المحور3: اختبار الفرضيات الاحصائية

المحاضرة 8- اختبار الفرضيات الإحصائية- الاختبارات المعلمية

إن اختبار الفرضيات الإحصائية يعتبر أهم جزء في نظرية اتخاذ القرار، نحدد بدقة في هذه المحاضرة ماذا نعني بالفرضية الإحصائية والاختبارات المتعلقة بها.

أولا: المفاهيم المتعلقة باختبار الفرضيات

1-1- الفرضية الإحصائية:

الفرضية الإحصائية هي مقولة أو إدعاء تتعلق بالمجتمع الإحصائي تحتمل الصحة أو الخطأ.

يتمثل اختبار الفرضيات بصياغة فرضية إحصائية ما، ونحن نهدف إلى رفضها أو إلغائها، هذه الفرضية تسمى بفرضية العدم (الفرضية الصفرية) ويرمز لها بالرمز H_0 ثم نقوم بصياغة فرضية جديدة أخرى تسمى بالفرضية البديلة ويرمز لها بالرمز H_1.

وتوجد ثلاث (03) صيغ ممكنة لاختبار الفرضيات وتكتب على الشكل التالي:

$$\begin{cases} H_0: \theta \leq \theta_0 \\ H_1: \theta > \theta_0 \end{cases} , \quad \begin{cases} H_0: \theta \geq \theta_0 \\ H_1: \theta < \theta_0 \end{cases} , \quad \begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases}$$

وفي حالة المتوسط نعوض المعلمة θ النسب نعوّض μ بـ P و μ_0 بـ P_0.

وتشير كل صورة للفرضية البديلة إلى جهة الاختبار ما إذا كان من جهة واحدة أو من جهتين. فالفرضية $H_1: \mu \neq \mu_0$ تعني أنّ متوسط المجتمع لا يساوي μ_0 وإنما أكبر منها أو أصغر منها وهي تعبّر عن أنّ شكل الاختبار من جهتين، بينما الفرضية $H_1: \mu < \mu_0$ تعني أنّ متوسط المجتمع أصغر من μ_0 وهي تعبّر عن أنّ شكل الاختبار من جهة اليسار، في حين أنّ الفرضية $H_1: \mu > \mu_0$ تعني أنّ متوسط المجتمع أكبر من μ_0 بمعنى أنّ شكل الاختبار من جهة اليمين.

نشير في هذا الصدد إلى أنّ المساواة تظهر فقط في فرضية العدم، كما أنّ الشيء الذي نريد تصديقه يتم وضعه في الفرضية البديلة H_1.

2-1- الخطأ من النوع الأول والخطأ من النوع الثاني

قد يحدث أحيانا أن نرفض الفرضية الصفرية وهي صحيحة أو نقبلها وهي خاطئة، وبذلك نكون أمام نوعين من الأخطاء:

- **الخطأ من النوع الأول:** وهو الخطأ الناتج عن رفض الفرضية H_0 وهي في الأصل صحيحة ونرمز له بالرمز α. وتسمى α بمستوى المعنوية (وهي عبارة عن احتمال)، ويحدد هذا المستوى من قبل الباحث حسب الدراسة وحسب الدقة المطلوبة. وتؤخذ قيمة α في الحسابات كاملة إذا كان الاختبار في اتجاه واحد سواء من جهة اليمين أو من جهة اليسار وتقسّم لنصفين إذا كان شكل الاختبار في اتجاهين. وتحسب قيمة α من العلاقة:

$$\alpha = P(H_0 \text{ رفض} / H_1 \text{صحيحة})$$

وتقرأ: احتمال رفض الفرضية H_0 علما أنّ H_1 صحيحة.

