المحاضرة 9- تابع لاختبار الفرضيات الإحصائية-

-3- اختبار الفرضيات للفرق بين متوسطي مجتمعين مرتبطين

بما أنّ متوسط مجتمع الفروق μ_D يساوي إلى فرق متوسطي المجتمعين $\mu_1 - \mu_2$ فإنّ مسألة اختبار الفرضيات ستكون بإحدى الأشكال التالية:

$$\begin{cases} H_0: \mu_D \leq 0 \\ H_1: \mu_D > 0 \end{cases} \quad , \quad \begin{cases} H_0: \mu_D \geq 0 \\ H_1: \mu_D < 0 \end{cases} \quad , \quad \begin{cases} H_0: \mu_D = 0 \\ H_1: \mu_D \neq 0 \end{cases}$$

ويجب أن نميز بين حالتين:

- في حالة $n \geq 30$: تكون إحصائية الاختبار هي:

$$Z = \frac{\bar{D} - \mu_D}{\sqrt{S_D^2/n}}$$

مثال:

من مجتمعين طبيعيين مرتبطين تم سحب عينتين عشوائيتين حجم كل واحدة يساوي إلى **36** وكان مجموع الفروق بين القيم المتناظرة للعينتين هو **100** والانحراف المعياري للفروق يساوي 16. والمطلوب، عند مستوى معنوية تساوي **5%** اختبر فرضية أن المجتمعين لهما نفس المتوسط.

الحل:

$$n = 36 \qquad \sum D_i = 100 \qquad S_D = 16 \qquad \alpha = 5\%.$$

<u>**شكل الاختبار**</u> هو:

$$\begin{cases} H_0: \mu_D = 0 \\ H_1: \mu_D \neq 0 \end{cases}$$

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

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

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

$$Z = \frac{\bar{D} - \mu_D}{\sqrt{S_D^2/n}} = \frac{\frac{100}{36} - 0}{\sqrt{16^2/36}} = 1.04$$

القرار:

- الطريقة الأولى:

قيمة Z المحسوبة تقع بين القيمتين الجدوليتين (بمعنى تقع في منطقة القبول) ، ومنه نقبل الفرضية H_0 وبالتالي لا يوجد فرق بين متوسطي المجتمعين، أي أنّ المجتمعين لهما نفس المتوسط.

- الطريقة الثانية:

الاختبار من جهتين أي $\mathbf{H_1: \mu_D \neq 0}$ وبالتالي:

$$\boldsymbol{p-value} = \mathbf{2.}\boldsymbol{P}(\mathbf{Z} > \mathbf{1.04}) = 2[1 - P(Z \leq 1.04)] = 2\,[1 - (0.850830)] = \mathbf{0.2983}$$

وبالتالي نقبل H_0 لأنّ **p- value** أكبر من مستوى المعنوية 10%. وبالتالي لا يوجد فرق بين متوسطي المجتمعين.

- في حالة $\boldsymbol{n} < 30$: تكون إحصائية الاختبار هي:

$$\boldsymbol{T} = \frac{\overline{\boldsymbol{D}} - \boldsymbol{\mu_D}}{\sqrt{\boldsymbol{S_D^2}/\boldsymbol{n}}} \sim \boldsymbol{t}(\boldsymbol{n}-\mathbf{1})$$

مثال :

ادعى طبيب جزائري اخترع دواء جديد لمعالجة مرضى السكري، أن المريض الذي يتناول هذا الدواء سيعرف تحسنا من الناحية الصحية مقارنة بالشخص الذي لم يتناول هذا الدواء، ولغرض اختبار صحة هذا الادعاء تم اختيار عشرين شخص مصاب بمرض السكري وتقسيمهم إلى عشرة أزواج، بحيث كل زوج يتألف من شخصين متساويين في القياسات، وكانت النتائج كما يلي:

أشخاص تناولوا الدواء (نسبة السكر في الدم بـــ ملغ/ ديسيلتر)	128	131	134	128	129	132	142	138	136	130
أشخاص لم يتناولوا الدواء (نسبة السكر في الدم بـــ ملغ/ ديسيلتر)	132	131	136	127	133	132	144	142	138	130

والمطلوب، عند مستوى معنوية 5% هل يوجد هناك تحسنا عند تناول الدواء الجديد.

