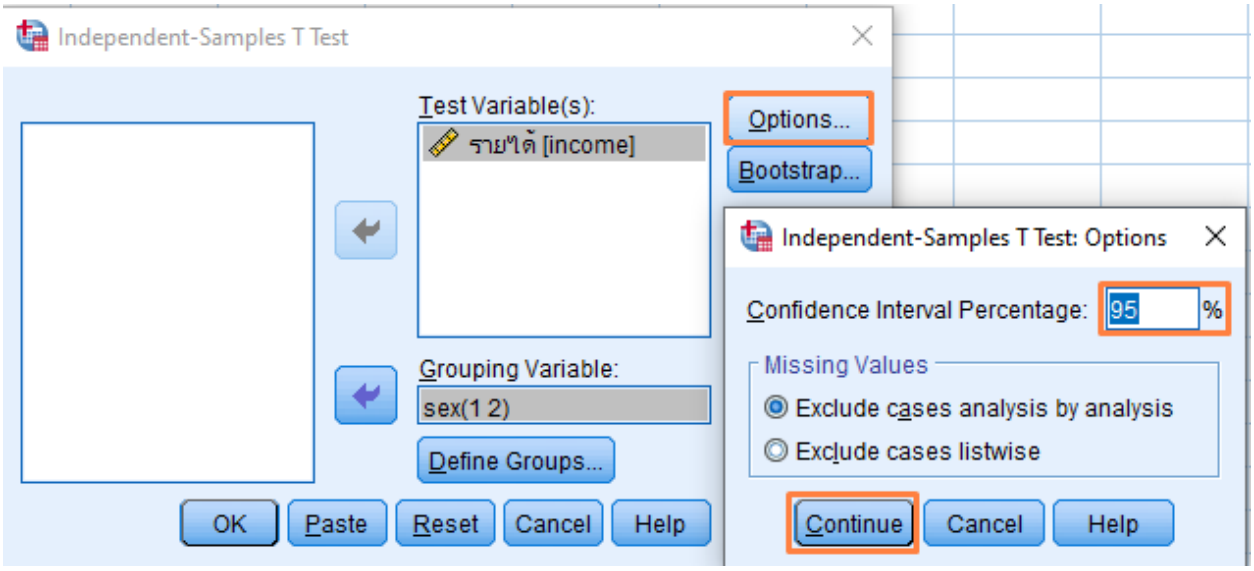
5. จากนั้นคลิก Define Groups... โดยใน Group 1: คือ เพศชาย แทนด้วยหมายเลข 1 และ Group 2: คือ เพศหญิง แทนด้วยหมายเลข 2 ดังภาพ



6. จากนั้นคลิก Continue จะได้ผลลัพธ์ดังภาพ



7. คลิกเลือก Option... โดยในช่อง Confidence Interval ใส่ 95% เนื่องจากโจทย์กำหนดระดับนัยสำคัญเป็น 0.05 จากนั้นคลิก Continue แล้วคลิก OK ดังภาพ



8. ผลลัพธ์ที่ได้จากการวิเคราะห์ เป็นดังนี้

Group Statistics

	เพศ	N	Mean	Std. Deviation	Std. Error Mean
รายได้	ชาย	12	28181.67	4623.519	1334.695
	หญิง	9	26677.78	4267.179	1422.393

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
รายได้	Equal variances assumed	.000	.994	.762	19	.456	1503.889	1974.145	-2628.043	5635.821
	Equal variances not assumed			.771	18.090	.451	1503.889	1950.542	-2592.583	5600.361

การสรุปผลการทดสอบ

ขั้นตอนที่ 1 ทดสอบค่าความแปรปรวนของประชากรทั้ง 2 กลุ่มเท่ากันหรือไม่

สมมติฐานการทดสอบดังนี้

$H_0 : \sigma_1^2 = \sigma_2^2$ (ความมแปรปรวนของรายได้ผู้ชายและผู้หญิงเท่ากัน)

$H_1 : \sigma_1^2 \neq \sigma_2^2$ (ความมแปรปรวนของรายได้ผู้ชายและผู้หญิงไม่เท่ากัน)