- **الخطأ من النوع الثاني:** وهو الخطأ الناتج عن قبول الفرضية $\mathbf{H_0}$ وهي في الأصل خاطئة ونرمز له بالرمز $\boldsymbol{\beta}$. وتحسب قيمة $\boldsymbol{\beta}$ بالعلاقة التالية:

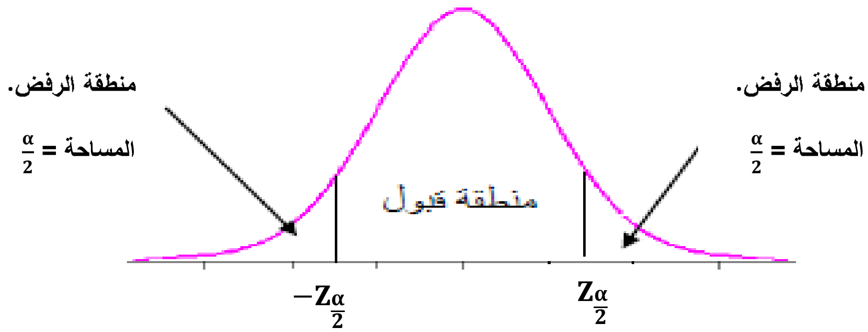
$$\boldsymbol{\beta} = \boldsymbol{P}(\boldsymbol{H_0}\ \text{قبول}\ / \mathbf{H_0}\text{خاطئة})$$

بطريقة مكافئة:

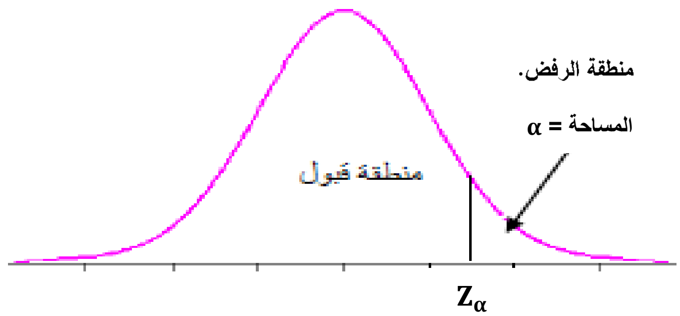
$$\boldsymbol{\beta} = \boldsymbol{P}(\boldsymbol{H_0}\ \text{قبول}\ / \mathbf{H_1}\text{صحيحة})$$

1-3- المنطقة الحرجة والقيمة الحرجة المعيارية:

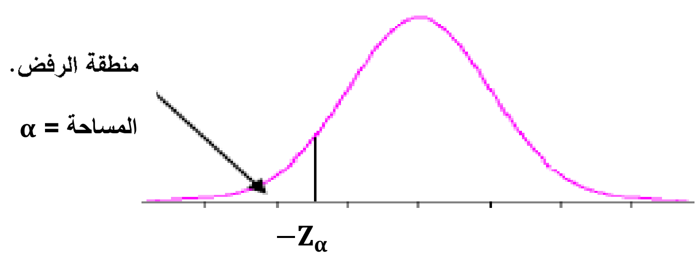
تعرّف **المنطقة الحرجة** بأنها المساحة التي تقع أسفل منحنى التوزيع الاحتمالي المستخدم في عملية التحليل الإحصائي، وهي تعني احتمال رفض H_0 وهي في الأصل صحيحة، وتسمى أيضا **بمنطقة الرفض** وتحدد أيضا حسب نوع الفرضية البديلة ويحدد قيمتها مستوى المعنوية $\boldsymbol{\alpha}$. بينما باقي المساحة أسفل المنحنى فتسمى بمنطقة القبول وقيم التوزيع الاحتمالي التي تفصل بين المنطقتين بالقيم الحرجة المعيارية، والرسم التالي يوضح هذه المفاهيم.



الرسم يوضح مناطق القبول والرفض في حالة $H_1: \mu \neq \mu_0$



الرسم يوضح مناطق القبول والرفض في حالة $H_1: \mu > \mu_0$



الرسم يوضح مناطق القبول والرفض في حالة $H_1: \mu < \mu_0$

ويتم رفض أو قبول H_0 حسب شكل الاختبار، **والقرار** يمكن أن يتخذ بطريقتين مختلفتين كما يلي:

- **الطريقة الأولى: مقارنة القيمة المحسوبة للإحصائية مع القيمة الجدولية لها:**

1- إذا كان الاختبار من جهتين أي $H_1: \mu \neq \mu_0$، نرفض H_0 إذا كانت القيمة المحسوبة للاختبار تقع في منطقة الرفض، أي تكون قيمة إحصائية الاختبار أكبر من القيمة الجدولية الموجبة أو أصغر من القيمة الجدولية السالبة.

2- إذا كان الاختبار من جهة واحدة نحو اليمين أي $H_1: \mu > \mu_0$ نرفض H_0 إذا كانت قيمة إحصائية الاختبار أكبر من القيمة الجدولية.

