المحاضرة الأولى: توزيعات المعاينة

1- مقدمة:

العينة هي عبارة عن فئة أو مجموعة جزئية من المجتمع يتم اختيار وتحليل بياناتها وذلك بهدف تقدير معالم المجتمع غير المعلومة، أو اختبار فروض بشأنها، وبشكل عام لاستنباط معلومات عن معالم المجتمع المسحوب منه العينة، نقوم بحساب الوسط الحسابي للعينة على سبيل المثال ويكون هذا الوسط تقديرا لمتوسط المجتمع المجهول، ويسمى هذا التقدير: **إحصاءة** (statistic).

والقاعدة العامة تقول أن أي دالة في المتغيرات العشوائية المكونة لعينة المشاهدات تسمى **إحصاءة**.

وعليه إذا كانت لدينا عينة عشوائية $(X_1,X_2,\ldots\ldots\ldots\ldots, Xn)$ حجمها n فإن الوسط الحسابي لهذه العينة هو الإحصاءة $\overline{x}$ ، حيث:

$$\overline{x} = \Sigma Xi / n \quad \ldots\ldots\ldots\ldots(1)$$

بينما تباين العينة Sample variance هو الإحصاءة S^2 :

$$S^2 = \Sigma(x_i - \overline{x})^2 / n-1 \quad \ldots\ldots.(2)$$

و ما يجب الإشارة إليه، هو أنه في حالة عدم معرفة تباين المجتمع فإنه يتم استخدام تباين العينة S^2 كتقدير له، أما الانحراف المعياري للعينة S فهو عبارة عن الجذر التربيعي لتباين العينة S^2، ويستخدم أيضا كتقدير للانحراف المعياري للمجتمع σ .

كذلك استخدمنا $(n -1)$ في مقام المعادلـــة (2) بدلا من n وذلك لكي يكون التقدير الناتج تقديرا غير متحيزا (Unbiased Estimate) للمعلمة σ^2، أي يكون:

$$E(S^2) = \sigma^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

إن العلاقة (3) تكون صحيحة إذا كان المجتمع المسحوب منه العينة غير محدود (غير منته)، أو عندما يكون حجم المجتمع (N) كبير جدا. أما إذا كان المجتمع محدود ومكون من القيم التالية: $(X_1, X_2, \ldots\ldots\ldots\ldots X_N)$ يكون:

$$E(S^2) = \frac{N}{N-1}\sigma^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

مع العلم أن:

$$\sigma^2 = \Sigma(X_i - \mu)^2 / \mathrm{N} \ \ldots\ldots\ldots\ldots \ (5)$$

$$\mu = \Sigma X_i / \mathrm{N} \ \ldots\ldots\ldots\ldots\ldots\ldots \ (6)$$

2- المعاينة بإرجاع والمعاينة بدون إرجاع:

إذا كان لدينا صندوق يحتوي على سبعة بطاقات مرقمة من 1 إلى 7 وقمنا بسحب بطاقة من هذا الصندوق، فانه لدينا الخيار في إرجاع هذه البطاقة أو عدم إرجاعها قبل إجراء عملية السحب الثانية. ففي الحالة الأولى فان البطاقة يمكن أن تظهر عدة مرات، بينما في الحالة الثانية يمكن أن تظهر البطاقة مرة واحدة فقط، وبالتالي ففي العينات التي يمكن أن نختار فيها مفردات المجتمع أكثر من مرة تسمى **بالمعاينة بإرجاع**. بينما إذا كانت المفردة في المجتمع لا يمكن اختيارها أكثر من مرة فتسمى **المعاينة بدون إرجاع**.

وفي الحالة العامة إذا كان لدينا مجتمعا من المفردات يتبع توزيعا معينا، ونريد اختيار عينة حجمها n من هذا المجتمع، فانه يمكن اختيار هذه العينة وفق الطريقتين التاليتين (السحب مع الإرجاع والسحب بدون إرجاع)، حيث نجد أن عدد العينات التي سوف يتم سحبها من هذا المجتمع وفق الطريقتين السابقتين يتم تحديده كما يلي:

N^n في حالة السحب مع الإرجاع.

C_N^n في حالة السحب بدون إرجاع.

3- توزيع المعاينة:

بفرض أننا قمنا باختيار عينة عشوائية حجمها n من مجتمع معين، ثم بعد ذلك قمنا بحساب مقياسا معينا لهذه العينة وليكن على سبيل المثال الوسط الحسابي $\overline{\mathrm{x}}$. ثم اخترنا عينة ثانية لها نفس الحجم n وقمنا بحساب نفس المقياس السابقة، واخترنا عينة ثالثة وحسبنا منها المقياس نفسه، وهكذا بالنسبة لجميع العينات التي يمكن سحبها من هذا المجتمع، فانه سيتوفر لدينا عدد كبير من القيم للوسط الحسابي $\overline{\mathrm{x}}$ ، هذه القيم تكون مجتمعا آخر عدد مفرداته اكبر بكثير من عدد مفردات المجتمع الأصلي.

ومن هذا المنطلق يمكن النظر إلى هذا المقياس على انه متغير عشوائي يأخذ قيما مختلفة (طبعا هي القيم التي حصلنا عليها من هذه العينات)، ويتبع توزيعا معينا قد يختلف عن توزيع المجتمع الأصلي، ويسمى هذا التوزيع **بتوزيع المعاينة** لهذا المقياس سواء كان هذا المقياس هو الوسط الحسابي للعينة $\overline{\mathrm{x}}$ أو الانحراف المعياري لها S.

وبالتالي يمكن القول بأن توزيع المعاينة لأية إحصاءة من إحصاءات العينة، ($\overline{\mathrm{x}}$، S^2) هو التوزيع الاحتمالي لجميع القيم الممكنة لهذه الإحصاءة، والتي نحصل عليها عند سحب كل العينات بنفس الحجم والطريقة ومن نفس المجتمع.

مثال1: إذا كان لدينا مجتمع مكون من خمس وحدات (N=5)، وكانت قيم ظاهرة معينة لهذه الوحدات هي: 1.5 ، 3 ، 6 ، 4.5 ، 7.5

المطلوب:

1- أحسب الوسط الحسابي (μ) والتباين (σ^2).

2- أحسب الوسط الحسابي $\overline{x}$ لجميع العينات البسيطة الممكنة والتي حجم كل منها ثلاث وحدات، وكذلك حدد جـــــدول التوزيع الاحتمـــــــــالي للوسط الحسابي ($\overline{x}$)، ومنه أحسب القيمة المتوقعة والتباين للمتغير ($\overline{x}$).

3- أحسب التباين S^2 لجميع العينات العشوائية الممكنة التي حجم كل منها ثلاث وحدات واكتب التوزيع الاحتمالي للمتغير S^2 وتحقق من أن:

$$E(S^2) = \frac{N}{N-1}\sigma^2$$

الحل :

1- الوسط الحسابي وتباين المجتمع:

$$\mu = \Sigma X_i \ / \ N = (1.5 + 3 + 6 + 4.5 + 7.5) / 5 = 4.5$$

$$=\Sigma(X_i - \mu)^2 \ / \ N\sigma^2$$

$$\sigma^2 = [(1.5\text{-}4.5)^2 +(3\text{-}4.5)^2 + (6\text{-}4.5)^2 +(4.5\text{-}4.5)^2 + (7.5\text{-}4.5)^2] / 5$$

$$\sigma^2 = 4.5$$

2- بما أن السحب تم بدون إرجاع فإن العينات الممكنة عددها عشرة وذلك حسب:

$$C_N^n = C_5^3 = \frac{5!}{3!2!} = 10$$

والجدول الموالي يوضح مختلف هذه العينات وأوساطها الحسابية.

الجدول (1) العينات العشرة الممكنة والوسط الحسابي المقابل لكل منهما

رقم العينة	القيم المختلفة للعينة x_i (x_1, x_2, x_3)	$\sum x_i$	$\overline{x}_i$
01	(1.5, 3, 6)	10.5	3.5
02	(1.5, 3, 4.5)	9	3
03	(1.5, 3, 7.5)	12	4

04	(1.5, 6, 4.5)	12	4
05	(1.5, 6, 7.5)	15	5
06	(1.5, 4.5, 7.5)	13.5	4.5
07	(3, 6, 4.5)	13.5	4.5
08	(3, 6, 7.5)	16.5	5.5
09	(3, 4.5 ,7.5)	15	5
10	(6, 4.5, 7.5)	18	6

من خلال هذا الجدول فإننا نستطيع تحديد جدول التوزيع الاحتمالي للوسط الحسابي ($\overline{x}$) لعينة عشوائية حجمها ثلاث وحدات، وذلك كما يلي:

الجدول (2): التوزيع الاحتمالي للوسط الحسابي ($\overline{x}$) لعينة عشوائية حجمها ثلاث وحدات.

$\overline{x}_i$	3	3.5	4	4.5	5	5.5	6
$P(\overline{x}_i)$	1/10	1/10	2/10	2/10	2/10	1/10	1/10

انطلاقا من هذا الجدول فإن القيمة المتوقعة للوسط الحسابي ($\overline{x}$) هي:

$$E(\overline{x}) = \Sigma \overline{x}_i\, P(\overline{x}_i) \text{ ....(7)}$$

$$E(\overline{x})=3(1/10)+3.5(1/10)+4(2/10)+4.5(2/10)+5(2/10)+5.5(1/10)$$

$$+6(1/10) = 4.5$$

مما سبق نستنتج أن: $\mathbf{E(\overline{x}) = \mu = 4.5}$

أما تباين الوسط الحسابي للعينة فيحسب بالعلاقة التالية:

$$Var(\overline{x}) = E(\overline{x})^2 - [E(\overline{x})]^2 \text{.........(8)}$$

حيث:

$$E(\overline{x})^2 = \Sigma \overline{x}_i^{\,2} \,.\, P(\overline{x}_i)$$

$$E(\overline{x})^2=3^2(1/10)+3.5^2(1/10)+4^2(2/10)+4.5^2(2/10)+5^2(2/10)+$$
$$(5.5)^2(1/10)+6^2(1/10) = 21$$

$$\Rightarrow Var(\overline{x}) = 21-(4.5)^2 = \mathbf{0.75}$$

3- حساب التباين S^2 لجميع العينات العشوائية الممكنة: ويتم ذلك كما يلي:

$$S^2 = \Sigma(x_i - \overline{x})^2 / n-1$$

$$= \Sigma(x_i - \overline{x})^2 / 3-1$$

يمكن تلخيص قيم S^2 المناظرة لكل عينة من العينات العشرة الممكنة في الجدول (3).