الحل:

شكل الاختبار هو:

$$\begin{cases} \mathbf{H_0}: \boldsymbol{\mu_D} \leq \mathbf{0} \\ \mathbf{H_1}: \boldsymbol{\mu_D} > \mathbf{0} \end{cases}$$

تكون الفروق ما بين قبل أخذ الدواء وبعد أخذه كما يلي:

أشخاص تناولوا الدواء (نسبة السكر في الدم بـ ملغ/ ديسيلتر)	128	131	134	128	129	132	142	138	136	130
أشخاص لم يتناولوا الدواء (نسبة السكر في الدم بـ ملغ/ ديسيلتر)	132	131	136	127	133	132	144	142	138	130
الفروق D_i	4	0	-2	1	-4	0	-2	-4	-2	0

وتكون:

$$\bar{D} = \frac{\sum D_i}{n} = -0.9$$

$$S_D^2 = \frac{\sum (D_i - \mu_D)^2}{n-1} = \frac{52.9}{9} = 5.88$$

وبذلك نحصل على:

$$T = \frac{\bar{D} - \mu_D}{\sqrt{\frac{S_D^2}{n}}} = \frac{-0.9 - 0}{\sqrt{\frac{5.88}{10}}} = -1.16$$

ومن جدول توزيع ستودنت عند درجة حرية تساوي 9 نجد:

$$t_{(\alpha, n-1)} = t_{(0.05,9)} = 1.833$$

القرار:

- الطريقة الأولى:

قيمة T المحسوبة أقل من القيمة الجدولية (بمعنى تقع في منطقة القبول) ، ومنه نقبل الفرضية H_0 وبالتالي لا يوجد تحسن عند تناول الدواء الجديد.

- الطريقة الثانية:

الاختبار من جهة اليمين أي $H_1: \mu_D > 0$ وبالتالي:

$$p - value = P(T > -1.16) = [1 - P(T > 1.16)] = [1 - (0.10)] = 0.90$$

وبالتالي نقبل H_0 لأنّ **p- value** أكبر من مستوى المعنوية 5%. وبالتالي لا يوجد تحسن عند تناول الدواء الجديد.

4-2- اختبار الفرضيات للنسبة P

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

$$\begin{cases} H_0: P \le P_0 \\ H_1: P > P_0 \end{cases} , \quad \begin{cases} H_0: P \ge P_0 \\ H_1: P < P_0 \end{cases} , \quad \begin{cases} H_0: P = P_0 \\ H_1: P \ne P_0 \end{cases}$$

وتكون إحصائية الاختبار هي:

$$Z = \frac{\widehat{P} - P_0}{\sqrt{P_0 . q_0 / n}} \sim N(0,1)$$

مثال :

حسب بيانات الديوان الوطني للإحصائيات فإن نسبة الأمية في الجزائر تعرف انخفاضا ملحوظا منذ العام 2007 بفضل الاستيراتيجية الوطنية لمحو الأمية، إذ بلغت في العام 2018 نسبة 9.44%. عند سحبنا لعينة عشوائية من المجتمع حجمها 500 وجدنا 52 شخص أمي. والمطلوب عند مستوى معنوية 5% اختبر فرضية أن نسبة الأمية في انخفاض مستمر منذ تبني الاستيراتيجية الوطنية لمحو الأمية.

الحل:

لدينا من معطيات المثال:

$$\widehat{\mathrm{P}} = \frac{52}{500} = \mathbf{10.4\%} \qquad \mathrm{P} = \mathbf{9.44\%} \qquad \mathbf{n = 500}$$

يكون شكل الاختبار هو:

$$\begin{cases} \mathrm{H_0}: P \geq 0.0944 \\ \mathrm{H_1}: \mathrm{P} < 0.0944 \end{cases}$$

القيمة الحرجة (الجدولية) سوف تكون:

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

أما قيمة إحصائية الاختبار فقيمتها تكون مساوية إلى:

$$Z = \frac{0.104 - 0.0944}{\sqrt{(0.0944).(0.9056) / 500}} = \mathbf{0.7341}$$

القرار:

- الطريقة الأولى:

قيمة Z المحسوبة أكبر من القيمة الجدولية لها، ومنه نقبل الفرضية H_0 (ونرفض الفرضية H_1) وبالتالي نسبة الأمية ليست في انخفاض مستمر منذ تبني الاستيراتيجية الوطنية لمحو الأمية.