3- إذا كان الاختبار من جهة واحدة نحو اليسار أي $\mathbf{H_1: \mu < \mu_0}$ نرفض H_0 إذا كانت قيمة إحصائية الاختبار أصغر من القيمة الجدولية.

- **الطريقة الثانية: باستخدام المعنوية المحسوبة p- value**

المعنوية المحسوبة p- value عبارة عن احتمال يحسب لقيمة الإحصائية، ويؤثر في حسابها التوزيع المستخدم في الاختبار وصيغة الفرضية البديلة وذلك كما يلي:

1- إذا كان الاختبار من جهتين أي $\mathbf{H_1: \mu \neq \mu_0}$ فإن:

$$p - value = 2.P(\text{المتغير العشوائي} > \text{قيمة الاحصائية إذا كانت موجبة})$$

أو

$$p - value = 2.P(\text{المتغير العشوائي} < \text{قيمة الاحصائية إذا كانت سالبة})$$

والقرار: نرفض H_0 إذا كانت **p- value** أقل من مستوى المعنوية ونقبلها إذا كانت **p- value** أكبر من أو تساوي مستوى المعنوية.

2- إذا كان الاختبار من جهة واحدة من جهة اليمين فإن:

$$p - value = P(\text{المتغير العشوائي} > \text{قيمة الاحصائية})$$

والقرار: نرفض H_0 إذا كانت **p- value** أقل من مستوى المعنوية ونقبلها إذا كانت **p- value** أكبر من أو تساوي مستوى المعنوية.

3- إذا كان الاختبار من جهة واحدة من جهة اليسار فإن:

$$p - value = P(\text{المتغير العشوائي} < \text{قيمة الاحصائية})$$

والقرار: نرفض H_0 إذا كانت **p- value** أقل من مستوى المعنوية ونقبلها إذا كانت **p- value** أكبر من أو تساوي مستوى المعنوية.

وتعتبر الطريقة الثانية عملية أكثر من الطريقة الأولى، حيث أنه في كل الحالات يتم قبول الفرضية الصفرية H_0 إذا كانت $p - value \geq \alpha$ وترفض الفرضية الصفرية إذا كانت $p - value < \alpha$.

4-1- خطوات اختبارات الفرضيات الاحصائية

توجد أربعة (04) خطوات لاختبار أية فرضية احصائية وهي:

- تحديد الفرضية الصفرية والفرضية البديلة لها.
- تحديد احصائية الاختبار وحساب قيمتها.
- استخراج القيمة أو القيم الجدولية لاحصائية الاختبار.
- مقارنة قيمة الاحصائية المحسوبة مع القيمة أو القيم الجدولية لها واتخاذ القرار.

ثانيا: اختبار الفرضيات الإحصائية

2-1- اختبار الفرضيات للمتوسط الحسابي للمجتمع μ

نميز بين حالتين:

2-1-1- عندما يكون تباين المجتمع σ^2 معلوما

نستخدم في هذه الحالة التوزيع الطبيعي، وتكون إحصائية الاختبار هي:

$$Z = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} \sim N(0,1)$$

مثال

من مجتمع يتوزع توزيعا طبيعيا بتباين يساوي **49**، تم سحب عينة عشوائية حجمها **100** فوجد أن متوسطها يساوي **80**. المطلوب، عند مستوى معنوية **5%** اختبر الفرضيات التالية:

- اختبر الفرضية القائلة بأن متوسط المجتمع يساوي **78**.
- اختبر الفرضية القائلة بأن متوسط المجتمع يزيد عن **78**.
- اختبر الفرضية القائلة بأن متوسط المجتمع يقل عن **78**.

الحل:

- **بالنسبة للحالة الأولى:**

__شكل الاختبار__ هو:

$$\begin{cases} H_0: \mu = 78 \\ H_1: \mu \neq 78 \end{cases}$$

__القيم الحرجة (الجدولية)__ سوف تكون:

$$-Z_{\frac{\alpha}{2}} = -1.96 \qquad\qquad Z_{\frac{\alpha}{2}} = 1.96$$

أما __قيمة إحصائية الاختبار__ تكون مساوية إلى:

$$Z = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} = \frac{80 - 78}{\sqrt{49/100}} = 2.85$$

__**والقرار**__: يمكن أن نتخذه بطريقتين مختلفتين:

➢ الطريقة الأولى:

قيمة Z المحسوبة أكبر من القيم الجدولية، ومنه نرفض الفرضية H_0 (ونقبل الفرضية H_1) وبالتالي متوسط المجتمع لا يساوي 78.

