การทดสอบสมมติฐานโดยใช้ t-test

2. การทดสอบค่าเฉลี่ยของกลุ่มตัวอย่าง 2 กลุ่ม (Two-sample test on means)

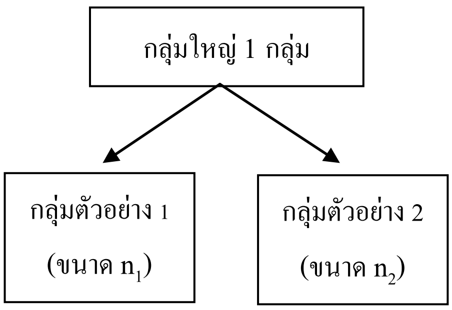
การเปรียบเทียบผลระหว่างกลุ่มตัวอย่าง 2 กลุ่ม ซึ่งข้อมูลที่รวบรวมได้เป็นข้อมูลในมาตราอันตรภาคหรือมาตราอัตราส่วน โดยนำค่าเฉลี่ยของกลุ่มตัวอย่าง ($\bar{X}$) แต่ละกลุ่มมาเปรียบเทียบกันเพื่อสรุปผลว่า ค่าเฉลี่ยของประชากร (μ) สองกลุ่มนั้นแตกต่างกันหรือไม่ ทดสอบความแตกต่างของค่าเฉลี่ยโดยใช้ Z-test เมื่อกลุ่มตัวอย่างแต่ละกลุ่มมีขนาดใหญ่ หรือใช้ t-test เมื่อกลุ่มตัวอย่างแต่ละกลุ่มมีขนาดเล็ก (การวิจัยทางการศึกษา กลุ่มตัวอย่างขนาดใหญ่ คือ $n \geq 100$)

การวิจัยทางการศึกษาที่เป็นการวิจัยเชิงทดลองนั้น กลุ่มตัวอย่างแต่ละกลุ่มมักมีขนาดเล็ก คือ $n < 100$ จึงใช้ t-test for independent samples ในการทดสอบค่าเฉลี่ยของกลุ่มตัวอย่าง 2 กลุ่มที่เป็นอิสระจากกัน และใช้ t-test for dependent samples ในการทดสอบค่าเฉลี่ยของกลุ่มตัวอย่าง 2 กลุ่มที่ไม่เป็นอิสระจากกันหรือกลุ่มตัวอย่างสัมพันธ์กัน

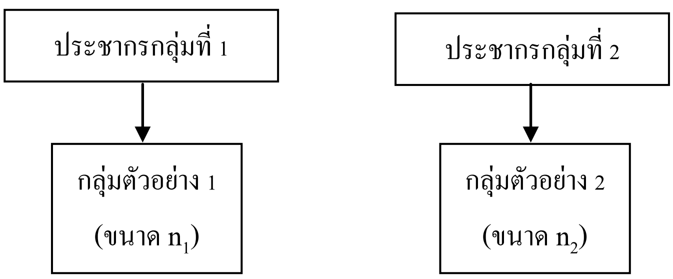
2.1 การทดสอบความแตกต่างระหว่างค่าเฉลี่ยของกลุ่มตัวอย่าง 2 กลุ่มที่เป็นอิสระจากกัน โดยใช้ t-test for independent samples

กลุ่มตัวอย่าง 2 กลุ่ม จะเป็นอิสระจากกัน ถ้าได้มาโดยวิธีต่อไปนี้

วิธีที่ 1 มีกลุ่มที่ต้องการศึกษา (Subjects) 1 กลุ่มใหญ่ แล้วสุ่มแยกเป็น 2 กลุ่มย่อย (Subgroup) เช่น สุ่มแยกเป็นกลุ่มทดลอง 1 กลุ่ม กลุ่มควบคุม 1 กลุ่ม ดังแผนภาพ



วิธีที่ 2 สุ่มกลุ่มตัวอย่าง 2 กลุ่ม มาจากประชากร 2 กลุ่ม ที่เป็นอิสระจากกัน โดยสุ่มกลุ่มตัวอย่างขนาด n_1 จากประชากรกลุ่มที่ 1 และสุ่มกลุ่มตัวอย่างขนาด n_2 จากประชากรกลุ่มที่ 2