الجدول (3): قيم S^2 المناظرة للعينات العشرة

رقم العينة	القيم المختلفة للعينة (x_1, x_2, x_3)	$\overline{x}_i$	$\Sigma(x_i - \overline{x})^2$	S^2
1	(1.5, 3, 6)	3.5	10.5	5.25
2	(1.5, 4.5, 3)	3	4.5	2.25
3	(1.5, 3, 7.5)	4	19.5	9.75
4	(1.5, 6, 4.5)	4	10.5	5.25
5	(1.5, 6, 7.5)	5	19.5	9.75
6	(1.5, 4.5 ,7.5)	4.5	18	9
7	(3, 6, 4.5)	4.5	4.5	2.25
8	(3, 6, 7.5)	5.5	10.5	5.25
9	(3, 4.5, 7.5)	5	10.5	5.25
10	(6, 4.5, 7.5)	6	4.5	2.25

وانطلاقا من هذا الجدول فإنه يمكننا الحصول على التوزيع الاحتمالي للمتغير S^2 كما هو موضح في الجدول (4) التالي:

الجدول (4): التوزيع الاحتمالي للمتغير S^2

S^2	2.25	5.25	9	9.75
$P(S^2)$	3/10	4/10	1/10	2/10

إذن من خلال الجدول (4) نجد:

$$E(S^2) = \sum S^2 . P(S^2) \text{..} (9)$$

$$= 2.25\left(\frac{3}{10}\right) + 5.25\left(4/10\right) + 9(1/10) + 9.75(2/10)$$

$$= 5.625$$

التحقق من أن: $E(S^2) = \frac{N}{N-1}\sigma^2$

من المعادلة (4) نجد: $\frac{N}{N-1}\sigma^2 = [\ 5 /(5-1)\ ]4.5 = 5.625$

إذن نستنتج أن: $\mathbf{E(S^2) = \frac{N}{N-1}\sigma^2}$

المحاضرة الثانية: توزيعات المعاينة

1-3. توزيع المعاينة للوسط الحسابي $\bar{x}$

1-1-3. توزيع المعاينة للوسط الحسابي $\bar{x}$ عند المعاينة من مجتمع طبيعي

بفرض أنه أخذنا عينة عشوائية $(x_1, x_2, \ldots\ldots, x_n)$ من مجتمع ما، وكان $\bar{x}$ هو الوسط الحسابي لهذه العينة. **فما هو توزيع** $\bar{x}$ ؟

للإجابة على هذا السؤال يجب التطرق إلى ما يلي:

1- التوزيع الاحتمالي للوسط الحسابي لعينة عشوائية مسحوبة من مجتمع محدود (حالة السحب دون إرجاع):

إذا كان لدينا مجتمع مكون من عدد محدود من القيم $(x_1, x_2, \ldots \ldots x_N)$ فإن الوسط الحسابي والتباين لهذا المجتمع يكونان على الشكل التالي:

$$\mu = \Sigma X_i \,/\, N$$

$$\mathrm{Var}(X) = \sigma^2 = \Sigma(X_i - \mu)^2 \,/\, N$$

وإذا كان $\bar{x}$ هو الوسط الحسابي لعينة عشوائية حجمها n مسحوبة من هذا المجتمع، فان القيمة المتوقعة والتباين للمتغير $\bar{x}$ هي:

$$\mathbf{E}(\bar{x}) = \mu_{\bar{X}} = \mu \qquad \ldots\ldots\ (10)$$

$$\mathbf{Var}\,(\bar{x}) = \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots(11)$$

ومنه فالإنحراف المعياري للمتغير $\bar{x}$ هو:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

المقدار $\left(\sqrt{\frac{N-n}{N-1}}\right)$ يسمى بمعامل تصحيح المجتمع المحدود، وهو دائما أقل من الواحد، وكلما اقترب من الواحد فيمكن الاستغناء عنه. وعادة يستعمل معامل التصحيح إذا تحقق الشرط:

$\mathbf{n \geq \% \ 5\ N}$

2- التوزيع الاحتمالي للوسط الحسابي $\overline{X}$ لعينة عشوائية حجمها n مسحوبة من مجتمع غير محدود (حالة السحب بإرجاع):

إذا كانت $(x_1,x_2......,x_n)$ هي مشاهدات عينة عشوائية مسحوبة من مجتمع غير محدود متوسطه μ وتباينه σ^2، فإن القيمة المتوقعة والتباين للمتغير $\overline{X}$ هما:

$$E(\overline{X}) = \mu_{\overline{X}} = \mu \quad .... (12)$$

$$Var(\overline{X}) = \sigma^2_{\overline{X}} = \frac{\sigma^2}{n} \quad ....(13)$$

من المعادلة (13) يمكن أن نجد الانحراف المعياري للوسط الحسابي $\overline{X}$ كما يلي:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

إن الانحراف المعياري للوسط الحسابي (أي: $\frac{\sigma}{\sqrt{n}}$) يسمى أيضا بالخطأ المعياري للوسط الحسابي، **وينقص هذا الخطأ كلما زاد حجم العينة**. ومن ثم فإنه يُتوقع أن تقترب $\overline{X}$ من μ كلما زاد حجم العينة n.

نظرية (1): إذا كانت $(x_1,x_2......,x_n)$ هي مشاهدات عينة عشوائية مسحوبة من مجتمع طبيعي وسطه μ وتباينه σ^2، فإن الوسط الحسابي $\overline{X}$ له توزيع طبيعي متوسطه μ ، وتباينه $\frac{\sigma^2}{n}$ ، أي أن المتغير العشوائي:

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad .... \quad (14)$$

يخضع للتوزيع الطبيعي المعياري وذلك مهما كان حجم العينة.

مثال(2): أخذت عينة عشوائية حجمها 9 من الأطفال حديثي الولادة في إحدى مستشفيات الوطن، فإذا علم أن وزن الطفل حديث الولادة يخضع للتوزيع الطبيعي ذو الوسط 2900 غ، وانحرافه المعياري 600غ.

المطلوب:

1- أوجد الوسط الحسابي والتباين والانحراف المعياري للوسط الحسابي لأوزان الأطفال في العينة.
2- أوجد احتمال أن الوسط الحسابي لأوزان الأطفال يزيد عن 3100غ.
3- أوجد احتمال أن الوسط الحسابي لأوزان الأطفال يقع ما بين 2700غ و 3200غ .

الحل:

بفرض أن $\overline{X}$ هو الوسط الحسابي لأوزان الأطفال في العينة.

1- لدينا: $\mu_{\overline{X}} = 2900g \ , \ \sigma^2 = (600)^2 = 360000$

$$\frac{\sigma^2}{n} = \frac{360000}{9} = 40000g\sigma^2{}_{\overline{X}} =$$

وعليه يكون:
$$\sigma_{\overline{X}} = \frac{600}{\sqrt{9}} = 200g$$

2- إيجـاد الاحتمـال التالـي:
$$P(\overline{X} > 3100)$$

نقوم أولا بحساب المتغير العشوائي Z:

$$= \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{3100 - 2900}{\frac{600}{\sqrt{9}}} = 1 Z_{3100}$$

ومنه نجد:

$$P(\overline{X} > 3100) = P(Z > 1) = 0.5 - P(0 < Z \leq 1)$$

$$= 0.5 - 0.3413 = 0.1587$$

3- إيجـاد الاحتمـال التالـي:

$$P(2700 < \overline{X} < 3200)$$

$$P(2700 < \overline{X} < 3200) = P\left(\frac{2700 - 2900}{200} < Z < \frac{3200 - 2900}{200}\right)$$

$$= P(-1 < Z < 1.5)$$

$$= P(0 < Z \leq 1.5) + P(-1 \leq Z < 0)$$

$$= 0.4332 + 0.3413 = 0.7745$$

3-1-2 . توزيع المعاينة للوسط الحسابي $\overline{X}$ عند المعاينة من مجتمع غير طبيعي:

إذا كان لدينا مجتمع وسطه μ وتباينه σ^2 معلوم، وأخذت منه جميـع العينات العشوائية ذات الحجم n، فما هو توزيع الوسط الحسابي $\overline{X}$ لهذه العينات حتى لم يكن توزيع المجتمع التوزيع الطبيـــعي ؟

للإجابة عن هذا السؤال يجب التطرق إلى النظرية التـالية:

نظرية (2): نظرية النهائية المركزية (نظرية تقارب التوزيعات):

إذا كانت $(x_1,x_2......,x_n)$ هي مشاهدات عينة عشوائية حجمها n أخذت من مجتمع إحصائي وسطه μ وتباينه σ^2، فإن توزيع المعاينة للوسط الحسابي $\overline{X}$ يقترب من التوزيع الطبيعي بمتوسط μ وتباين $\sigma^2_{\overline{X}}$ كلما زاد حجم العينة تدريجيا، ويكفي لذلك أن يصل حجم العينة إلى 30؛ بعبارة أخرى فإن المتغير (Z) حيث:

$$\mathbf{Z} = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

يقترب من التوزيع الطبيعي المعياري كلما كان حجم العينة أكبر أو يساوي 30.

تعتبر نظرية النهاية المركزية واحدة من أهم النظريات في النواحي التطبيقية للإحصاء.

مثال (3): مجتمع كبير متوسطه 75 وانحرافه المعياري 13, سحبت منه عينة عشوائية بسيطة حجمها 51.

المطلوب:

1- أحسب إحتمال أن يكون متوسط العينة أصغر من 78.
2- أحسب إحتمال أن لا يبعد متوسط العينة عن متوسط المجتمع بأكثر من 4%.

الحل:

$$\mu_{\overline{X}} = \mu = 75$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{13}{\sqrt{51}} = 1.82$$

1- حساب الاحتمال التالي: $\boldsymbol{P}(\overline{X} < 78)$

نقوم أولا بحساب المتغير العشوائي Z:

$$Z_{78} = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{78-75}{1.82} = 1.65$$

ومنه يكون:

$$P\left(\overline{X} < 78\right) = P(Z < 1.65) = 0.5 + P(0 < Z \leq 1.65) = 0.5 + 0.4505 = 0.9505$$

2 - حساب إحتمال أن لا يبعد متوسط العينة عن متوسط المجتمع بأكثر من 4 %

لدينا: $\% 4 \times 75 = 3$

إذن المراد هو حساب الاحتمال التالي:

$$P(72 < \bar{x} < 78) = P\left(\frac{72-75}{1.82} < Z < \frac{78-75}{1.82}\right)$$

$$= P(-1.65 < Z < 1.65)$$

$$= 2P(0 < Z \leq 1.65)$$

$$= 0.901$$

3-1-3. توزيع المعاينة لـ $\bar{x}$ عندما يكون الانحراف المعياري للمجتمع)σ(غير معلوم: مقدمة لـتوزيــع T

في الكثير من التطبيقات المعلقة باستعمال الوسط الحسابي $\bar{x}$ ، نجد أن الانحراف المعياري للمجتمع لا يكون معلوما، وبالتالي يكون المتغير (Z) دالة في مؤشر غير معلوم ، ومن ثم لا يمكن تحديد قيمة (Z) لعينة محددة. وبالتالي يبدو أن هذا سوف يخلق مشكلة.

لمعالجة هذه المشكلة نقوم باستبدال σ بالتقدير S في المعادلة (14) وبالتالي نحصل على الإحصاءة **T** ، حيث:

$$T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \quad \text{............ ....} \quad (15)$$

ويكون **T** خاضع لتوزيع **ستيودنت (t)** على درجات حرية (n-1)

إن توزيع **T** يشبه التوزيع الطبيعي المعياري من حيث أنه مُتماثل ومُركز حول الصفر، ولكنه أكثر تشتتا واختلافا، وبالتالي إلى أي مدى يزيد التشتت عندما نستبدل σ بـ S ؟ هذا التشتت يعتمد على حجم العينة؛

- فإذا كانت n كبيرة بدرجة كافية أي $(n \geq 30)$ فإن S تكون تقديرا دقيقا جدا لـ σ ويكون التشتت في **T** قليل جدا.