➢ الطريقة الثانية:

الاختبار من جهتين أي $H_1: \mu \neq \mu_0$ وبالتالي:

$$\boldsymbol{p-value = 2.P(Z > 2.85)} = 2[1 - P(Z \leq 2.85)] = 2\,[1 - (0.997814] = \mathbf{0.0043}$$

والقرار: نرفض H_0 لأنّ **p- value** أقل من مستوى المعنوية 5%. وبالتالي متوسط المجتمع لا يساوي 78.

- **بالنسبة للحالة الثانية**

يكون <u>شكل الاختبار</u> هو:

$$\begin{cases} \mathbf{H_0}: \boldsymbol{\mu} \leq \mathbf{78} \\ \mathbf{H_1}: \boldsymbol{\mu} > 78 \end{cases}$$

<u>القيمة الحرجة (الجدولية)</u> سوف تكون:

$$\boldsymbol{Z_\alpha = 1.64}$$

أما **<u>قيمة إحصائية الاختبار</u>** فقيمتها تكون مساوية إلى:

$$\boldsymbol{Z} = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} = \frac{80 - 78}{\sqrt{49/100}} = \mathbf{2.85}$$

<u>القرار</u>:

➢ الطريقة الأولى:

الاختبار من جهة واحدة نحو اليمين أي $\mathbf{H_1}: \boldsymbol{\mu} > \boldsymbol{\mu_0}$ وبالتالي نرفض $\boldsymbol{H_0}$ لأنّ قيمة إحصائية الاختبار أكبر من القيمة الجدولية. بمعنى متوسط المجتمع لا يساوي 78.

➢ الطريقة الثانية:

الاختبار من جهة واحدة من جهة اليمين وبالتالي:

$$\boldsymbol{p-value = P(Z > 2.85) = 0.0021}$$

والقرار: نرفض $\boldsymbol{H_0}$ لأنّ **p- value** أقل من مستوى المعنوية 5%. بمعنى متوسط المجتمع لا يساوي 78.

- **الحالة الثالثة:**

يكون <u>شكل الاختبار</u> هو:

$$\begin{cases} \mathbf{H_0}: \boldsymbol{\mu} \geq \mathbf{78} \\ \mathbf{H_1}: \boldsymbol{\mu} < 78 \end{cases}$$

<u>القيمة الحرجة (الجدولية)</u> سوف تكون:

$$\boldsymbol{-Z_\alpha = -1.64}$$

أما <u>قيمة إحصائية الاختبار</u> فقيمتها تكون مساوية إلى:

$$Z = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} = \frac{80 - 78}{\sqrt{49/100}} = \mathbf{2.85}$$

<u>والقرار</u>:

➢ الطريقة الأولى:

الاختبار من جهة واحدة نحو اليسار أي $\mathbf{H_1: \mu < \mu_0}$ وبالتالي نقبل H_0 لأنّ قيمة إحصائية الاختبار أكبر من القيمة الجدولية.

➢ الطريقة الثانية:

الاختبار من جهة واحدة من جهة اليسار وبالتالي:

$$\boldsymbol{p - value = P(Z < 2.85) = 0.9978}$$

والقرار: نقبل H_0 لأنّ **p- value** أكبر من مستوى المعنوية 5%.

2-1-2- عندما يكون تباين المجتمع σ^2 مجهولا

- **إذا كان حجم العينة كبيرا $(n \geq 30)$:**

نستبدل تباين المجتمع بتباين العينة، ويبقى التوزيع توزيع طبيعي. وتكون إحصائية الاختبار هي:

$$\boldsymbol{Z = \frac{\bar{X} - \mu_0}{\sqrt{S^2/n}} \sim N(0,1)}$$

مثال:

من سجلات الوفيات لإحدى البلديات لولاية ميلة أخذت عينة من **100** شخص متوفي فوجد أن متوسط العمر لهؤلاء كان **71.8** سنة بانحراف معياري قدره **8.9** سنة. والمطلوب اختبر صحة الفرضية القائلة بأن متوسط العمر أكبر من **70** سنة وذلك عند مستوى معنوية **5%.**

