المحور الثالث: نظرية الاختبار (الفروض)

1- مفهوم احتمالات الخطأ والفرضيات:

الاختبار الإحصائي عبارة عن آلية تتيح الاختيار بين فرضيتين بناء على نتائج العينة بافتراض أن H_0 وH_1 فرضيتان إحداهما فقط صحيحة، حيث تتعلق قاعدة القرار باختيار إما H_0 أو H_1، وهناك أربعة حالات يمكن توضيحها في الجدول التالي:

جدول رقم 01:

	H_0 صحيحة	**H_1 صحيحة**
قبول H_0	$1-\alpha$	β
قبول H_1	α	$1-\beta$

حيث α و β هما احتمالي خطأ من النوع الأول و من النوع الثاني:

- α هو احتمال قبول H_1 عندما تكون H_0 صحيحة؛
- β هو احتمال قبول H_0 عندما تكون H_1 صحيحة.

ملاحظة: تعتبر احتمالات الخطأ هذه متعارضة، فكلما كانت α كبيرة (صغيرة على التوالي)، كلما كانت β صغيرة (كبيرة على التوالي)

- **تعريف:** قوة الاختبار تساوي $1-\beta$، وفي ممارسة الاختبارات الإحصائية من القاعدة تثبيت α، بحيث تعتبر النسب التالية: 1 %، 5 %، 10 % الأكثر تداولا، وتسمى H_0 الفرضية المرجعية، كما تسمى أيضا الفرضية الصفرية (فرضية العدم)، بينما تسمى H_1 الفرضية البديلة.

2- مفردات الاختبار الإحصائي:

اعتمادا على الفرضيتين H_0 وH_1 المراد اختبارهما نختار متغير القرار، ويجب معرفة قانون هذا المتغير في فرضية واحدة على الأقل تكون في أغلب الأحيان H_0، حتى لا يتم إدخال مجاهيل جديدة في المشكلة الاحصائية، ويسمى بالتوزيع المرجعي.

- **المنطقة الحرجة:** نسمي المنطقة الحرجة (منطقة الرفض) مجموعة قيم متغير القرار التي تؤدي إلى رفض الفرضية الصفرية H_0 لصالح الفرضية البديلة H_1، وتتوافق مع المجالات التي تكون فيها الاختلافات كبيرة جدا نتيجة لاختيار العينات العشوائية؛

- **منطقة القبول:** نسمي منطقة القبول المنطقة المكملة للمنطقة الحرجة، وتتوافق منطقة القبول مع المجال الذي تعزى فيه الاختلافات الملحوظة بين الإنجازات والنظري إلى تقلبات أخذ العينات. ويعتمد بناء الاختبار على تحديد المنطقة الحرجة مسبقا دون معرفة نتيجة التجربة.

3- الاختبارات الثنائية والاختبارات الأحادية:

قبل القيام بالاختبار الإحصائي لا بد من تحديد المشكلة المطروحة، واعتمادًا على الفرضيات التي يتم صياغتها، فإننا نطبق إما: اختبارًا ثنائي الجانب أو اختبارًا أحادي الجانب.

- نطبق الاختبار الثنائي عندما نبحث عن الفرق بين قيمتي معلمتين، أو بين قيمة معلمة وقيمة معطاة، ويمكن أن تكون منطقة الرفض للفرضية المرجعية عبارة عن اتحاد منطقتين موجودتين في كل ذيل من التوزيع؛

- بينما ينطبق الاختبار أحادي الجانب عندما نريد معرفة ما إذا كانت المعلمة أكبر (أو أقل) من معلمة أخرى أو قيمة معينة، وتقع منطقة الرفض للفرضية الرئيسية على جانب واحد فقط من التوزيع الاحتمالي المرجعي.