- أما إذا كانت n صغيرة جدا فإن S تصبح تقديرا غير دقيقا لـ σ، ومنه تُظهر **T** تباينا أكثر، وعليه فإن التشتت في توزيع **T** يعتمد على حجم العينة n.

أي أنه كلما زادت n فإن توزيع **T** يُظهر تشتت أقل وأقل ويصبح مشابها أكثر فأكثر للتوزيع الطبيعي المعياري، وفي الحقيقة فإنهما يُصبحان متطابقين من الناحية النظرية كلما كبرت n واقتربت من الما لا نهاية.

وهذا يعني أنه إذا كانت n كبيرة بدرجة كافية فإن التوزيع الطبيعي المعياري يُعد تقريبا جيدا لتوزيـــــــــع **T**، ويُمكـــن أن يستخدم بدلا منه إذا شئنا ذلك، حيث نجد أن القاعدة المقبولة على نطاق واسع أن التقريب يعد مقبولا إذا كانت $n \geq 30$. أما إذا كان حجم العينة ليس كبير بدرجة كافية أي (n<30) فهذا ما سوف نعالجه في النظرية التالية:

نظرية 03: إذا أخذت عينة عشوائية من مجتمع يخضع للتوزيع الطبيعي $N\left(\mu,\sigma^2\right)$ وكان σ^2 غير معلوم، وحجم العينة أقل من 30، وكان $\bar{X}$ هو الوسط الحسابي لهذه العينة، و S هو إنحرافها المعياري فإن توزيع المعاينة للمتغير **T**، حيث:

$$T = \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$$

هو توزيع **ستيودنت** بدرجات حرية $(n-1)$.

مثال (4): وكالة لحماية البيئة حددت متوسطا لمعدل الأميال / جالون على الطرق السريعة قدره 45 وذلك لنوع معين من السيارات. اشترت منظمة مستقلة للمستهلكين إحدى هذه السيارات واختبرتها لتتحقق من معدل الوكالة، وتم ذلك بقيادة السيارة لمسافة 100 ميل في 25 رحلة مختلفة، حيث سجلت القيم الفعلية للأميال المقطوعة لكل جالون في كل رحلة. ومن خلال 25 رحلة هاته حسب المتوسط والانحراف المعياري فكان: 43.5 ، 2.5 ميل/جالون على التوالي.

وهناك اعتقاد بأن التوزيع الفعلي للأميال / جالون على الطرق السريعة لهذا النوع من السيارات يقترب من التوزيع الطبيعي.

المطلوب:

1- مفترضا أن معدل الوكالة (45 ميل/جالون) متحققا لهذه السيارة، أوجد احتمال أن متوسط الأميال/جالون في العينة يجب أن يكون 43.5 أو أقل ؟

2- اعتمادا على بيانات العينية الحالية، هل هناك سبابا مقنعا للمنظمة لكي تشك في أن معدل الوكالة متحققا لهذه السيارة.

الحل:

لدينا:

$$\mu = 45 \quad , \quad s = 2.5 \quad , \quad \bar{X} = 43.5 \quad , \quad n = 25$$

$$\mu_{\bar{X}} = \mu = 45$$

$$\sigma_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{2.5}{\sqrt{25}} = 0.5$$

1- حساب الاحتمال التـالي: $P\left(\bar{X} \le 43.5\right)$

بما أن الانحراف المعياري للمجتمع σ غير معلوم، وحجم العينة أقل من الثلاثون يكون من البديهي أن نتجه إلى حساب T، أي أن توزيع المعاينة للوسط الحسابي $\bar{X}$ في هذه الحالة يخضع لتوزيع **ستيودنت**.

حيث:

$$T = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{43.5 - 45}{0.5} = -3 \quad , \quad \upsilon = 24$$

ومنه:

$$P(T \leq -3) = P(\overline{x} \leq 43.5) = 0.005$$

وبالتالي نجد:

$$P(T > -3) = P(\overline{x} > 43.5) = 0.995$$

2- اعتمادا على الإجابة في السؤال الأول فإنه من البديهي أن نشك في معدل الوكالة، ومع ذلك وقبل أن نلوم الوكالة لمعدلها المرتفع وغير المناسب، فإن فكرة إجراء أبحاث إضافية هو أمر جيد، والتعارض المشاهد ربما يكون ببساطة نتيجة للفروق بين طريقتي القياس للأميال في المنظمتين (أي وكالة لحماية البيئة ومنظمة المستهلكين).

المحاضرة الثالثة: توزيعات المعاينة

3-2 . توزيع المعاينة للفرق بين متوسطي عينتين مستقلتين

3-2-1. حالة المعاينة من مجتمعين طبيعيين ذو تباينين معلومين

نظرية (4):

إذا كان $\bar{X}_1$, $\bar{X}_2$ هما متوسطا عينتين عشوائيتين مستقلتين حجمها على التوالي $n_1 n_2$, تم سحبهما من مجتمعين طبيعيين ذو متوسطين μ_1, μ_2 وتباينين σ_1^2 , σ_2^2 معلومين على الترتيب، فإن الفرق $(\bar{X}_1 - \bar{X}_2)$ يكون له توزيع طبيعي متوسطه وتباينه يأخذان الشكل التالي:

$$E(\bar{X}_1 - \bar{X}_2) = \mu_{\bar{X}_1 - \bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = \boldsymbol{\mu_1 - \mu_2} \ldots\ldots \quad (16)$$

$$V(\bar{X}_1 - \bar{X}_2) = \sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \ldots\ldots \frac{\sigma_2^2}{n_2} \quad (17)$$

وبالتالي فإن الانحراف المعياري للفرق $(\bar{X}_1 - \bar{X}_2)$ هو:

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

ومن ثم يكون المتغير Z الذي يساوي:

$$Z = [(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)] \Big/ \sqrt{\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}} \quad \ldots\ldots\ldots(18)$$

له توزيع طبيعي معياري $N(0,1)$ وذلك مهما كان حجم العينتين.

ملاحظة: قد يكون من المفيد أحيانا الحديث عن توزيع المعاينة لمجموع إحصائتين مثلا $(\bar{X}_1 + \bar{X}_2)$ فإن الوسط الحسابي والتباين لهذا النوع من التوزيع يعطى بالصيغتين التاليتين:

$$\mu_{\bar{X}_1 + \bar{X}_2} = \mu_{\bar{X}_1} + \mu_{\bar{X}_2} = \boldsymbol{\mu_1 + \mu_2} \ldots\ldots\ldots\ldots\ldots\ldots \quad (19)$$

$$= \frac{\sigma_1^2}{n_1}\sigma^2_{\bar{X}_1+\bar{X}_2} + \quad \ldots\ldots\ldots\ldots\ldots \frac{\sigma_2^2}{n_2} (20)$$

مثال (5): إذا كانت رواتب المعلمين في وزارة التربية والتعليم تخضع لتوزيع طبيعي وسطه 33000 دج، وانحرافه المعياري 5165 دج، ورواتب المعلمين في المدارس الخاصة تخضع لتوزيع طبيعي وسطه 25740 دج، وانحرافه المعياري 5663 دج.

أخذت عينة عشوائية من المعلمين في الوزارة حجمها 16 معلما وعبرنا عن وسطها الحسابي بالرمز $\bar{X}_1$، وأخذت عينة عشوائية من معلمي المدارس الخاصة حجمها 10 معلمين وعبرنا عن وسطها الحسابي بالرمز $\bar{X}_2$.

المطلوب: أوجد احتمال أن يزيد $\bar{X}_1$ عن $\bar{X}_2$ بمقدار 8000 دج.

الحل:

نريد إيجاد الاحتمال التالي:

$$\boldsymbol{P}\left(\bar{X}_1-\bar{X}_2\geq 8000\right)$$

بتطبيق النظرية (4) نجد :

$$Z = [(\bar{X}_1 - \bar{X}_2)-(\mu_1 - \mu_2)] / \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

$$Z = [(8000)-(33000 - 25740)] / \sqrt{\frac{5165^2}{16}+\frac{5663^2}{10}}$$

$$Z = 0.33$$

ومنه نجد:

$$P(Z \geq 0.33) = 0.5 - 0.1293 = 0.3707 \; P\left(\bar{X}_1-\bar{X}_2\geq 8000\right) =$$

2-2-3. حالة المعاينة من مجتمعين غير طبيعيين وذو تباينين معلومين

نظرية (5): إذا كان $\bar{X}_1$ هو الوسط الحسابي لعينة حجمها n_1 أخذت من مجتمع وسطه μ_1 وتباينه σ_1^2 معلوم، وكان $\bar{X}_2$ هو الوسط الحسابي لعينة حجمها n_2 أخذت من مجتمع آخر

مستقل عن الأول وسطه μ_2 وتباينه σ_2^2 معلوم أيضا، وكان حجم العينتين كبير بدرجة كافية، فان توزيع المعـــــاينة لـ $(\bar{x}_1 - \bar{x}_2)$ يخضع تقريبا للتوزيع الطبيعي بمتوسط وتباين تم تحديدهما في المعادلتين (16) و(17) على الترتيب، وعليه فإن توزيع المتغير **Z** حيث:

$$Z = [(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)] / \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

هو التوزيع الطبيعي المعياري.

مثال(6): إذا كان الإنفاق الاستهلاكي العائلي اليومي في ولاية باتنة وسطه الحسابي 1500دج وتباينه 600 دج، وكان الإنفاق الاستهلاكي العائلي اليومي في ولاية بسكرة وسطه الحسابي 1000دج وتباينه 450 دج، فإذا سحبنا من ولاية باتنة عينة عشوائية حجمها 150 عائلة، ومن ولاية بسكرة عينة عشوائية حجمها 100 عائلة، وكانت العينتان مستقلتان والمجتمعين غير خاضعين للتوزيع الطبيعي.

المطلوب: أوجد احتمال أن يكون الفرق بين متوسطي العينتين أكثر من 510 دج.

الحل:

بما أن تبايني المجتمعين معلومين، وحجم العينتين كبير بدرجة كافية، فإن المتغير $(\bar{x}_1 - \bar{x}_2)$ سيتوزع توزيعا قريبا من التوزيع الطبيعي بالرغم من أن توزيع المجتمعين غير طبيعي.

ومنه المطلوب هو إيجاد الاحتمال التالي:

$$\boldsymbol{P}(\bar{X}_1 - \bar{X}_2 > 510)$$

بتطبيق النظرية (5) نجد:

$$Z = [(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)] / \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$Z = [(510) - (1500 - 1000)] / \sqrt{\frac{600}{150} + \frac{450}{100}}$$

$$Z = 3.43$$

$$P(\bar{X}_1 - \bar{X}_2 > 510) = P(Z > 3.43) = 0.5 - 0.4997$$

$$= 0.0003$$

المحاضرة الرابعة: توزيعات المعاينة

3-2-3. توزيع المعاينة للفرق $(\bar{X}_1 - \bar{X}_2)$ عندما يكون تبايني المجتمعين مجهولين والعينتان مستقلتان وصغيرتا الحجم:

إن قيمة التباينات في المجتمعات غالبا تكون مجهولة، وعليه عند تقدير الانحراف المعياري للفرق $(\bar{X}_1 - \bar{X}_2)$ علينا أن نستخدم تبايني العينتين، بالإضافة إلى ذلك إذا كانت العينتين مستقلتين وصغيرتا الحجم فإن توزيع المعاينة لـ $(\bar{X}_1 - \bar{X}_2)$ ليس له توزيع طبيعي ،بل له توزيع **ستيودنت** (t) بدرجات حرية υ.