الحل:

يكون **<u>شكل الاختبار</u>** هو:

$$\begin{cases} \mathbf{H_0: \mu \leq 70} \\ \mathbf{H_1: \mu} > 70 \end{cases}$$

<u>القيمة الحرجة (الجدولية)</u> سوف تكون:

$$\boldsymbol{Z_\alpha = 1.64}$$

أما <u>قيمة إحصائية الاختبار</u> فقيمتها تكون مساوية إلى:

$$Z = \frac{\bar{X} - \mu_0}{\sqrt{S^2/_n}} = \frac{71.8 - 70}{\sqrt{79.21/_{100}}} = \mathbf{2.022}$$

القرار:

➢ الطريقة الأولى:

نرفض H_0 لأنّ قيمة إحصائية الاختبار أكبر من القيمة الجدولية. بمعنى متوسط عمر المتوفي يزيد عن 70 سنة.

➢ الطريقة الثانية:

الاختبار من جهة واحدة من جهة اليمين وبالتالي:

$$p - value = P(Z > 2.022) = 0.0216$$

والقرار: نرفض H_0 لأنّ **p- value** أقل من مستوى المعنوية 5%. بمعنى متوسط عمر المتوفي يزيد عن 70 سنة.

- **إذا كان حجم العينة صغيرا** $(n < 30)$:

نستبدل تباين المجتمع بتباين العينة، والتوزيع هو توزيع ستودنت بدرجة حرية $(n-1)$. وتكون إحصائية الاختبار هي:

$$T = \frac{\bar{X} - \mu_0}{\sqrt{S^2/_n}} \sim t(n-1)$$

2-2- اختبار الفرضيات للفرق بين متوسطي مجتمعين طبيعيين مستقلين

إذا كان لدينا مجتمعان يتوزعان توزيعا طبيعيا بحيث $X_1 \sim N(\mu_1, \sigma_1^2)$ و $X_2 \sim N(\mu_2, \sigma_2^2)$، وأردنا وضع فرضيات إحصائية للفرق بين متوسطي المجتمعين $(\mu_1 - \mu_2)$ فستكون في إحدى الأشكال التالية:

➢ اختبار الفرضيات من جهتين (طرفين): $\begin{cases} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 \neq 0 \end{cases}$

➢ اختبار الفرضيات من جهة اليمين: $\begin{cases} H_0: \mu_1 - \mu_2 \leq 0 \\ H_1: \mu_1 - \mu_2 > 0 \end{cases}$

➢ اختبار الفرضيات من جهة اليسار: $\begin{cases} H_0: \mu_1 - \mu_2 \geq 0 \\ H_1: \mu_1 - \mu_2 < 0 \end{cases}$

والقرار في قبول أو رفض الفرضية الصفرية H_0 يمكن أن تتخذ بطريقتين مختلفين:

➢ **مقارنة القيمة المحسوبة للإحصائية مع القيمة الجدولية لها:**

1- إذا كان الاختبار من جهتين أي $H_1: \mu_1 - \mu_2 \neq 0$، نرفض H_0 إذا كانت القيمة المحسوبة للاختبار تقع في منطقة الرفض، أي تكون قيمة إحصائية الاختبار أكبر من القيمة الجدولية الموجبة أو أصغر من القيمة الجدولية السالبة.

2- إذا كان الاختبار من جهة واحدة نحو اليمين أي $\mathbf{H_1: \mu_1 - \mu_2 > 0}$ نرفض H_0 إذا كانت قيمة إحصائية الاختبار أكبر من القيمة الجدولية.

3- إذا كان الاختبار من جهة واحدة نحو اليسار أي $\mathbf{H_1: \mu_1 - \mu_2 < 0}$ نرفض H_0 إذا كانت قيمة إحصائية الاختبار أصغر من القيمة الجدولية.

- **باستخدام المعنوية المحسوبة p- value**

1- إذا كان الاختبار من جهتين أي $\mathbf{H_1: \mu_1 - \mu_2 \neq 0}$ فإن:

$$p - value = 2.P(\text{المتغير العشوائي} > \text{قيمة الاحصائية إذا كانت موجبة})$$

أو

$$p - value = 2.P(\text{المتغير العشوائي} < \text{قيمة الاحصائية إذا كانت سالبة})$$

والقرار: نرفض H_0 إذا كانت **p- value** أقل من مستوى المعنوية ونقبلها إذا كانت **p- value** أكبر من أو تساوي مستوى المعنوية.