4- الاختبارات الاحصائية:

4-1 اختبار العلاقة بين متوسط عينة ومتوسط مجتمع:

4-1-1 حالة معلومية الانحراف المعياري δ_X للمجتمع وحجم العينة كبير:

إذا أردنا دراسة أو اختبار العلاقة بين متوسط عينة $\bar{X}$ ومتوسط مجتمع μ، ومعرفة ما إذا كان الفرق معنويا أي $(\bar{X}-\mu)$ أو غير معنوي بدرجة ثقة معينة نتبع الخطوات التالية:

- تحديد درجة الثقة ومستوى المعنوية فمثلا إذا كانت درجة الثقة $(1-\alpha)=0.95\ \%$ فإن: قيمة Z الجدولية تساوي 1.96

- تحديد الفرض ومنطقة كل فرض وهذه الفروض هي:

- H_0: فرضية العدم (الفرضية الصفرية) ومعناه عدم وجود فرق معنوي بين المتوسطين وعليه فإن العينة تنتمي إلى المجتمع μ ونكتب $N(\mu\ ,\ \delta)$ وتكون منطقة قبول الفرض

$H_0 : \mu = \mu_0$ هي المنطقة المحددة بـــــــ: $-Z\frac{\alpha}{2} < Z < +Z\frac{\alpha}{2}$

- H_1: الفرضية البديلة ومعناه عدم وجود فرق معنوي بين المتوسطين أي أن: $(\bar{X}-\mu)$ مهم، وتظهر لنا ثلاثة بدائل في H_1 كالتالي:

الحالة الأولى: $\mu \neq \mu_0$ **:H_1** فإننا **نرفض H_0 إذا كانت:** (الجدولية) $| Z |> Z_{\frac{\alpha}{2}}$ (المحسوبة)

الحالة الثانية: $\mu > \mu_0$ **:H_1** فإننا **نرفض H_0** إذا كانت: **(الجدولية)** $|Z |> Z_{\frac{\alpha}{2}}$ **(المحسوبة)**

الحالة الثالثة: $\mu < \mu_0$ **:H_1** فإننا **نرفض H_0** إذا كانت: **(الجدولية)** $|Z |< Z_{\frac{\alpha}{2}}$ **(المحسوبة)**

والمقصود هنا بـــ: Z (المحسوبة) هي دالة الاختبار Z ويمكن إيجادها من العلاقة التالية:

$$Z=\frac{\bar{X}-\mu}{\frac{\delta}{\sqrt{n}}}$$

مثال: أخذت عينة بطريقة عشوائية تتكون من 100 كيس إسمنت من مصنع الإسمنت، ووجد أن متوسط وزن كيس الإسمنت 49.5 كغ

المطلوب: هل يمكن أن نستنتج أن متوسط هذه العينة يتماشى مع المتوسط العام لوزن الكيس و الذي يساوي 50 كغ إذا كان الانحراف المعياري $\delta=5$ بدرجة ثقة 95%.

الحل:

$n=100>300$ (العينة كبيرة)

$\delta=5$ (الانحراف المعياري معلوم)

نطبق قانون الاختبار Z حيث: $Z=\frac{\bar{X}-\mu}{\frac{\delta}{\sqrt{n}}}$

(الحسابية)$Z=\frac{49.5-50}{\frac{5}{\sqrt{100}}}=\frac{-0.5}{0.5}=-1$

$|Z |= |-1 |= 1$

قاعدة اتخاذ القرار:

إذا كانت: (النظرية)Z < $|Z|$ (الحسابية) نرفض H_0

بما أن: (النظرية) $Z= 1.96 > |Z |=1$ (الحسابية) فإننا نقبل H_0

4-1-2 حالة الانحراف معياري δ_X للمجتمع مجهول وحجم العينة صغير:

إذا كانت $x_1,x_2,x_3,\dots\dots x_n$ قيم مشاهدات العينة العشوائية مأخوذة من توزيع طبيعي وكانت δ غير معلومة فإننا نستخدم قيمة دلالة الاختبار T بحيث:

$$T=\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$$

وإذا كانت $H_0 : \mu = \mu_0$ فهناك ثلاثة بدائل في H_1 كالتالي:

الحالة الأولى: $H_1: \mu \neq \mu_0$ فإننا **نرفض H_0 إذا كانت:** (النظرية) $t(1-\frac{\alpha}{2}, n-1)$ < | T | (المحسوبة)

الحالة الثانية: $H_1: \mu > \mu_0$ فإننا **نرفض H_0** إذا كانت: (النظرية) $t(1-\alpha, n-1)$ < | T | (المحسوبة)

الحالة الثالثة: $H_1: \mu < \mu_0$ فإننا **نرفض H_0** إذا كانت: (النظرية) $t(1-\alpha, n-1)$ > | T | (المحسوبة)

مثال: أخذت عينة عشوائية حجمها 25 مفردة من توزيع طبيعي $N(\mu,\delta)$ ، وإذا كان الانحراف المعياري للعينة يساوي 3، وأردنا أن نختبر H_0 حيث: $\overline{X} = 60$ مقابل: $H_1 > 55$ على مستوى الدلالة: $\alpha = 1\ \%$

الحل:

(النظرية) $t(1-\alpha, n-1)$, $t(0.995,24)$, $t=2.797$

(الحسابية) $T=\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} = \frac{60-55}{\frac{3}{\sqrt{25}}} = \frac{5}{\frac{3}{5}} = \frac{25}{3} = 8.33$

قاعدة اتخاذ القرار:

إذا كانت: (النظرية)t < $|T|$ (الحسابية) نرفض H_0

بما أن: (النظرية) $t = 2.797$ < $|T| = 8.33$ (الحسابية) فإننا نرفض H_0 على مستوى الدلالة: $\alpha = 1\ \%$

2-4 اختبار الفرضيات للفرق بين وسطين:

1-2-4 اختبار الفرضيات للفرق بين وسطين مع معلومية δ_{X1} و δ_{X2} (n1>30, n2>30):

إذا أخذت عينة عشوائية أولى حجمها n_1 ووسطها الحسابي $\bar{X}_1$ تتوزع توزيعا طبيعيا، ثم أخذت عينة عشوائية ثانية حجمها n_2 ووسطها الحسابي $\bar{X}_2$ من توزيع طبيعي مستقل عن الأول، وكانت δ_{X1} و δ_{X2} معلومتين، وأردنا اختبار $H_0 : \mu_1 = \mu_2$ مقابل الفرضية البديلة H_1 فإنه:

- ✓ إذا كانت: $H_1 : \mu_1 \neq \mu_2$ فإننا نرفض H_0 على مستوى الدلالة α إذا كانت:
  - الجدولية Z < | Z | (المحسوبة)

حيث أن: (المحسوبة) $Z = \frac{(\bar{x}1 - \bar{x}2)}{\sqrt{\frac{\delta 12}{n1} + \frac{\delta 22}{n2}}}$

- ✓ **إذا كانت:** $H_1 : \mu_1 < \mu_2$ فإننا نرفض H_0 على مستوى الدلالة α **إذا كانت:**
  - الجدولية Z > | Z | المحسوبة
- ✓ **إذا كانت:** $H_1 : \mu_1 > \mu_2$ فإننا نرفض H_0 على مستوى الدلالة α **إذا كانت:**
  - الجدولية Z < | Z | المحسوبة

مثال:

ليكن لدينا نوعين من السجائر وحسبنا متوسط النيكوتين في السجائر فوجدنا في النوع الأول $\bar{X}_1 = 24.1$ في حجم عينة يساوي 40 سيجارة والتباين $\delta^2_{X1} = 1.44$، فيما وجد في النوع الثاني $\bar{X}_2 = 23.8$ في حجم عينة يساوي 50 سيجارة والتباين $\delta^2_{X2} = 1.96$

اختبر على مستوى الدلالة α=5%:

$$\begin{cases} H0: \mu 1 = \mu 2 \\ H1: \mu 1 \neq \mu 2 \end{cases}$$

الحل:

(المحسوبة) $Z = \frac{(\bar{x}1 - \bar{x}2)}{\sqrt{\frac{\delta 12}{n1} + \frac{\delta 22}{n2}}} = \frac{24.1 - 23.8}{\sqrt{\frac{1.44}{40} + \frac{1.96}{50}}} = \frac{0.3}{0.27} = 1.11$

$Z_{0.025}=1.96$ (الجدولية)

بما أن: (المحسوبة) $|Z| < Z_{0.025}$ (الجدولية)

فإننا نقبل H_0 ونرفض H_1 وعليه: $\mu_1 = \mu_2$

2-2-4 اختبار الفرضيات للفرق بين وسطين إذا كان δ_{X1} و δ_{X2} مجهولان ($n1<30$, $n2<30$):

إذا كانت قيم المشاهدات $x_1, x_2, x_3, \ldots\ldots x_n$ عينة عشوائية من توزيع طبيعي $N(m_1, \delta_{X1})$ وكانت المشاهدات $y_1, y_2, y_3, \ldots\ldots y_n$ عينة عشوائية من توزيع طبيعي $N(m_2, \delta_{X2})$ مستقل عن التوزيع الأول وكان $\delta_{X1} = \delta_{X2}$ مجهولان وأردنا اختبار :

H_0: $\mu_1 = \mu_2$ على مستوى الدلالة α مقابل :

- **H_1 : $\mu_1 \neq \mu_2$** فإننا نرفض H_0 إذا كانت: $t\left\{1-\frac{\alpha}{2}, n1+n2-2\right\}$ النظرية $< |T|$ المحسوبة
- **H_1 : $\mu_1 < \mu_2$** فإننا نرفض H_0 إذا كانت: $t\{1-\alpha, n1+n2-2\}$ النظرية $< |T|$ المحسوبة
- **H_1 : $\mu_1 > \mu_2$** فإننا نرفض H_0 إذا كانت: $t\{1-\alpha, n1+n2-2\}$ النظرية $> |T|$ المحسوبة

علما أن: $T = \frac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n1}+\frac{1}{n2}}}$ (المحسوبة)

و: $S^2_c = \frac{S21(n1-1)+S22(n2-1)}{(n1+n2)-2}$

مثال: لتكن عينتان عشوائيتان من توزيعين مستقلين $N(m_1, \delta_{X1})$ و $N(m_2, \delta_{X2})$ وكانت كل من δ_{X1} و δ_{X2} مجهولتان ووجد أن حجم العينة الأولى $n_1 = 6$ وحجم العينة الثانية $n_2 = 8$ وكان الوسط الحسابي للعينة الأولى يساوي 16 وكان الوسط الحسابي للعينة الثانية يساوي 20 وكان $S^2_1 = 9$ بينما $S^2_2 = 16$

المطلوب: اختبار الفرضية $H_0: \mu_1 = \mu_2$ على مستوى الدلالة $\alpha = 5\ \%$ مقابل :

- $H_1 : \mu_1 \neq \mu_2$
- $H_1 : \mu_1 < \mu_2$
- $H_1 : \mu_1 > \mu_2$

الحل: $n_1 = 6,\ \ n_2 = 8,\ \ \bar{X}_1 = 16,\ \bar{X}_2 = 20,\ \ S^2_1 = 9,\ \ S^2_2 = 16$

$$S^2_c = \frac{S21(n1-1)+S22(n2-1)}{(n1+n2)-2} = \frac{9(6\text{-}1)+16(8\text{-}1)}{(6+8)\text{-}2} = \frac{9(5)+16(7)}{(14)\text{-}2} = \frac{9(5)+16(7)}{12} = \frac{9(5)+16(7)}{12}$$

$$= \frac{157}{12} = 13.08$$

$$\sqrt{S^2_c} = \sqrt{13.08} = 3.61$$

(المحسوبة) $T = \dfrac{\bar{X}-\bar{y}}{sc\sqrt{\frac{1}{n1}+\frac{1}{n2}}} = \dfrac{16-20}{\sqrt{13.08}\sqrt{\frac{1}{6}+\frac{1}{8}}} = \dfrac{-4}{\sqrt{13.08}\sqrt{\frac{1}{6}+\frac{1}{8}}} = \dfrac{-4}{3.61\ (0.53)} = \dfrac{-4}{1.91} = -\ 2.09$