وبالتالي، عند دراسة توزيع المعاينة للفرق $(\bar{X}_1 - \bar{X}_2)$ لما يكون تبايني المجتمعين مجهولين وحجم العينتين صغير لدينا حالتين هما:

1- تبايني المجتمعين متساويين.
2- تبايني المجتمعين غير متساويين.

1-حالة تبايني المجتمعين متساويين ومجهولين

ليكن σ_1^2 هو التباين الخاص بالمجتمع الأول وهو مجهول القيمة، وكان σ_2^2 هو التباين الخاص بالمجتمع الثاني وهو مجهول القيمة أيضا، وكان $\sigma_1^2 = \sigma_2^2$ ، فإن التباين σ^2 يمثل التباين المشترك لهما، وبالتالي فهو مجهول أيضا، حيث نضع:

$$\sigma^2 = \sigma_1^2 = \sigma_2^2$$

ومنه تكون صيغة الانحراف المعياري للفرق $(\bar{X}_1 - \bar{X}_2)$ على الشكل التالي:

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}} = \sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \quad \ldots. (21)$$

وبما أن التباين المشترك σ^2 مجهول القيمة فإننا نستطيع تقديره باستخدام تبايني العينيتين $(s_1^2 ، s_2^2)$ كما في الصيغة الموالية، حيث يكون تقدير التباين هو متوسط مرجع للقيم s_1^2 و s_2^2 وتكون الترجيحات مبنية على أساس حجم العينات. ولكي يكون تقدير التباين تقديرا غير متحيز لـ σ^2 فإننا يجب أن نستخدم درجات الحرية $(n_1 - 1)$ و $(n_2 - 1)$ كترجيحات بدلا من إستخدام العينيتين: n_1 ، n_2 بشكل مباشر.

وبناء على ذلك فإن مقدر التباين المشترك (σ^2) هو s_p^2 ويُحسب بالصيغة التالية:

$$s_p^2 = \frac{(n_1-1)\, s_1^2+(n_2-1)\, s_2^2}{(n_1-1)+(n_2-1)}$$

ومنه يكون:

$$s_p^2 = \frac{(n_1-1)\, s_1^2+(n_2-1)\, s_2^2}{n_1+ n_2 -2}$$

حيث يسمى الإحصاء s_p^2 **بتباين العينة التجميعي** "pooled sample variance" وذلك لأنه مكون من تجميع المعلومات عن العينتين معا.

وتأسيسا على ما سبق، يكون تقدير الخطأ أو الانحراف المعياري للفرق $(\bar{x}_1 - \bar{x}_2)$ على الشكل التالي:

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (22)$$

نظرية (06): إذا كان $\bar{x}_1$ هو الوسط الحسابي لعينة عشوائية حجمها n_1 سحبت من مجتمع وسطه μ_1 وتباينه σ_1^2، وكان $\bar{x}_2$ هو الوسط الحسابي لعينة عشوائية حجمها n_2 سحبت من مجتمع آخر مستقل عن المجتمع الأول وسطه μ_2 وتباينه σ_2^2، وكان حجم العينتين صغير وتبايني المجتمعين مجهولين ومتساويين، فإن المتغير:

$$[(\bar{x}_1 - \bar{x}_2)-(\mu_1 - \mu_2)] \Big/ \sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \quad \ldots\ldots (23)$$

يخضع لتوزيع **ستيودنت (t)** بدرجة حرية: $\upsilon = n_1 + n_2 - 2$

مثال (07): سحبت عينة عشوائية حجمها 16 وحدة من مجتمع طبيعي وسطه 30 وتباينه σ_1^2 مجهول، وسحبت أيضا عينة عشوائية حجمها 25 وحدة من مجتمع طبيعي آخر مستقل عن الأول وسطه 28 وتباينه σ_2^2

مجهول أيضا. وكان $\bar{x}_1$ و $\bar{x}_2$ هما الوسطين الحسابيين للعينتين الأولى والثانية على الترتيب، وتباين العينة الأولى هو 4، وتباين العينة الثانية هو 7.

المطلوب: إذا كان تبايني المجتمعين متساويين، أوجد احتمال أن الفرق بين متوسطي العينتين يكون أقل من 3.

الحل:

لدينا:

$$n_1 = 16 \text{ ، } n_2 = 25 \text{، } s_1^2 = 4 \text{ ، } s_2^2 = 7$$
$$= 30\mu_1 \text{ ، } = 28 \quad \mu_2$$

بما أن تبايني المجتمعين مجهولين ومتساويين فإننا نستخدم توزيع ستيودنت t بدرجة حرية $(n_1 + n_2 - 2)$ وذلك لإيجاد الاحتمال التالي:

$$P(\bar{X}_1 - \bar{X}_2 < 3)$$

لدينا:

$$= P\left(\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} < \frac{(3) - (30 - 28)}{\sqrt{s_p^2(\frac{1}{16} + \frac{1}{25})}}\right) P(\bar{X}_1 - \bar{X}_2 < 3)$$

الآن نقوم بحساب s_p^2 ، حيث:

$$s_p^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}$$

$$s_p^2 = \frac{(16-1)4 + (25-1)7}{16+25-2} = 5.84$$

ومنه يكون:

$$= P\left(T < \frac{1}{\sqrt{5.84(\frac{1}{16} + \frac{1}{25})}}\right) = P(T < 1.30) \boldsymbol{P}(\bar{X}_1 - \bar{X}_2 < 3)$$

عند درجات الحرية: $(n_1 +_2 - 2) = 16+25\text{-}2 = 39$

نجـد: $\boldsymbol{P}(\bar{X}_{\mathbf{1}} - \bar{X}_{\mathbf{2}} < 3) = 0.9$

2- حالة تبايني المجتمعين غير متساويين ومجهولين:

نظرية (7): إذا كان $\bar{X}_1$ و $\bar{X}_2$ هما الوسطين الحسابين لعينتين مستقلتين صغيرتا الحجم تم سحبهما من مجتمعين ذو متوسطين μ_1 و μ_2، وذو تباينين مجهولين وغير متساويين فإننا نستطيع تقدير الانحراف المعياري للفرق $(\bar{X}_1 - \bar{X}_2)$ باستخدام تبايني العينتين مباشرة $(s_1^2 \text{ و } s_2^2)$ وذلك كما يلي:

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \quad \ldots\ldots\ldots\ldots (24)$$

وعليه فإن المتغير:

$$\left(\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (25)$$

يخضع تقريبا لتوزيع **ستيودنت** بدرجة حرية لها الصيغة المركبة التالية :

$$\upsilon = \left(\frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{(n_1-1)}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{(n_2-1)}}\right) \quad \ldots\ldots\ldots\ldots\ldots (26)$$

مثال(08): إذا كانت نقاط طلبة السنة ثانية (ل.م.د) إدارة أعمال في مقياس الإحصاء لديها وسط حسابي قدره 15، وكانت نقاط طلبة السنة ثانية (ل.م.د) محاسبة في نفس المقياس لديها وسط حسابي قدره 10. وقمنا بسحب عينة من طلبة السنة ثانية (ل.م.د) إدارة أعمال حجمها 25 طالب، وقمنا أيضابسحب عينة من طلبة السنة ثانية (ل.م.د) محاسبة حجمها 20 طالبا. فإذا كان تباين العينة الأولى هو 6 وتباين العينة الثانية هو 4.

المطلوب: أوجد احتمال أن يكون الفرق بين متوسطي العينتين أكثر من 6، هذا إذا كان تبايني المجتمعين مجهولين وغير متساويين.

الحل:

لدينا:

$$n_1 = 25 \text{ ، } n_2 = 20 \text{ ، } s_1^2 = 6 \text{ ، } s_2^2 = 4$$
$$\mu_1 = 15 \text{ ، } \mu_2 = 10$$

بما أن تبايني المجتمعين مجهولين وغير متساويين فإننا نستخدم توزيع t بدرجة حرية لها الصيغة المركبة الموضحة في العلاقة (26) وذلك لإيجاد الاحتمال التالي:

$$P(\bar{X}_1 - \bar{X}_2 > 6)$$

ومنه يكون:

$$= P\left(\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} > \frac{(6) - (15 - 10)}{\sqrt{\frac{6}{25} + \frac{4}{20}}}\right) P(\bar{X}_1 - \bar{X}_2 > 6)$$

$$= P(T > 1.51)$$

عند درجات الحرية:

$$\upsilon = \left(\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{(n_1 - 1)} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{(n_2 - 1)}}\right) = \left(\frac{\left(\frac{6}{25} + \frac{4}{20}\right)^2}{\frac{\left(\frac{6}{25}\right)^2}{(25 - 1)} + \frac{\left(\frac{4}{20}\right)^2}{(20 - 1)}}\right) = 43$$

نجد:

$$= 0.05 P(\bar{X}_1 - \bar{X}_2 > 6)$$

3-3. توزيع المعاينة لتباين العينة S^2

نعلم جيدا أن S^2 تقيس الاختلافات ومن ثم فهي تدل على التشتت بين القيم في عينة عشوائية، حيث أن التباين يعتبر من أهم المقاييس الهامة مثله في ذلك مثل مقاييس النزعة المركزية، وبالتالي فإن أهمية S^2 للاستدلال عن σ^2 تُضاهي أهمية $\bar{X}$ عند الاستدلال عن μ. وفيما يلي سنحدد الوسط الحسابي والانحراف المعياري للتباين S^2، ثم نبحث عن توزيع المعاينة لـ S^2.

1-3-3. الوسط الحسابي والانحراف المعياري لتباين العينة

بناء على ما سبق فإن تباين العينة (S^2) والتي حجمها n هو:

$$S^2 = \frac{\Sigma\left(x_i - \bar{X}\right)^2}{n - 1}$$

والقيمة المتوقعة لتباين العينة (متوسط القيم لجميع العينات الممكنة ذات الحجم n) هو:

$$E(S^2) = \sigma^2$$

أما الانحراف المعياري لتباين العينة فيعطى بالصيغة التالية:

$$\sigma_{S^2} = \sigma^2 . \sqrt{\frac{2}{n-1}} \;\ldots\ldots\ldots\ldots\ldots\ldots\; (27)$$

مثال (09): إذا كان لدينا مجتمع كبير جدا وسطه الحسابي هو 50 وانحرافه المعياري هو 0.5

المطلوب: أحسب القيمة المتوقعة والانحراف المعياري للتباين S^2 لجميع العينات العشوائية الممكنة ذات الحجم 10 والتي يمكن سحبها من هذا المجتمع.