สถิติที่ใช้ทดสอบใช้ F โดยดูได้จาก Column ของ Levene's Test for Equality of Variances เนื่องจากค่า Sig. = 0.994 > 0.05 ดังนั้น ยอมรับ H_0 นั่นคือ ความแปรปรวนของรายได้ผู้ชายและผู้หญิงเท่ากัน ที่ระดับนัยสำคัญ 0.05

ขั้นตอนที่ 2 ทดสอบความแตกต่างระหว่างรายได้ผู้ชายและผู้หญิง

สมมติฐานการทดสอบดังนี้

$H_0 : \mu_1 - \mu_2 = 0$

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

เนื่องจากความแปรปรวนของรายได้ผู้ชายและผู้หญิงเท่ากัน ดังนั้นเลือกใช้ค่า Sig. แถว Equal Variances Assumed (แถวบน) นั่นคือ Sig. = 0.456 > 0.05 ดังนั้น ยอมรับ H_0 นั่นคือ รายได้ของผู้ชายและจากผู้หญิงไม่แตกต่างกัน ที่ระดับนัยสำคัญ 0.05 และค่าประมาณแบบช่วงของ $\mu_1 - \mu_2$ ที่ระดับความเชื่อมั่น 95% โดยมีค่าเท่ากับ $2628.43 < |\mu_1 - \mu_2| < 5635.821$ นั่นคือ ผลต่างระหว่างรายได้ผู้ชายและผู้หญิงจะอยู่ในช่วง 2628.43 ถึง 5635.821 บาท

แบบฝึกหัดปฏิบัติ

1. จากการสุ่มตรวจสอบจำนวนสินค้าที่ขายได้ต่อวัน จำนวน 30 วัน (หน่วยเป็นชิ้น) ได้ข้อมูลดังนี้

32	30	33	41	31	38	35	32	38	42
33	36	38	39	35	30	32	38	37	40
31	34	37	33	38	39	31	35	37	41

จงทดสอบว่าจำนวนสินค้าที่ขายได้โดยเฉลี่ยมากกว่า 32 ชิ้นต่อวันหรือไม่ ที่ระดับนัยสำคัญ 0.01

2. ในชั่วโมงเรียนวิชาคอมพิวเตอร์อาจารย์ได้สอนให้นักศึกษาเรียนวิธีการพิมพ์สัมผัส โดยก่อนการสอนอาจารย์ให้นักศึกษาทดลองพิมพ์ตามปกติและทำการจับเวลา หลังจากนั้นอาจารย์สอนวิธีพิมพ์สัมผัสและให้นักศึกษาทดลองพิมพ์โดยใช้เวลาเท่าเดิม ได้ผลดังนี้

จำนวนคำ / คนที่	ก่อนเรียนวิธีการพิมพ์สัมผัส	หลังเรียนวิธีการพิมพ์สัมผัส
1	12	25
2	21	31
3	15	29
4	18	32
5	19	28
6	22	35
7	17	28
8	18	27
9	13	22
10	14	29

จงทดสอบว่าจำนวนคำที่นักศึกษาพิมพ์ได้ก่อนเรียนวิธีการพิมพ์สัมผัสน้อยกว่าจำนวนคำที่นักศึกษาพิมพ์ได้หลังเรียนวิธีการพิมพ์สัมผัสหรือไม่ ที่ระดับนัยสำคัญ 0.05

3. จากการลงพื้นที่สำรวจราคาเครื่องสำอางยี่ห้อหนึ่งในจังหวัดสุรินทร์และศรีสะเกษ (หน่วยเป็นบาท) ได้ข้อมูลดังนี้

ร้านค้าที่	จังหวัดสุรินทร์	จังหวัดศรีสะเกษ
1	370	330
2	325	295
3	319	299
4	299	325
5	349	289
6	285	298
7	299	299
8	325	300
9	329	299
10	315	329

จงทดสอบว่าราคาเครื่องสำอางในในจังหวัดสุรินทร์และศรีสะเกษแตกต่างกันหรือไม่ ที่ระดับนัยสำคัญ 0.05