- الطريقة الثانية:

الاختبار من جهة اليسار أي $\mathbf{H_1}: \mathbf{P} < \mathbf{0.0944}$ وبالتالي:

$$\boldsymbol{p - value} = P(\mathrm{Z} < 0.7341) = \mathbf{0.7673}$$

وبالتالي نقبل H_0 لأنّ **p- value** أكبر من مستوى المعنوية 5%. وبالتالي فرضية نسبة الأمية في الجزائر في انخفاض مستمر منذ تبني الاستيراتيجية الوطنية لمحو الأمية مرفوضة.

2-5- اختبار الفرضيات للفرق بين نسبتين $(P_1 - P_2)$

رأينا في الفصلين الأول والثاني أنه إذا سحبنا عينة كبيرة n_1 من مجتمع يخضع لتوزيع برنولي (ذي الحدين $(X_1 \sim b(1, P_1)$ وسحبنا عينة كبيرة أخرى n_2 من مجتمع آخر مستقل عن المجتمع الأول ويخضع لتوزيع برنولي (ذي الحدين $(X_2 \sim b(1, P_2)$ فإنّ الفرق بين النسبتين في العينيتن $(\widehat{P_1} - \widehat{P_2})$ هو متغير عشوائي يخضع تقريبا للتوزيع الطبيعي.

إذا أردنا اختبار الفرضيات لتساوي نسب المجتمعين فتكون لدينا ثلاث (03) صيغ ممكنة وهي:

$$\begin{cases} H_0: P_1 = P_2 \\ H_1: P_1 \neq P_2 \end{cases} \quad , \quad \begin{cases} H_0: P_1 \geq P_2 \\ H_1: P_1 < P_2 \end{cases} \quad , \quad \begin{cases} H_0: P_1 \leq P_2 \\ H_1: P_1 > P_2 \end{cases}$$

وتكون إحصائية الاختبار هي:

$$Z = \frac{(\widehat{P}_1 - \widehat{P}_2) - P_1 - P_2}{\sqrt{\dfrac{\widehat{P}_1(1 - \widehat{P}_1)}{n_1} + \dfrac{\widehat{P}_2(1 - \widehat{P}_2)}{n_2}}} \sim N(0,1)$$

مثال:

بهدف مقارنة نوعين من الأدوية لتخفيف الألم بعد إجراء العمليات الجراحية، تم استخدام الدواء A على عينة من 100 مريض وادعى منهم 48 مريضا بأن الدواء A قد خفف الألم لديهم بعد خضوعهم للعمليات الجراحية، واستخدام الدواء B على عينة من 120 مريضا وأدعى منهم 54 مريضا بأن الدواء B قد خفف الألم لديهم بعد خضوعهم للعمليات الجراحية. والمطلوب، عند مستوى معنوية 5% هل يوجد فرق حقيقي ما بين الدواء A والدواء B في تخفيف الألم عند المرضى لدى خضوعهم للعمليات الجراحية.

الحل:

من معطيات المثال لدينا:

$$n_1 = 100 \qquad \widehat{P}_1 = \frac{48}{100} = 0.48 \qquad \alpha = 5\%.$$

$$n_2 = 120 \qquad \widehat{P}_2 = \frac{54}{120} = 0.45$$

نريد اختبار الفرضية:

$$\begin{matrix} H_0: P_1 = P_2 \\ H_1: P_1 \neq P_2 \end{matrix}$$

تكون قيمة إحصائية الاختبار هي:

$$Z = \frac{(\widehat{P}_1 - \widehat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{\widehat{P}_1(1-\widehat{P}_1)}{n_1} + \frac{\widehat{P}_2(1-\widehat{P}_2)}{n_2}}} = \frac{(0.48 - 0.45) - 0}{\sqrt{\frac{0.48(0.52)}{100} + \frac{0.45(0.55)}{120}}} = \mathbf{0.44}$$

الاختبار من جهتين والتوزيع توزيع طبيعي، ومستوى المعنوية تساوي 5% وبذلك تكون القيم الحرجة هي:

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

القرار:

- الطريقة الأولى:

قيمة Z المحسوبة تقع بين القيمتين الجدوليتين (بمعنى تقع في منطقة القبول) ، ومنه نقبل الفرضية H_0 وبالتالي لا يوجد فرق بين الدوائين A و B.