الحل:

$$E(S^2) = (0.5)^2 = 0.25$$

$$\sigma_{S^2} = \sigma^2 . \sqrt{\frac{2}{n-1}} = 0.25 . \sqrt{\frac{2}{10-1}} = 0.1178$$

2-3-3. توزيع المعاينة لـ S^2: مقدمة لتوزيع المعاينة للمتغير χ^2 (كآي تربيع)

نظرية (8): إذا سحبنا عينة عشوائية $(x_1,x_2\ldots\ldots,x_n)$ من مجتمع له توزيع طبيعي (هذا شرط أساسي) وسطه μ وتباينه σ^2، وكانت S^2 تمثل تباين العينة، فإن المتغير:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \;\ldots\ldots\ldots\ldots\ldots\; (28)$$

له توزيع **كآي تربيع** (Chi-square statistic) بدرجات حرية: $\upsilon = n-1$

المتغير السابق (χ^2) هو دالة في S^2. والرمز χ هو حرف يوناني **كآي**. أما درجات الحرية $(n-1)$ المقترنة بالإحصاء كآي تربيع تعكس حقيقة أن هناك $(n-1)$ من درجات الحرية مقترنة بتباين العينة S^2.

وما يجب الإشارة إليه هو أن **متوسط** توزيع كآي تربيع هو$(n-1)$:، أي:

$$\mathbf{E(\chi^2) = n - 1} \;\ldots\ldots\ldots\ldots\ldots\ldots\; (29)$$

وبخصوص منحنى توزيع χ^2 فإنه غير متماثل ، وهو منحنى ملتو الى اليمين (موجب الالتواء) ويقل هذا الالتواء كلما زادت قيمـــة درجـــــة الحرية. بعبـــارة أخرى كلما زادت قيمة درجة الحرية كلما اقترب منحنى χ^2 من التوزيع الطبيعي. إضافة إلى ذلك فإن قيم المتغير العشوائي χ^2 لا تكون سالبة، حيث

يبدأ منحنى التوزيع من الصفر على المحور الأفقي ويمتد إلى اليمين، وعند القيم الكبيرة يقترب من المحور الأفقي ولكن لا يلاقيه.

تحسب قيمة المتغير χ^2لأي عينة باستخدام الصيغة (28) ويكون احتمال أن تكون χ^2 **أكبر** من أي عدد يساوي **المساحة الواقعة تحت منحنى** χ^2 على يمين ذلك العدد. وعادة يُرمز لقيمةχ^2 التي تكون المساحة الواقعة على يمينها مساوية α والمناظرة لدرجات حرية υ بالرمز: $\chi^2{}_{(,\upsilon)}$

مثال(10): إذا كانت S^2 هو تباين عينة عشوائية ذات الحجم 4 وحدات مسحوبة من مجتمع له توزيع طبيعي تباينه25.

المطلوب:

1- أوجد احتمال أن تباين العينة سوف يكون 2.5 أو أقل.
2- أوجد احتمال أن تباين العينة سوف يتعدى 66.

الحل:

1- بتطبيق الصيغة (28) نجد:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{3S^2}{25}$$

لها توزيع كآي تربيع بدرجات حرية: $\upsilon = n-1 = 4-1= 3$ ،

ومنه يكون:

$$P(S^2 \le 2.5) = 1 - P(S^2 > 2.5)$$

لدينا:

$$P(S^2>2.5) = P\left(\frac{(n-1)S^2}{\sigma^2} > \frac{3(2.5)}{25}\right)$$

$$= P(\chi^2>0.3)$$

وباستخدام جدول توزيع **كآي تربيع** نجد:

$$P(\chi^2>0.3) \cong 0.95$$

أي أن:

$$= P(\chi^2>0.3) \cong 0.95 P(S^2>2.5)$$

وفي الأخير نجد:

$$P(S^2 \leq 2.5) = 1 - P(S^2 > 2.5)$$

$$= 1 - 0.95 \; = 0.05$$

2- إيجاد الاحتمال التالي:

$$P(S^2 > 66) = P\left(\frac{(n-1)S^2}{\sigma^2} > \frac{3(66)}{25}\right) = P(\chi^2 > 7.92)$$

عند درجات الحرية: $\upsilon = 3$ ، نجد: $\boldsymbol{P}(\chi^2 > 7.92) \cong 0.05$

أي أن:

$$\boldsymbol{P}(S^2 > 66) = \boldsymbol{P}(\chi^2 > 7.92) \cong 0.05$$

المحاضرة الخامسة: توزيعات المعاينة

3-4. توزيع المعاينة لنسبة $\frac{S_1^2}{\sigma_1^2}$ إلى $\frac{S_2^2}{\sigma_2^2}$: توزيع F

من أجل المقارنة بين تباين مجتمعين فإننا نحتاج النسبة بين تباين عينتين مأخوذتين من هذين المجتمعين. وسنتطرق إلى توزيع هذه النسبة في حالة المعاينة من مجتمعين طبيعيين مستقلين.

نظرية(09): إذا كانت S_1^2 ، S_2^2 هما تبايني عينتين مستقلتين حجمهما n_1 و n_2 مسحوبتين من مجتمعين لهما توزيعين طبيعيين ذو التباينين σ_1^2، σ_2^2 على الترتيب. فإن المتغير:

$$\mathbf{F} = \frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ (30)$$

يكون له توزيع **فيشر** (F) بدرجات حرية $(\upsilon_1،\upsilon_2)$ أي $(n_1-1،n_2-1)$.

وما يجب الإشارة إليه أيضا هو أن توزيع F هو دالة في درجات الحرية، حيث يكون لتوزيع F نوعين من درجات الحرية:

- درجات حرية مقترنة بتباين العينة S_1^2 في **البسط**، ويُرمز لها بـ:

$$\upsilon_1 = n_1 - 1$$

- درجات حرية مقترنة بتباين العينة S_2^2 في **المقــام**، ويُرمز لها بـ:

$$\upsilon_2 = n_2 - 1$$

أي أن توزيع F يتحدد تماما بدلالة درجات الحرية، ولا يتوقف على أي معالم أخرى. حيث يتمركز حول القيمة واحد، ويرجع ذلك إلى أن تباين المجتمعين يتم تقديرهما بتبايني العينتين، ومنه فمن المتوقع أن يكون كل من: S_1^2/σ_1^2 ، S_2^2/σ_2^2 قريبا من القيمة واحد، لذلك فإن النسبــة:

$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$ تقترب أيضا من الواحد الصحيح.

توزيع F هو توزيع ملتوي إلى اليمين ومداه نظريا يكون من الصفر إلى ما لا نهاية، أي أن قيم المتغير F لا تكون سالبة، كذلك نجد أن توزيع F يقترب من التوزيع الطبيعي كلما زادت درجتا الحرية.

إن توزيع F مثل توزيع Z و توزيع T و توزيع χ^2 ، نجده متواجدا في معظم البرامج الإحصائية الجاهزة، ولكي نستخدمه يجب أن نحدد أولا الاحتمال المرغوب فيه (محدد بالمساحة المظلة التي تقع على يسار القيمة الجزئية المطلوبة). بعد ذلك نبحث عن:

- درجات الحرية الخاصة بالبسط وهي: $(\upsilon_1 = n_1 - 1)$
- درجات الحرية الخاصة بالمقام وهي: $(\upsilon_2 = n_2 - 1)$

ومنه قيمة F التي تنتج من تقاطع درجتي حرية البسط والمقام هي القيمة الجزئية المطلوبة، حيث أن المساحة التي تقع على يسار تلك القيمة تمثل الاحتمال المطلوب.

مثال(11): إذا كان: $n_1 = 16$، $n_2 = 20$، ونرغب في إيجاد إحتمال أن يأخذ المتغير F قيمة لا تزيد عن: 1/ 0.36 ، 2/ 2.23

الحل:

لدينا:

$$\upsilon_1 = n_1 - 1 = 16 - 1 = 15\ ,\quad \upsilon_2 = n_2 - 1 = 20 - 1 = 19$$

1- باستخدام جدول توزيع F يكون:

$$P\left(F_{(15,19)} \leq 0.36\right) = 0.025$$

وبالمقابل يكون:

$$P\left(F_{(15,19)} > 0.36\right) = 0.975$$

2 - بنفس الطريقة نجد:

$$P\left(F_{(15,19)} \leq 2.23\right) = 0.95$$

وبالمقابل يكون:

$$P\left(F_{(15,19)} > 2.23\right) = 0.05$$

3-5. توزيع المعاينة لنسبة العينـة

يحتاج الباحث في أغلب الدراسات لمعرفة نسبة ظاهرة معينة في المجتمع محل الدراسة، كنسبة المدخنين في ولاية بسكرة، نسبة الذكور في جامعة محمد خيضر ببسكرة، نسبة الوحدات التالفة في إنتاج مصنع معين، نسبة الأيام التي تزيد فيها الحرارة عن 40 درجة مئوية خلال فصل الصيف في منطقة معينة،...،الخ، ففي كل حالة من هذه الحالات نجد أن المجتمع محل الدراسة منقسم إلى قسمين؛

قسم تتوافر فيه الظاهرة محل الدراسة (الخاصية المدروسة)، والقسم الثاني لا تتوافر فيه هذه الظاهرة. ومجتمعات من هذا النوع يكون فيها المتغير وصفيا أي نوعيا لا نستطيع قياسه كميا، وبالتالي تعاد صياغته وتحويله إلى متغير عشوائي نرمز له بالرمز **x**، ونتعامل في هذا النوع من المجتمعات مع نسبة الظاهرة محل الدراسة في المجتمع، ويرمز لها بالرمز **P** ، ويطلق عليها **نسبة المجتمع** ، وتحسب بالعلاقة التالية:

P = **(عدد مفردات المجتمع التي تتحقق فيها الظاهرة المدروسة) / (العدد الكلي لمفردات المجتمع)**

وبالتالي فإن P تمثل إحتمال ظهور هذه الظاهرة في المجتمع، ويُرمز لإحتمال عدم ظهور هذه الظاهرة في المجتمع بالرمز q، حيث أن حدث ظهور الظاهرة، وحدث عدم ظهورها، هما حدثان مكملان لبعضهما البعض، إذن:

$$q = \mathbf{1} - P$$

تعتبر النسبة P من أهم معالم المجتمع التي يرغب الباحث في معرفتها لكي يستطيع وصف المجتمع محل الدراسة وصفا جيدا، ولكن في الكثير من الأحيان لا نستطيع تحديد نسبة المجتمع لعدم توافر بيانات عن كل مفردات المجتمع، ولذلك نقوم بالاستدلال عليها، أي استنتاجها باستخدام نسبة الظاهرة محل الدراسة في العينة العشوائية المسحوبة من هذا المجتمع ، ويرمز **لنسبة العينة** بالرمز $\bar{p}$ وتحسب بالعلاقة التالية:

$\bar{\mathbf{P}}$ = **(عدد مفردات العينة التي تتوفر فيها الظاهرة المدروسة) / (العدد الكلي لمفردات العينة)**

نسبة العينة $(\bar{p})$، كأي إحصائية تتغير قيمتها من عينة لأخرى، وبالتالي فهي متغير عشوائي له توزيع احتمالي يطلق عليه **توزيع المعاينة لنسبة العينة**.

