2.1. توزيع المعاينة للفرق بين الوسطين الحسابيين لعينتين

يمكن التمييز بين حالتين:

1.2.1.توزيع الفرق بين متوسطي عينتين مستقلتين

التوزيعات الطبيعية المستقلة المعلومة التباين

إذا كانت $X_1, X_2, \dots, X_{n_1}$ عينة من توزيع وسطه μ_1 وتباينه σ_1^2، و $Y_1, Y_2, \dots, Y_{n_2}$ عينة أخرى من توزيع آخر مستقل عن الأول متوسطه μ_2 وتباينه σ_2^2 حيث:

$$V(\bar{X}) = \frac{\sigma_1^2}{n_1}$$

$$V(\bar{Y}) = \frac{\sigma_2^2}{n_2}$$

من الاستقلالية:

$$V(\bar{X} - \bar{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

ومنه:

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

التوزيعات الطبيعية المستقلة المجهولة التباين

حالة العينات الكبيرة

إذا كانت $X_1, X_2, \dots, X_{n_1}$ عينة من توزيع وسطه μ_1 وتباينه σ_1^2 و $Y_1, Y_2, \dots, Y_{n_2}$ عينة أخرى من توزيع آخر مستقل عن الأول متوسطه μ_2 وتباينه σ_1^2 ، وكان σ_2^2 و σ_2^2 مجهولين يمكن استبدالهما بتقديريهما غير المتحيزين s_2^2 و s_1^2 ، فإذا كان حجم العينتين يفوق أو يساوي 30 فإن توزيع $(\bar{X} - \bar{Y})$ هو:

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right)$$

حالة العينات الصغيرة

إذا كانت $X_1, X_2, \dots, X_{n_1}$ عينة من توزيع وسطه μ_1 وتباينه σ_1^2، و $Y_1, Y_2, \dots, Y_{n_2}$ عينة أخرى من توزيع آخر مستقل عن الأول متوسطه μ_2 وتباينه σ_1^2 ، وكان σ_2^2 و σ_2^2 مجهولين وحجم العينتين $n_2, n_1 < 30$ فإن توزيع $(\bar{X} - \bar{Y})$ هو:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

حيث:

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

أمثلة:

المثال الأول:	المثال الثاني:
سحبت عينة من كلية الاقتصاد حجمها 36 طالبا تباين معدلات توجيههم هو 4 وعينة أخرى من كلية الآداب حجمها 42 طالبا تباين معدلات توجيههم هو 3.5، فإذا كان المتوسطان الحقيقيان للمعدلات في كل من الكليتين متساويين أحسب احتمال أن يكون الفرق بين الوسطيين الحسابيين للعينتين مساويا لـ 1.5.	مصنع ينتج 700 كغ من العجائن الغذائية كمعدل يومي، سحبت منه عينة عشوائية تمثل إنتاج 20 يوما فبلغ انحرافها المعياري 40 كغ، في حين ينتج مصنع آخر 500كغ كمعدل يومي، سحبت منه عينة عشوائية تمثل إنتاج 15 يوماً فبلغ انحرافها المعياري 20 كغ. المطلوب حساب الاحتمال التالي: $P(\bar{X} - \bar{Y} \leq 210)$
الحل: σ_1^2 و σ_2^2 مجهولان و $n_1, n_2 \geq 30$ ومنه يمكن استبدالهما بتقديريهما غير المتحيزين s_1^2 و s_2^2 ، فينتج لدينا: $Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \backsim N(0,1)$ والاحتمال المطلوب هو: $P(\bar{X} - \bar{Y} = 1.5) = P\left(\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{1.5 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}\right)$ $= P(Z = 3.41) = 0.9997$	σ_1^2 و σ_2^2 مجهولان و $n_1, n_2 < 30$ ومنه: $T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \backsim t_{n_1+n_2-2}$ والاحتمال المطلوب هو: $P(\bar{X} - \bar{Y} \leq 210)$ $= P\left(\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq \frac{210 - (\mu_1 - \mu_2)}{S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right)$ $= P\left(T \leq \frac{210 - (700 - 500)}{27.41\sqrt{\frac{1}{20} + \frac{1}{15}}}\right)$ $= P(T \leq 1.07) = 0.850$

2.2.1. توزيع الفرق بين متوسطي عينتين غير مستقلتين

أحيانا تكون العينتان اللتان يتم سحبهما مرتبطتين، فمثلا لقياس فاعلية برنامج تدريبي على عينة من الطلبة يتم قياس مستواهم قبل التعرض للبرنامج وبعده، فتشكل القراءات قبل وبعد التعرض للبرنامج عينتين مرتبطتين، ولكي نجد توزيع الفرق بين متوسطي المجتمعين الذين سحبنا منهما العينتين نلجأ لتحويل المسألة لإيجاد توزيع متوسط واحد.

فإذا كانت $X_1, X_2, \dots, X_{n_1}$ عينة من توزيع وسطه μ_1 وتباينه σ_1^2، و $Y_1, Y_2, \dots, Y_{n_2}$ عينة أخرى من توزيع آخر مستقل عن الأول متوسطه μ_2 وتباينه σ_2^2، و X_i و Y_i القيمتين المتناظرتين في العينتين، نضع $d_i = x_i - y_i$ وعليه تشكل $D_1, D_2, \dots, D_n$ عينة من مجتمع طبيعي وسطه $\mu_D = \mu_1 - \mu_2$ وتباينه σ_D^2 ونفرض أن متوسط العينة $\overline{D}$ وتباينها S_D^2.

وفي هذه الحالة يمكن التمييز بين حالتين:

- حجم العينة $n \geq 30$:

$$\overline{D} \backsim N\left(\mu_D, \frac{\sigma_D}{\sqrt{n}}\right)$$

وبمعلومية تباين عينة الفروق ووسطها كمقدرين لوسط مجتمع الفروق وتباينه يكون:

$$Z = \frac{\overline{D} - \mu_D}{\frac{\sigma_D}{\sqrt{n}}} \backsim N(0,1)$$

- حجم العينة $n < 30$:

$$T = \frac{\overline{D} - \mu_D}{\frac{S_D}{\sqrt{n}}} \backsim t_{n-1}$$

مثال:

في تجربة لبيان تحسن أداء العمال، تم سحب عينة عشوائية بحجم 16 عاملا في المصنع فكان قياس الكفاءة قبل وبعد دخولهم دورة تحسين الأداء كما هو موضح في الجدول التالي:

x_i	7	8	7	7	8	8	9	9	9	8	7	7	7	8	8	8
y_i	5	5	5	4	7	8	5	6	7	8	5	6	5	7	7	5

المطلوب: أحسب احتمال أن الفرق في الأداء قبل وبعد الدورة لا يقل عن 2.5.

الحل:

نحسب الفروق $d_i = x_i - y_i$ كما هو موضح في الجدول الموالي:

d_i	2	3	2	3	1	0	4	3	2	0	2	1	2	1	1	3

لدينا :

حجم العينة $n < 30$ ومنه:

$$T = \frac{\overline{D} - \mu_D}{\frac{S_D}{\sqrt{n}}} \sim t_{n-1}$$

ويكون الاحتمال المطلوب حسابه كما يلي:

$$P(\overline{D} > 2.5) = P\left(\frac{\overline{D} - \mu_D}{\frac{S_D}{\sqrt{n}}} > \frac{2.5 - \mu_D}{\frac{S_D}{\sqrt{n}}}\right)$$
$$= P(T > 2.71) = 1 - P(T \leq 2.71)$$
$$= 1 - 0.99 = 0.01$$