- **الحالة الأولى:** $H_1 : \mu_1 \neq \mu_2$

(النظرية) $t\left\{1-\frac{\alpha}{2}, \mathrm{n1}+\mathrm{n2}-2\right\}$, $t\{0.975,12\}$, $t=2.179$

بما أن: $t = 2.\ 179$ (الجدولية) $> 2.09 = |\ T\ |$ (المحسوبة)

فإننا نقبل H_0 ونرفض H_1 وعليه: $\mu_1 = \mu_2$

- **الحالة الثانية:** $H_1 : \mu_1 < \mu_2$

(النظرية) $t\{1-\alpha, n1+n2-2\}$, $t\{0.95,12\}$, $t=1.782$

بما أن: $t = 1.782$ (الجدولية) $< 2.09 = |\ T\ |$ (المحسوبة)

فإننا نقبل H_0 ونرفض H_1 وعليه: $\mu_1 = \mu_2$

- **الحالة الثالثة:** $H_1 : \mu_1 > \mu_2$

(النظرية) $t\{1-\alpha, n1+n2-2\}$, $t\{0.95,12\}$, $t=1.782$

بما أن: $t = 1.782$ (الجدولية) $< 2.09 = |\ T\ |$ (المحسوبة)

فإننا نرفض H_0 ونقبل H_1 وعليه: $\mu_1 > \mu_2$

3-4 اختبار الفرضيات للنسبة:

إن اختبارات الفرضيات التي لها علاقة بالنسب مطلوبة في عدة مجالات فعلى سبيل المثال يهتم الصناعيون بمعرفة نسب التالف في إنتاجهم، وسنعتبر مسألة اختبار الفرضيات في نسبة النجاح لتجربة توزيع ثنائي الحد لقيم المشاهدات: $x_1, x_2, x_3, \ldots\ldots x_n$ مساوية لـــ: $H_0 : P=P_0$ حيث أن P_0 هي نسبة النجاح، وستكون الفرضية المقابلة إما:

- **$H_1 : P\neq P_0$:** فإننا نرفض H_0 إذا كانت: الجدولية $Z >$ | Z | (الحسابية)
- **$H_1 : P< P_0$:** فإننا نرفض H_0 إذا كانت: الجدولية $Z <$ | Z | (الحسابية)
- **$H_1 : P> P_0$:** فإننا نرفض H_0 إذا كانت: الجدولية $Z >$ | Z | (الحسابية)

حيث أن: $Z=\frac{P-P0}{\sqrt{\frac{P0(1-P0)}{n}}}$ (الحسابية)

مثال: يدعي صياد بأنه يصيب 75% من الطيور التي يطلق عليها النار فهل توافق هذا الادعاء، إذا كان في يوم ما قد أسقط 80 طيرا من أصل 120 طيرا أطلق النار عليها مستخدما $\alpha = 0.05$

الحل:

$$\begin{cases} H0: P = 0.75 \\ H1: P \neq 0.75 \end{cases}$$

$Z_{0.975}=1.96,\ P=\frac{80}{120} = 0.67$ (الجدولية)

$Z=\frac{P-P0}{\sqrt{\frac{P0(1-P0)}{n}}} = \frac{0.67-0.75}{\sqrt{\frac{0.75(0.25)}{120}}} = \frac{-0.18}{\sqrt{0.00156}} = \frac{-0.18}{0.04} = -4.5$ (الحسابية)

بما أن: $Z_{0.975}=1.96$ (الجدولية) $<4.5=$ | Z | **(الحسابية)**

فإننا نرفض H_0 وعليه فإن: $H_1: P \neq 0.75$