وسنجد أنه توجد علاقة بين الوسط الحسابي وتباين توزيع المعاينة لنسبة العينة واللذين نرمز لهما على التوالي، بالرمز $\mu_{\bar{p}}$ والرمز $\sigma^2_{\bar{p}}$ ، وبين نسبة المجتمع، فنجد أن:

$$\mu_{\bar{p}} = P \;\ldots\ldots\ldots\ldots\; (31)$$

وعندما يكون المجتمع لا نهائيا أو مجتمعا غير محدودا أو كانت عملية السحب تتم مع الإرجاع (أي $n \leq 0.05N$) فإن:

$$\sigma_{\bar{p}}^2 = \frac{\mathbf{Pq}}{\mathbf{n}} \;\ldots\ldots\ldots\ldots (32)$$

أما إذا كان المجتمع محدود أو عملية السحب تتم دون إعادة (أي $n>0.05N$) فإن:

$$\sigma_{\bar{p}}^2 = \frac{\mathbf{Pq}}{\mathbf{n}}\left(\frac{\mathbf{N-n}}{\mathbf{N-1}}\right) \;\ldots\ldots\ldots\ldots (33)$$

لقد عرفنا العلاقة بين الوسط الحسابي وتباين توزيع المعاينة لنسبة العينة، وبين نسبة المجتمع P، ولكن لاستنتاج قيمة المعلمة P باستخدام نسبة العينة p ، يجب معرفة طبيعة توزيع المعاينة للنسبة p ، وبما أن التغير الذي يحصل في قيمة p سببه تغير عدد مفردات العينة التي تتوافر فيها الظاهرة محل الدراسة من عينة إلى أخرى فقط، لأن كل العينات التي نستنتج منها توزيع المعاينة حجمها ثابت ويساوي n .

وبما أن توزيع عدد المفردات التي تتوافر فيها الظاهرة محل الدراسة في العينة (عدد المحاولات الناجحة في العينة) تتبع توزيـــع ذي الحدين (Binomial Distribution) بمعلمتين n و p في حالة سحب مفردات العينة مع الإرجاع.

ونعلم أنه وفقا **لنظرية النهاية المركزية**، نستطيع استخدام التوزيع الطبيعي كتقريب جيد لتوزيع ذات الحدين عندما يكون حجم العينة كبيرا.

وبالتالي عندما تكون n كبيرة، نستطيع استخدام التوزيع الطبيعي كتقريب جيد لتوزيع المعاينة لعدد مفردات العينة التي تتوافر فيها الظاهرة محل الدراسة، والذي هو نفسه توزيع المعاينة لنسبة العينة، وذلك كما هو واضح في النظرية التالية:

نظرية (10): وفقا لنظرية النهاية المركزية، توزيع المعاينة للنسبة p يقترب من التوزيع الطبيعي عندما يكون حجم العينة كبيرا بدرجة كافية، ويتحقق ذلك عندما يكون كل من np و nq على الأقل 5؛

أي إذا كان : $nq \geq 5$, $np \geq 5$ ، فإن المتغير العشوائي Z ، حيث:

$$\mathbf{Z} = \frac{p - \mu_{\bar{p}}}{\sqrt{\sigma_{\bar{p}}^2}} \;\ldots\ldots\ldots\ldots\; (34)$$

سيتبع توزيعا قريبا من التوزيع الطبيعي المعياري.

مثال (12): إذا علمت أن نسبة الأسر التي تقيم في شقق في ولاية ما 58.27 %، فإذا سحبنا عينة عشوائية من هذه الولاية تشمل 40 أسرة.

فما هو احتمال أن تكون نسبة الأسر التي تقيم في شقق في هذه العينة تتراوح بين 55 % و 70 % ؟

الحل:

البيانات المتوافرة لدينا هي:

نسبة المجتمع: P = 58.27 % حجم العينة: n = 40

والاحتمال المطلوب هو: $\boldsymbol{P}(0.55 \leq \bar{p} \leq 0.70) = ?$

بما أن:

$$np = 40\,(0.5827) = 23.31 \quad , \quad nq = 40\,(0.4173) = 16.69$$

أي أن كلا من np و nq أكبر من 5، وبالتالي فإن توزيع المعاينة للنسبة p سيكون قريبا من التوزيع الطبيعي بمتوسط وتباين قدرهما على التوالي كما يلي:

$$\mu_{\bar{p}} = P = 0.5827$$

$$\sigma^2_{\bar{p}} = \frac{\mathbf{Pq}}{\mathbf{n}} = [(0.5827)(1-0.5827)] / 40 = 0.0061$$

نستطيع التعبير عن الاحتمال المطلوب كما يلي:

$$\boldsymbol{P}(0.55 \leq \bar{p} \leq 0.70) = P(z_1 \leq Z \leq z_2)$$

حيث:

$$z_1 = \frac{p_1 - \bar{P}}{\sqrt{\sigma^2_{\bar{p}}}} = \frac{0.55 - 0.5827}{\sqrt{0.0061}} = -0.4$$

$$z_2 = \frac{p_2 - \bar{P}}{\sqrt{\sigma^2_{\bar{p}}}} = \frac{0.70 - 0.5827}{\sqrt{0.0061}} = 1.5$$

إذن:

$$\boldsymbol{P}(0.55 \leq \bar{p} \leq 0.70) = \boldsymbol{P}(-0.42 \leq Z \leq 1.50)$$

$$= 0.1628 + 0.4332 = 0.5960$$

3- 6. توزيع المعاينة للفرق بين نسبتي عينتين $(\bar{p}_1 - \bar{p}_2)$

إذا كانت دراستنا خاصة بمقارنة نسبة ظاهرة معينة في مجتمعين مختلفين، أي محاولة معرفة الفرق بين النسبتين $(P_1 - P_2)$، حيث P_1 ترمز لنسبة الظاهرة في المجتمع الأول، وP_2 ترمز لنسبة الظاهرة في المجتمع الثاني، وعند عدم توافر بيانات عن مفردات كل من المجتمعين، نقوم بالاستدلال على المعلمة $(P_1 - P_2)$ أي استنتاجها باستخدام الفرق بين نسبتي العينتين العشوائيتين المسحوبتين من هذين المجتمعين، أي $(p_1 - p_2)$، حيث p_1 هي نسبة الظاهرة في العينة العشوائية المسحوبة من المجتمع الأول، وp_2 هي نسبة الظاهرة في العينة العشوائية المسحوبة من المجتمع الثاني، ولذلك يجب دراسة توزيع المعاينة لهذه الإحصائية، والذي يطلق عليه **"توزيع المعاينة للفرق بين نسبتي عينتين"**.

فإذا سحبنا من المجتمع الأول كل العينات العشوائية ذات الحجم $_1$ ، وحسبنا نسبة الظاهرة المدروسة p_1 لكل عينة، وسحبنا من المجتمع الثاني كل العينات العشوائية ذات الحجم n_2 ، وحسبنا نسبة الظاهرة المدروسة p_2 لكل عينة، و كانت العينات المسحوبة من المجتمع الأول مستقلة عن
العينات المسحوبة من المجتمع الثاني، وحسبنا كل **الفروق** بين نسب عينات المجتمع الأول ونسب عينات المجتمع الثاني، أي حسبنا كل قيم $(p_1 - p_2)$، فسنحصل على توزيع المعاينة للفرق بين نسبتي عينتين $(p_1 - p_2)$ ، وإذا حسبنا الوسط الحسابي $\mu_{p_1-p_2}$ ، والتباين $\sigma^2_{p_1-p_2}$ لهذا التوزيع، فنجد أن هناك علاقات تربط هذين المقياسين مع نسبة المجتمع الأول ونسبة المجتمع الثاني، وذلك كما يلي:

$$\mu_{p_1-p_2} = \mathbf{P_1 - P_2} \ldots\ldots\ldots\ldots\ldots\ldots \quad (35)$$

فإذا كان المجتمع غير محدود أو كان السحب مع الإرجاع، و $_1$ / n_1 ، n_2 / N_2 كليهما أقل من أو يساوي 0.05 فإن:

$$\sigma^2_{p_1-p_2} = \frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2} \ldots\ldots\ldots\ldots\ldots\ldots \ (36)$$

أما إذا كان المجتمع محدود أو كان السحب دون إرجاع، و n_1/N_1 ، n_2/N_2 كليهما أو احدهما أكبر من 0.05 ، فإن:

$$\sigma^2_{p_1-p_2} = \frac{P_1 q_1}{n_1}\left(\frac{N_1 - n_1}{N_1 - 1}\right) + \frac{P_2 q_2}{n_2}\left(\frac{N_2 - n_2}{N_2 - 1}\right) \ldots\ldots\ldots\ (37)$$

ومن هنا نصل إلى النظرية التالية:

نظرية (11): إذا كان لدينا عينتان مستقلتان كبيرتا الحجم تم سحبهما من مجتمعين، ووفقا لنظرية النهاية المركزية، يكون توزيع المعاينة للفرق بين نسبتي العينتين $(p_1 - p_2)$ توزيعا قريبا من التوزيع الطبيعي بوسط حسابي وتباين تم توضيحهما في العلاقتين (35) و (36) على الترتيب. ومن ثم فإن المتغير العشوائي Z حيث:

$$Z = \frac{(\bar{p}_1 - \bar{p}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}} \ldots\ldots\ldots\ldots\ldots\ldots(38)$$

سيتبع توزيعا قريبا من التوزيع الطبيعي المعياري.

مثال (13): عن الكتاب الإحصائي الصادر عن الديوان الوطني للإحصاء، نجد أن العدد الكلي للسكان الذين أعمارهم تتراوح بين 10 سنوات و30سنة، موزعين حسب الجنس كما يلي: 2157136 ذكرا منهم 221914 يحملون شهادة جامعية، و 2067508 أنثى منهن 144423 يحملن شهادة جامعية، فإذا سحبنا من هذين المجتمعين عينتين عشوائيتين مستقلتين، الأولى من الذكور حجمها 2000 ذكر، والثانية من الإناث حجمها 1500 أنثى.

المطلوب:

أوجد احتمال أن يكون الفرق بين نسبتي العينتين أكبر أو يساوي 5 %.

الحل:

بافتراض أن المجتمع الأول يمثل مجتمع الذكور، والمجتمع الثاني يمثل مجتمع الإناث، نجد أن:

P_1: تمثل نسبة حاملي شهادة جامعية في المجتمع الأول $= \frac{221914}{2157136} = 0.10$

P_2: تمثل نسبة حاملي شهادة جامعية في المجتمع الثاني $= \frac{144423}{2067508} = 0.07$

n_1 : حجم العينة الأولى = 2000 ذكر.

n_2 : حجم العينة الثانية = 1500 أنثى.

والاحتمال المطلوب هو : $= ?$

$$\boldsymbol{P}\,[(\bar{p}_1 - \bar{p}_2) \geq 0.05]$$

بما أن n_1 و n_2 كبيرتان، فإن توزيع المعاينة للإحصائيّة $(\bar{p}_1 - \bar{p}_2)$ سيكون قريبا من التوزيع الطبيعي، وبالتالي فإن الإحتمال المطلوب يتم حسابه كما يلي:

$$\boldsymbol{P}\,[(\bar{p}_1 - \bar{p}_2) \geq 0.05\,] = P(Z \geq z)$$

حيث أن:

$$z = \frac{(\bar{p}_1 - \bar{p}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}} = \frac{(0.05) - (0.10 - 0.07)}{\sqrt{\frac{0.10x0.90}{2000} + \frac{0.07x0.93}{1500}}}$$

$$= 2.13$$

إذن:

$$P\left[(\bar{p}_1 - \bar{p}_2) \geq 0.05\right] = P(Z \geq 2.13) = 0.5 - 0.4834 = 0.0166$$

المحاضرة السادسة / الفصل الثاني
" إختبار الفرضيات"

1- مقدمة:

بالرغم من أهمية موضوع تقدير المعالم الذي تناولنا دراسته في المحور السابق إلا أنه غالب ما يكون الاهتمام منصبا ليس على مجرد تقدير المعالم ولكن على عملية وضع القواعد التي تسمح بالتوصل إلى قرار بقبول أو رفض فرض عن معالم مجتمع أو أكثر وهذا ما يسمى **باختبار الفرضيات**.