2- إذا كان الاختبار من جهة اليمين فإن:

$$p - value = P(\text{المتغير العشوائي} > \text{قيمة الاحصائية})$$

والقرار: نرفض H_0 إذا كانت **p- value** أقل من مستوى المعنوية ونقبلها إذا كانت **p- value** أكبر من أو تساوي مستوى المعنوية.

3- إذا كان الاختبار من جهة اليسار فإن:

$$p - value = P(\text{المتغير العشوائي} < \text{قيمة الاحصائية})$$

والقرار: نرفض H_0 إذا كانت **p- value** أقل من مستوى المعنوية ونقبلها إذا كانت **p- value** أكبر من أو تساوي مستوى المعنوية.

في ما يلي سنذكر الحالات المختلفة في اختبار الفرضيات للفرق بين متوسطي مجتمعين طبيعيين مستقلين:

2-2-1- عندما يكون تباين المجتمعين معلومين (σ_2^2 و σ_1^2 معلومين)

تكون إحصائية الاختبار في هذه الحالة هي:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0.1)$$

2-2-2- عندما يكون تباين المجتمعين مجهولين (σ_2^2 و σ_1^2 مجهولين)

نستبدل تباين المجتمعين σ_2^2 و σ_1^2 بتباين العينتين S_2^2 و S_1^2. ويجب أيضا أن نميز بين حالتين:

2-2-2-1- عندما يكون σ_2^2 و σ_1^2 مجهولين وغير متساويين

- **إذا كان** $\mathbf{n_1, n_2} \geq \mathbf{30}$: تكون إحصائية الاختبار في هذه الحالة هي:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0.1)$$

- **إذا كان** $\mathbf{n_1, n_2} < 30$ **أو أحدهما صغير**: تكون إحصائية الاختبار هي:

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t(n_1 + n_2 - 2)$$

2-2-2-2- **عندما يكون** $\boldsymbol{\sigma_2^2}$ **و** $\boldsymbol{\sigma_1^2}$ مجهولين و متساويين

نستبدل قيمة كل من σ_1^2 و σ_2^2 بـ التباين المشترك S_P^2. وقيمة إحصائية الاختبار تكون هي:

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t(\vartheta = n_1 + n_2 - 2)$$

مثال :

لنفرض أنه لديك المعطيات التالية:

$\bar{X}_1 = 31.9$ $\quad \bar{X}_2 = 32.1$ $\quad S_1^2 = 6.806$ $\quad S_2^2 = 3.408$ $\quad S_P^2 = 5.261$ $\quad \alpha = 5\%$

والمطلوب هل يوجد فرق معنوي بين متوسطي المجتمعين؟.

الحل:

شكل الاختبار هو:

$$\begin{cases} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 \neq 0 \end{cases}$$

القيم الحرجة (الجدولية) سوف تكون:

$$-t_{\frac{\alpha}{2},\vartheta} = t_{0.975,11} = -2.201 \qquad t_{\frac{\alpha}{2},\vartheta} = t_{0.975,11} = 2.201$$

أما قيمة إحصائية الاختبار فقيمتها تكون مساوية إلى:

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{-0.2 - 0}{\sqrt{5.261\left(\frac{1}{7} + \frac{1}{6}\right)}} = -0,16$$

القرار:

- الطريقة الأولى:

قيمة $\boldsymbol{T}$ المحسوبة تقع بين القيمتين الجدوليتين أي تقع في منطقة القبول، ومنه نقبل الفرضية H_0 وبالتالي لا يوجد فرق بين متوسطي المجتمعين.

- الطريقة الثانية:

الاختبار من جهتين أي $\mathbf{H_1: \mu_1 - \mu_2 \neq 0}$ وبالتالي:

$$\boldsymbol{p-value = 2.P(\mathrm{T} > -0.16) = 0.80}$$

وبالتالي نقبل H_0 لأنّ **p- value** أكبر من مستوى المعنوية 5%. وبالتالي لا يوجد فرق بين متوسطي المجتمعين.