วิธีที่ 3 จำแนกกลุ่มตัวอย่างตามตัวแปรอิสระที่ศึกษา เช่น เพศ ชั้นปี ฯลฯ

2.1.1 t-test for independent samples ในกรณีที่ assume ว่า $\sigma^2_1 = \sigma^2_2$

ใช้สูตร
$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$df = n_1 + n_2 - 2$$

เมื่อ $\bar{X}_1$, $\bar{X}_2$ แทนค่าเฉลี่ยของกลุ่มตัวอย่างกลุ่มที่ 1, 2

S_p^2 แทนความแปรปรวนร่วม (Pooled variance)

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

n_1 , n_2 แทนขนาดของกลุ่มตัวอย่างกลุ่มที่ 1, 2

df แทนชั้นแห่งความเป็นอิสระ (degree of freedom)

ข้อตกลงเบื้องต้นของการใช้สูตรนี้

1) กลุ่มตัวอย่างทั้งสองกลุ่มต้องเป็นอิสระจากกัน และต้องได้มาจากการสุ่มจากกลุ่มประชากรที่มีการแจกแจงเป็นโค้งปกติ
2) คุณลักษณะที่ต้องการศึกษาภายในกลุ่มต้องเป็นอิสระจากกัน
3) ความแปรปรวนของกลุ่มประชากรทั้งสองกลุ่มเท่ากัน ($\sigma^2_1 = \sigma^2_2$) แต่ไม่ทราบค่า

หมายเหตุ สูตรนี้เขียนได้อีกลักษณะหนึ่งดังนี้

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$df = n_1 + n_2 - 2$$

2.1.2 t-test for independent samples ในกรณีที่ assume ว่า $\sigma^2_1 \neq \sigma^2_2$ เมื่อมีเหตุผลที่เชื่อได้ว่า $\sigma^2_1 \neq \sigma^2_2$

ใช้สูตร $$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$

เมื่อ $\bar{X}_1$, $\bar{X}_2$ แทนค่าเฉลี่ยของกลุ่มตัวอย่างกลุ่มที่ 1, 2

S_1^2 , S_2^2 แทนความแปรปรวนของกลุ่มตัวอย่างกลุ่มที่ 1, 2

n_1 , n_2 แทนขนาดของกลุ่มตัวอย่างกลุ่มที่ 1, 2

df แทนชั้นแห่งความเป็นอิสระ (degree of freedom)

ข้อตกลงเบื้องต้นของการใช้สูตรนี้

1) กลุ่มตัวอย่างทั้งสองกลุ่มต้องเป็นอิสระจากกัน และต้องได้มาจากการสุ่มจากกลุ่มประชากรที่มีการแจกแจงเป็นโค้งปกติ
2) คุณลักษณะที่ต้องการศึกษาภายในกลุ่มต้องเป็นอิสระจากกัน
3) มีเหตุผลพอที่จะเชื่อได้ว่าความแปรปรวนของกลุ่มประชากรทั้งสองกลุ่มไม่เท่ากัน ($\sigma^2_1 \neq \sigma^2_2$)
4) ขนาดของกลุ่มตัวอย่างไม่เท่ากัน ($n_1 \neq n_2$)

หมายเหตุ

ในกรณีที่ไม่สามารถตัดสินใจได้ว่า $\sigma^2_1 = \sigma^2_2$ หรือ $\sigma^2_1 \neq \sigma^2_2$ ควรต้องมีการทดสอบความแปรปรวนก่อน ด้วยการทดสอบค่าเอฟ หรือ **F-test**

ในการทดสอบความแปรปรวน ว่า $\sigma^2_1 = \sigma^2_2$ หรือไม่ ตั้ง H_0 และ H_1 ดังนี้

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

สูตรการทดสอบค่าเอฟคือ $F = \frac{S_1^2}{S_2^2}$ เมื่อ $S_1^2 > S_2^2$ โดยมี $df_1 = n_1-1$, $df_2 = n_2-1$

หรือ $F = \frac{S_2^2}{S_1^2}$ เมื่อ $S_2^2 > S_1^2$ โดยมี $df_1 = n_2-1$, $df_2 = n_1-1$

นั่นคือ เอาความแปรปรวนที่มีค่ามากกว่าเป็นตัวเศษ และให้ df ของตัวเศษเป็น df_1 df ของตัวส่วน เป็น df_2 เสมอ