فمثلا يمكن للمسؤول عن الرقابة الإحصائية على جودة المنتجات أن يقرر ما إذا كانت الوحدات المنتجة من صنف معين تحقق الموصفات المطلوبة أم لا، فإذا كانت الموصفات المطلوبة تحدد أن وزن الوحدة المنتجة من صنف معين تساوي 1كلغ ، وبالتالي فان المسؤول عن الرقابة الإحصائية في هذه الحالة يرغب في اختبار الفرض التالي : $\mu = 1$ ، حيث μ يرمز لمتوسط مجتمع الوحدات المنتجة من هذا الصنف.

وعلى العموم فإن الفرض الإحصائي هو ادعاء أو اعتقاد يتعلق بقيمة غير معلومة للمؤشر (أي للمعلمة) ويُراد إختبار مدى صحتها، كذلك فإن الفرض الإحصائي هو كل عبارة صحتها أو عدم صحتها بحاجة إلى قرار.

2-خطوات إختبارات الفروض:

لإجراء إختبار الفرضيات هناك خمس خطوات يجب القيام بها، ويمكن توضيحها في الشكل التالي:

الشكل (1) خطوات إختبار الفرضيات

وضع الفرض العدمي و الفرض البديل

إختيار مستوى المعنوية

حساب الإختبار الإحصائي الملائم
(حساب دالة الاختبار)

ومن أجل فهم واستيعاب الخطوات السابقة، يجب دراسة وتحليل المفاهيم التالية:

- ❖ الفرض العدمي والفرض البديل.
- ❖ الخطأ من النوع الأول والخطأ من النوع الثاني.
- ❖ مستوى المعنوية.
- ❖ إحصاء الاختبار.

وفيما يلي سوف نتناول تحليلا موجزا لهذه المفاهيم.

2-1. الفرض العدمي والفرض البديل

"Null and Alternative Hypothesis"

الفرض العدمي يُرمز له بالرمز H_0، وهو صفة مميزة لا تحتاج إلى إثبات، فنحن نفترض أنـه صحيحا ما لم يظهر بوضوح أنه غير صحيح، فمثلا عادة ما تكون عملية تعبئة المنظف في الأكيـاس بمتوسـط 500غ . في أي وقت من عملية التعبئة، نفترض أن هذا هو المتوسط ما لم يكن دليل العينة يشـير بشكل واضح إلى أن المتوسط قد انحرف عن ما هو محدد، وبالتالي يوضع الفرض العدمي على النحو التالي:

$$H_0 : \mu = 500$$

وعلى العموم ، الفرض العدمي يتصف بحقيقة أننا نعامله و كأنه صحيح ما لم يكن ادعائه يتناقض بوضوح مع بيانات العينة، لذا إذا كان الفرض العدمي غير مناقض بوضوح، فإنه لا يوجد سببا كافيا لرفضه. وغالبا ما ننظر إلى الفرض العدمي على أنه نقطة البداية في عملية التحليل.

أما **الفرض البديل** فيرمز له بالرمز H_1، و هو بديل لحالة الفرض العدمي، أي هو الفرض الذي يمكن قبوله عند رفض الفرض العدمي، ففـــــــي مثـــال عملية التعبئة، يكون

الفرض البديل هو أن متوسط العملية الحالي ليس 500غ. و بالتالي يوضع الفرض البديل على النحو التالي:

$$H_1: \mu \neq 500$$

والفرض البديل يأخذ ثلاثة أشكال وهي:

- الفرض البديل **ذو الذيلين (الطرفيين)**: وفيه تكون معلمة المجتمع θ لا تساوي قيمة معينة θ_0 .

ففي مثال عملية التعبئة دائما يكون $H_1 : \mu \neq 500$ هذا من جهة. ومن جهة أخرى نضع نصف قيمةα في كل من طرفي توزيع دالة الاختبار، أي تكون $\frac{\alpha}{2}$ على كل طرف. فنقبل الفرض H_0 إذا وقعت قيمة دالة الاختبار في منطقة قبول H_0 ، و نرفض H_0 إذا وقعت قيمة دالة الاختبار في منطقة رفض H_0.

- الفرض البديل ذو **الطرف الأعلى (الطرف الأيمن)**: وفيه تكون معلمة المجتمع θ أكبر من قيمة معينة θ_0، ففي المثال السابق يكون الفرض البديل: $H_1 : \mu > 500$ هذا من جهة. ومن جهة أخرى نضع قيمة α في الطرف الأيمن (الأعلى) من توزيع دالة (اقتران) الاختبار. فنقبل H_0 إذا وقعت قيمة دالة الاختبار في منطقة قبول H_0، ونرفض H_0 إذا وقعت قيمة دالة الاختبار في منطقة رفض H_0 .

- الفرض البديل **ذو الطرف الأدنى (الطرف الأيسر)**: وفيه تكون معلمة المجتمع θ أصغر من قيمة معينة θ_0 ، ففي المثال السابق يكون الفرض البديل: $H_1 : \mu < 500$ ، ونضع قيمة αفي الطرف الأيسر من توزيع دالة الاختبار.

بنفس الطريقة السابقة، نقبل H_0 إذا وقعت قيمة دالة الاختبار في منطقة قبول H_0 ، ونرفض H_0 إذا وقعت قيمة دالة الاختبار في منطقة رفض H_0.

2-2 . الخطأ من النوع الأول و الخطأ من النوع الثاني:

من المعلوم أنه عند اتخاذ أي قرار إحصائي فإن ذلك ينطوي على أخطاء بنسب معينة ، حيث أنـه من المحتمل أن نرفض فرضية معينة في حين أنها صحيحة ، والعكس صحيح. لهذا فان هناك نوعان مـن الأخطاء الإحصائية وهي:

-الخطأ من النوع الأول (α): هو الخطأ الذي نقع فيه عندما نرفض الفرضية الصفرية H_0 بالرغم من صحتها، ويُرمز لإحتمال وقوع هذا الخطأ بالرمز α، ونسميه "**مستوى دلالة الاختبار**" أو" **مستوى معنوية الاختبار**".

-الخطأ من النوع الثاني (β): هو الخطأ الذي نقع فيه عندما نقبل الفرضية الصفرية بالرغم من عدم صحتها، ويُرمز إلى إحتمال هذا الخطأ بالرمز β.

وعلى العموم يمكن توضيح نوعي الخطأ من خلال الجدول التالي:

الجدول (01) جدول توضيحي لنوعي الخطأ

	القرار	
الفرض العدمي	قبول الفرض العدمي	رفض الفرض العدمي
الفرض العدمي H_0 صحيح	قرار صحيح	خطأ من النوع الأول (α)
الفرض العدمي H_0 خاطئ	خطأ من النوع الثاني (β)	قرار صحيح

بصفة عامة يمكن توضيح الخصائص التالية للعلاقـة بين الخطأين:

- يرتبط الخطأ من النوع الأول (α) ارتباطا عكسيا مع الخطأ من النوع الثاني (β)، أي أن إنخفاض إحتمال أحدهما يؤدي إلى زيادة إحتمال الآخر.

- زيادة حجم العينة يؤدي إلى تناقص كلا النوعين من الخطأ.

- إذا كان الفرض العدمي H_0 غير صحيح، فإن قيمة الخطأ من النوع الثاني (β) تكون أكبر ما يمكن عندما تكون القيمة الحقيقية للمعلمة تكاد تتطابق مع القيمة الافتراضية لها، والعكس صحيح عندما يكون الفرق بين القيمتين الحقيقية والافتراضية، فإن قيمة الخطأ من النوع الثاني (β) تقل.

3-2 . مستوى المعنوية أو مستوى الدلالة:

في اختبار فرضية معينة، فإن أقصى إحتمال والذي يمكن أن نتحمل بـه خطأ من النـوع الأول يسمى **مستوى المعنوية للاختبـار**، هذا الإحتمال يرمز له بالرمز α ويحدد بشكل عام قبل سحـب أي عينة، لكـي لا تتأثر النتائج التي حصلنا عليها في إختبارنا.

ومن الناحية العملية فإننا نستخدم عادة مستوى المعنوية 0.05 أو 0.01 ، و إذا كان هناك قيم أخرى يتم استخدامها.

إن استخدامنا لمستوى المعنوية 0.05 أو 5% في اختبار فرضية معينة، فهذا يعني أن هناك حوالي 5 فرص من أصل 100 فرصة أننا سوف نرفض الفروض وهي صحيحة؛ بمعنى أننا سنكون واثقين بنسبة 95% في أننا سنتخذ القرار الصحيح، و بالمقابل فإنه من الممكن أن نكون على خطأ باحتمال قدره 0,05.

4-2. إحصاء الاختبار Test Statistic

إحصاء الاختبار هو إقتران إحصائي يساعدنا على إتخاذ قرار حول فرضية إحصائية معينة، ويتم حساب قيمته من بيانات العينة، وبالتالي فهو عبارة عن متغير عشوائي تتغير قيمته بتغير بيانات العينـة الإحصائية التي نأخذها من المجتمع الإحصائي.

ومن القيمة التي نحصل عليها لإحصاء الاختبار نقرر ما إذا كان هناك سببا قويا لرفض H_0 وقبول H_1 أم لا.

المحاضرة السابعة/ الفصل الثاني
" إختبار الفرضيات"
يتبع ....

3- إختبار الفروض حول وسط المجتمع μ

3-1. إختبار الفروض حول وسط المجتمع μ، عندما يكون هذا المجتمع طبيعي ذو تباين معلوم:

نظرية (1): إذا كانت ($x_1,x_2......,x_n$) عينة عشوائية مسحوبة من مجتمع يتوزع توزيعا طبيعيا وسطه μ وتباينه σ^2 **معلوم**، فإن إحصاء الاختبار المناسب هو: $Z=\frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$، وأردنا إختبار الفرضية الصفرية التالية:

$$H_0:\ \mu=\mu_0$$

مقابل الفرضية البديلة:

1. $H_1:\ \mu\neq\mu_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$Z<-z_{\alpha/2} \quad \text{أو} \quad Z>z_{\alpha/2}$$

2. $H_1:\ \mu>\mu_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$Z>z_{\alpha}$$

3. $H_1:\ \mu<\mu_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$Z<-z_{\alpha}$$

مثال: تخضع أوزان عبوات أحد مساحيق الغسيل لتوزيع طبيعي إنحرافه المعياري 7غ ومعدله μ، عند مستوى الدلالة 0.05 .