- الطريقة الثانية:

الاختبار من جهتين أي $\mathbf{H_1: P_1 \neq P_2}$ وبالتالي:

$$p - value = 2.P(Z > 0.44) = 2[1 - P(Z \leq 0.44)] = 2\,[1 - (0.670031)] = \mathbf{0.6599}$$

وبالتالي نقبل H_0 لأنّ **p- value** أكبر من مستوى المعنوية 5%. وبالتالي لا يوجد فرق بين الدوائين A وB.

6-2- اختبار الفرضيات لتباين مجتمع يتوزع توزيعا طبيعيا (σ^2)

تكلمنا سابقا في الفصلين الأول والثاني أنه إذا كان $X \sim N(\mu, \sigma^2)$ وكان S^2 هو تباين العينة فإن $\frac{(n-1)S^2}{\sigma^2}$ يخضع لتوزيع مربع كاي بدرجة حرية $n-1$ أي $\chi = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$.

وعليه لاختبار الفرضيات حول تباين المجتمع σ^2 توجد ثلاث (03) صيغ ممكنة هي:

$$\begin{cases} H_0: \sigma^2 \leq \sigma_0^2 \\ H_1: \sigma^2 > \sigma_0^2 \end{cases} , \qquad \begin{cases} H_0: \sigma^2 \geq \sigma_0^2 \\ H_1: \sigma^2 < \sigma_0^2 \end{cases} , \qquad \begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 \neq \sigma_0^2 \end{cases}$$

وتكون إحصائية الاختبار هي:

$$\chi = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

مثال :

بعد الإعلان عن علامات الامتحان النهائي لطلبة الثانية علوم تسيير في مادة الإحصاء 3 والتي تتوزع توزيعا طبيعيا بتباين قدره 9، ادعى مجموعة من الأساتذة الجامعيين أنّ إتباع طريقة جديدة في تدريس مادة الإحصاء 3 سوف يقلل من التباين في علامات الطلبة، ولغرض التحقق من هذا الإدعاء سحبنا عينة مكونة من 25 طالب، وبعد بتدريسهم على حسب الطريقة الجديدة أجرينا لهم امتحان، فكان تباين علاماتهم يساوي 7.8. والمطلوب عند مستوى معنوية تساوي 5% هل فعلا الطريقة الجديدة المتبعة في التدريس تقلل من التباين في علامات الطلبة؟.

الحل:

من معطيات المثال لدينا:

$\sigma_0^2 = 9$ $\quad n = 25$ $\quad S^2 = 7.8$ $\quad \alpha = 5\%$

نريد اختبار الفرضية:

$$\mathrm{H_0}: \sigma^2 \geq 9$$
$$\mathrm{H_1}: \sigma^2 < 9$$

تكون قيمة إحصائية الاختبار هي:

$$\chi = \frac{(\mathbf{n}-\mathbf{1})\mathbf{S}^2}{\boldsymbol{\sigma}^2} = \frac{24 \times 7.8}{9} = 20.8$$

من جدول توزيع مربع كاي عند مستوى معنوية $\alpha = 5\%$ تكون القيمة الحرجة مساوية إلى:

$$\chi^2_{(1-\alpha,n-1)} = \chi^2_{(0.95,24)} = 13.85$$

القرار:

- الطريقة الأولى:

قيمة χ^2 المحسوبة أكبر من القيمة الجدولية لها، ومنه نقبل الفرضية H_0 وبالتالي تباين المجتمع ليس اقل من 9، بمعنى آخر الطريقة الجديدة المتبعة في التدريس لم تؤد إلى تقليص التشتت في علامات مادة الإحصاء 3.

- الطريقة الثانية:

الاختبار من جهة اليسار أي $\mathrm{H_1}: \sigma^2 < 9$ وبالتالي:

$$p-value = P(\chi^2 < 20{,}8) = 0.10$$

وبالتالي نقبل H_0 لأنّ **p- value** أكبر من مستوى المعنوية 5%. أي أنّ فرضية الطريقة الجديدة المتبعة في التدريس تقلّص التشتت في علامات مادة الإحصاء 3 مرفوضة.