เช่น กลุ่มตัวอย่างที่ 1 มี $n_1= 30$ กลุ่มตัวอย่างที่ 2 มี $n_2= 25$

1) ถ้าคำนวณ S^2 ของกลุ่มตัวอย่างที่ 1 ได้ค่ามากกว่า S^2 ของกลุ่มตัวอย่างที่ 2 ($S^2_1 > S^2_2$) ให้เอา S^2_1 เป็นตัวเศษ S^2_2 เป็นตัวส่วน ตัวเศษมี $n_1= 30$ ดังนั้น $df_1= 30-1=29$ ตัวส่วนมี $n_2= 25$ ดังนั้น $df_2= 25-1=24$

2) ถ้าคำนวณ S^2 ของกลุ่มตัวอย่างที่ 2 ได้ค่ามากกว่า S^2 ของกลุ่มตัวอย่างที่ 1 ($S^2_2 > S^2_1$) ให้เอา S^2_2 เป็นตัวเศษ S^2_1 เป็นตัวส่วน ตัวเศษมี $n_2= 25$ ดังนั้น $df_1= 25-1=24$ ตัวส่วนมี $n_1= 30$ ดังนั้น $df_2= 30-1=29$

3) หาค่าวิกฤตจากตารางค่าวิกฤต F (Critical values of F) ซึ่งมีแนบท้ายเล่มตำราสถิติเพื่อการวิจัย จากตัวอย่างข้างต้น ถ้า $S^2_1 > S^2_2$ จะได้ $df_1= 30-1=29$ $df_2 = 25-1=24$ แต่ $df_1= 29$ ไม่ได้แสดงไว้ในตาราง จึงเลือก df_1 ที่ใกล้เคียงที่สุดคือ 30 ที่ $df_1=30$ พบ $df_2 = 24$ นั้นแสดงค่าวิกฤต F = 1.94 ที่ $\alpha = 0.05$ และค่าวิกฤต F= 2.58 ที่ $\alpha = 0.01$ การทดสอบนี้มุ่งที่จะแสดงว่า $\sigma^2_1 = \sigma^2_2$ จึงควรเลือกค่าวิกฤตที่ทำให้ยอมรับ H_0 ได้ยากกว่า เพื่อลดโอกาสที่จะเกิดความคลาดเคลื่อนในการยอมรับ H_0 ดังนั้นจึงเลือกค่าวิกฤต F = 1.94

df_1 (degrees of freedom ของตัวเศษ)

df_2 (degrees of freedom ของตัวส่วน) — DEGREES OF FREEDOM FOR DENOMINATOR	DEGREES OF FREEDOM FOR NUMERATOR: 1	…	24	30	…	500	∞
1	161		249	250		254	254
	4052		6234	6258		6361	6366
. . .							
24	4.26		1.98	1.94		1.74	1.73
	7.82		2.66	2.58		2.23	2.21
. . .							
∞	3.84		1.52	1.46		1.11	1.00
	6.64		1.79	1.69		1.15	1.00

ในแถวคู่แต่ละคู่ แถวบนเป็นค่าวิกฤต F ที่ $\alpha = 0.05$ แถวล่างเป็นค่าวิกฤต F ที่ $\alpha = 0.01$

4) นำค่า F ที่คำนวณได้จากสูตรไปเปรียบเทียบกับ ค่าวิกฤต F ที่ได้จากตาราง Critical values of F

(1) ถ้า F ที่ได้จากการคำนวณ $\geq$ ค่าวิกฤต F ที่ได้จากตาราง แสดงว่าความแปรปรวนของประชากรสองกลุ่มแตกต่างกัน $(\sigma_1^2 \neq \sigma_2^2)$ จะต้องใช้สูตรการทดสอบค่าที (t-test) ในข้อ 2.1.2 ซึ่งเรียกว่า Nonpooled t-test หรือ Separate variance t-test

(2) ถ้า F ที่ได้จากการคำนวณ $<$ ค่าวิกฤต F ที่ได้จากตาราง แสดงว่าความแปรปรวนของประชากรสองกลุ่มเท่ากัน $(\sigma_1^2 = \sigma_2^2)$ จะต้องใช้สูตรการทดสอบค่าที (t-test) ในข้อ 2.1.1 ซึ่งเรียกว่า Pooled variance t-test หรือ Pooled t-test

ที่มา : http://www.krupai.net/stat/stat.htm