2-6- اختبار الفرضيات للنسبة بين تبايني مجتمعين مستقلين يتوزعان توزيعا طبيعيا $\left(\frac{\sigma_1^2}{\sigma_2^2}\right)$

في هذه الحالة تكون صيغ اختبار الفرضيات كالتالي:

$$\begin{cases} H_0: \sigma_1^2 \le \sigma_2^2 \\ H_1: \sigma_1^2 > \sigma_2^2 \end{cases} , \quad \begin{cases} H_0: \sigma_1^2 \ge \sigma_2^2 \\ H_1: \sigma_1^2 < \sigma_2^2 \end{cases} , \quad \begin{cases} H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 \ne \sigma_2^2 \end{cases}$$

أو بعبارة أخرى:

$$\begin{cases} H_0: \dfrac{\sigma_1^2}{\sigma_2^2} \le 1 \\ H_1: \dfrac{\sigma_1^2}{\sigma_2^2} > 1 \end{cases} , \quad \begin{cases} H_0: \dfrac{\sigma_1^2}{\sigma_2^2} \ge 1 \\ H_1: \dfrac{\sigma_1^2}{\sigma_2^2} < 1 \end{cases} , \quad \begin{cases} H_0: \dfrac{\sigma_1^2}{\sigma_2^2} = 1 \\ H_1: \dfrac{\sigma_1^2}{\sigma_2^2} \ne 1 \end{cases}$$

وتكون إحصائية الاختبار كما رأينا في الفصلين الأول والثاني هي:

$$f = \frac{S_1^2}{S_2^2} . \frac{\sigma_2^2}{\sigma_1^2} \sim F(n_1 - 1, n_2 - 1)$$

مثال (3- 13):

إذا كانت أطوال شجيرات الزيتون في المزرعتين A و B تتبع التوزيع الطبيعي، وسحبنا عينة من المزرعة A حجمها 20 ومن المزرعة B حجمها 25 علما أنّ العينتين مستقلتين، ووجدنا أن تباين أطوال الشجيرات في المزرعة A يساوي 49 وفي المزرعة B يساوي 36. والمطلوب عند مستوى معنوية تساوي 5% هل يوجد فرق بين تباين مجتمع أطوال شجيرات الزيتون في المزرعة A وتباين مجتمع أطوال شجيرات الزيتون في المزرعة B؟.

الحل:

من معطيات المثال لدينا:

$$\alpha = 5\% \qquad S_1^2 = 49 \qquad n_1 = 20$$
$$S_1^2 = 36 \qquad n_2 = 25$$

نريد اختبار الفرضية:

$$\begin{cases} H_0: \dfrac{\sigma_1^2}{\sigma_2^2} = 1 \\ H_1: \dfrac{\sigma_1^2}{\sigma_2^2} \ne 1 \end{cases}$$

تكون قيمة إحصائية الاختبار هي:

$$f = \frac{S_1^2}{S_2^2} . \frac{\sigma_2^2}{\sigma_1^2} = \frac{49}{36} . 1 = 1.36$$

<u>القيم الحرجة (الجدولية) عند درجتي حرية</u> $n_1 - 1 = 19$ و $n_2 - 1 = 24$ سوف تكون:

$$\boldsymbol{F}_{\left(\frac{\alpha}{2},\, n_1-1, n_2-1\right)} = \boldsymbol{F}_{(0.025,19,24)} = \mathbf{2.33}$$

$$\boldsymbol{F}_{\left(1-\frac{\alpha}{2},\, n_1-1, n_2-1\right)} = \boldsymbol{F}_{(0.975,19,24)} = \frac{\mathbf{1}}{\boldsymbol{F}_{(0.025,24,19)}} = \frac{\mathbf{1}}{\mathbf{2.45}} = \mathbf{0.408}$$

القرار:

- الطريقة الأولى:

قيمة $\boldsymbol{f}$ المحسوبة تقع بين القيم الجدولية لها، ومنه نقبل الفرضية H_0 وبالتالي لا يوجد فرق بين تباين مجتمع أطوال شجيرات الزيتون في المزرعة $\boldsymbol{A}$ وتباين مجتمع أطوال شجيرات الزيتون في المزرعة $\boldsymbol{B}$.

- الطريقة الثانية:

الاختبار من جهتين أي $\mathbf{H_1} : \frac{\sigma_1^2}{\sigma_2^2} \neq \mathbf{1}$ وبالتالي:

$$\boldsymbol{p-value} = 2.P\left(\mathrm{F}_{(19,24)} > 1.36\right) = \mathbf{0.50}$$

وبالتالي نقبل H_0 لأنّ **p- value** أكبر من مستوى المعنوية 5%. أي لا يوجد فرق بين تباين مجتمع أطوال شجيرات الزيتون في المزرعة $\boldsymbol{A}$ وتباين مجتمع أطوال شجيرات الزيتون في المزرعة $\boldsymbol{B}$.