المطلوب:

اختبر الفرضية الصفرية: $H_0:\ \mu=200$ ، مقابل الفرضية البديلة: $H_1:\ \mu\neq200$، إذا كان الوسط الحسابي لعينة حجمها 25 هو 208.

الحل:

$$H_0: \quad \mu=200$$
$$H_1: \mu\neq 200$$
$$\alpha = 0.05$$

بما أن الفرضية البديلة لا تحدد إتجاها واحدا، فإن الاختبار المناسب هو ذو طرفين، والقيم الحرجة تكون:

$z_{\alpha/2} = z_{0.025} = 1.96$ ، $-z_{\alpha/2} = -z_{0.025} = -1.96$

نرفض H_0 عند مستوى الدلالة 0.05 إذا كان:

$Z > 1.96$ أو $Z < -1.96$

والآن نقوم بحساب Z من خلال المعادلة التالية : $Z = \dfrac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

فنجده يساوي: $Z = \dfrac{208-200}{7/\sqrt{25}} = 5.71$

نلاحظ أن $5.71 > 1.96$ ، أي أن قيمة Z تقع في منطقة الرفض (المنطقة الحرجة)، لذلك نرفض H_0 ونقبل H_1 عند مستوى المعنوية 0.05، أي أن قيمة $\mu > 200$ ، لأن Z وقعت في منطقة الرفض اليُمنى، أي على يمين $z_{\alpha/2}$.

3-2. اختبار الفروض حول وسط المجتمع، عندما يكون هذا المجتمع طبيعي ذو تباين مجهول، وحجم العينة كبير

نظرية (2): إذا تم أخذ عينة عشوائية كبيرة الحجم من مجتمع يتوزع توزيعا طبيعيا وسطه μ وتباينه σ^2 **مجهول**، فإن إحصاء الاختبار المناسب هو:

$$Z = \frac{\overline{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

وأردنا اختبار الفرضية الصفرية: $H_0: \quad \mu = \mu_0$ مقابل الفرضية البديلة:

1. $H_1: \quad \mu\neq\mu_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$Z > z_{\alpha/2}$ أو $Z < -z_{\alpha/2}$

2. $H_1: \quad \mu>\mu_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$Z > z_{\alpha}$$

3. $H_1: \quad \mu<\mu_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$Z < -\ z_{\alpha}$$

3- 3. اختبار الفروض حول وسط المجتمع، عندما يكون هذا المجتمع طبيعي ذو تباين مجهول، وحجم العينة صغير

نظرية (3): إذا تم أخذ عينة عشوائية صغيرة الحجم من مجتمع يتوزع توزيعا طبيعيا وسطه μ وتباينه σ^2 **مجهول**، فإن إحصاء الاختبـار المناسب هو:

$$T = \frac{\overline{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

وهو توزيـع t بدرجات حريـة $(\upsilon = n-1)$.

وأردنا اختبار الفرضية الصفرية: $H_0: \mu = \mu_0$ ، مقابل الفرضية البديلة:

1. $H_1: \mu \neq \mu_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$T < -t_{(\frac{\alpha}{2},\upsilon)} \quad \text{أو} \quad T > t_{(\frac{\alpha}{2},\upsilon)}$$

2. $H_1: \quad \mu>\mu_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$T > \ t_{(\alpha,\upsilon)}$$

3. $H_1: \quad \mu<\mu_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$T < -t_{(\alpha,\upsilon)}$$

مع الإشارة إلى أن القيمة الحرجة $t_{(\frac{\alpha}{2})}$ هي قيمة على المحور الأفقي لتوزيع t ذي درجـات حرية $(n-1)$ ويقع إلى يمينها مساحة قدرها $\frac{\alpha}{2}$. أما القيمة الحرجة $t_{(\alpha)}$ فهي قيمة على المحور الأفقي لتوزيع t ذي درجات حريـة $(n-1)$ ويقع إلى يمينها مساحة قدرها α .

مثال: استنتج أحد الباحثين أن معدل عدد الساعات التي يقضيـها طلبة إحدى الجامعات في الدراسة أثناء أسبوع الامتحانات هو 50 ساعة، والمطلوب هو اختبار هذه الفرضية مقابل

فرضية أن معدل عدد الساعات يختلف عن 50 ساعة، إذا كان الوسط الحسابي لعدد الساعات التي قضاها 10 طلاب أثناء ذلك الأسبوع هو 7.51 ساعة بإنحراف معياري 6.3 ساعة. مع العلم أن مستوى الدلالة هو 0.05، وتوزيع عدد الساعات الدراسية يقترب من التوزيع الطبيعي.

الحل:

$$H_0: \quad \mu=50$$
$$H_1: \quad \mu \neq 50$$
$$\alpha = 0.05$$

بما أن الفرضية البديلة لا تحدد اتجاها واحدا، فان الاختبار المناسب هو ذو طرفين، وبما أن توزيع المجتمع طبيعي، وتباينه غر معلوم، وحجم العينة صغير، نستعمل توزيع t بدرجات حرية $(n-1)$ ، ولذلك فالقيم الحرجة هي:

$$t_{(0.025,9)} = 2.262$$

$$-t_{(0.025,9)} = -2.262$$

نرفض الفرضة الصفرية عند مستوى الدلالة 0.05 إذا كان:

$$T < -2.262 \quad \text{أو} \quad T > 2.262$$

والآن نقوم بحساب قيمة T من خلال المعادلة التالية:

$$T = \frac{\overline{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

أي :

$$T = \frac{51.7-50}{6.3/\sqrt{10}} = 0.85$$

نلاحظ أن: $(-2.262 < 0.85 < 2.262)$؛ أي أن قيمة T تقع في منطقة القبول، لذلك نقبل H_0 ونرفض H_1 عند مستوى المعنوية 0.05، أي أن إستنتاج الباحث صحيح ($\mu = 50$).

4- اختبار الفروض حول الفرق بين متوسطي مجتمعين

4-1. عندما يكون تبايني المجتمعين معلومين

نظرية (4): إذا كان $\overline{X}_1$ هو الوسط الحسابي لعينة عشوائية حجمها n_1 تم سحبها من مجتمع له توزيع طبيعي متوسطه μ_1 وتباينه **معلوم** σ_1^2، وكان $\overline{X}_2$ هو الوسط الحسابي لعينة

عشوائية حجمها n_2 مستقلة عن الأولى تم سحبها من مجتمع له توزيع طبيعي متوسطه μ_2 وتباينه σ_2^2 **معلوم** أيضا.

وأردنا اختبار الفرضية الصفرية: $H_0: (\mu_1-\mu_2) = d_0$ مقابل الفرضية البديلة:

1. $H_1:(\mu_1-\mu_2) \neq d_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$Z < -z_{\alpha/2} \quad \text{أو} \quad Z > z_{\alpha/2}$$

2. $H_1:(\mu_1-\mu_2) > d_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$Z > z_{\alpha}$$

3. $H_1:(\mu_1-\mu_2) < d_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$Z < -z_{\alpha}$$

مع العلم أن قيمة Z يتم حسابها وفقا للصيغة التالية:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

في معظم الحالات تكون الفرضية الصفرية في هذا النوع من الاختبار على الشكل: H_0 : $(\mu_1-\mu_2) = 0$ أي أن وسطي المجتمعيين متساويين.

مثال: أخذت عينة عشوائية حجمها 16 من مجتمع طبيعي تباينه 9 فوُجد أن وسطها الحسابي 69، ثم أخذت عينة عشوائية أخرى حجمها 20 من مجتمع طبيعي آخر مستقل عن الأول تباينه 25، فوُجد أن وسطها الحسابي 71.

المطلوب: إختبر الفرضية $H_0 : \mu_1 = \mu_2$ مقابل الفرضية $H_1 : \mu_1 \neq \mu_2$ عند مستوى المعنوية 0.05.

الحل:

$$H_0 : \mu_1 = \mu_2 \Rightarrow H_0 : (\mu_1-\mu_2) = 0$$
$$H_1 : \mu_1 \neq \mu_2 \Rightarrow H_1 : (\mu_1-\mu_2) \neq 0$$
$$\alpha = 0.05$$

بما أن الفرضية البديلة لا تُحدد اتجاها واحدا، فإن الاختبار المناسب هو ذو طرفين، والقيم الحرجة تكون:

$$-z_{\alpha/2} = -z_{0.025} = -1.96 \quad ، \quad z_{\alpha/2} = z_{0.025} = 1.96$$

نرفض H_0 عند مستوى الدلالة 0.05 إذا كان:

$$Z < -1.96 \quad \text{أو} \quad Z > 1.96$$

والآن نقوم بحساب Z من خلال المعادلة التالية:

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

أي :

$$Z = \frac{(69 - 71) - 0}{\sqrt{\frac{9}{16} + \frac{25}{20}}} = -1.486$$

نلاحظ أن: $-1.486 > -1.96$ ؛ أي أن قيمة Z تقع في منطقة قبول الفرضية الصفرية، لذلك نقبل H_0 ونرفض H_1 عند مستوى المعنوية 0.05، أي أن وسطي المجتمعين متساويين عند نفس مستوى المعنوية.

المحاضرة الثامنة / الفصل الثاني
" إختبار الفرضيات"
يتبع ....

4-2. عندما يكون تبايني المجتمعين مجهولين وحجمي العينتين كبير بدرجة كافية:

نظرية (5): إذا كان $\overline{X}_1$ هو الوسط الحسابي لعينة عشوائية حجمها n_1 **كبير** تم سحبها من مجتمع متوسطه μ_1 وتباينه σ_1^2 **مجهول**، وكان $\overline{X}_2$ هو الوسط الحسابي لعينـة عشوائية أخرى حجمها n_2 **كبير** أيضا مستقلة عن الأولى تم سحبها من مجتمع متوسطه μ_2 وتباينه σ_2^2 **مجهول** أيضا. فان إحصاء الإختبار المناسب هو Z .

وأردنا اختبار الفرضية الصفرية: $H_0 : (\mu_1-\mu_2) = d_0$ مقابل الفرضية البديلة:

1. $H_1:(\mu_1-\mu_2) \neq d_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$Z > z_{\alpha/2} \quad \text{أو} \quad Z < -z_{\alpha/2}$$

2. $H_1:(\mu_1-\mu_2) > d_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$Z > z_{\alpha}$$

3. $H_1:(\mu_1-\mu_2) < d_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$Z < -z_{\alpha}$$

مع العلم أن قيمة Z يتم حسابها وفقا للصيغة التالية:

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

مثال: أخذت عينة عشوائية من مجتمع معين، وكان حجمها 50، ووسطها الحسابي 57.5 وانحرافها المعياري 6.2، ثم أخذت عينة عشوائية أخرى مستقلة عن الأولى من مجتمع آخر، وكان حجمها 60، ووسطها الحسابي 54.4 وانحرافها المعياري 10.6 .

المطلوب:

هل نستطيع أن نستنتج أن متوسط المجتمع الأول أكبر من متوسط المجتمع الثاني عند مستوى الدلالة 0.05.

الحل: نريد اختبار ما يلي:

$$H_0 : (\mu_1 - \mu_2) = 0$$
$$H_1 : (\mu_1 - \mu_2) > 0$$
$$\alpha = 0.05$$

نلاحظ أن الفرضية البديلة ذات طرف واحد، إذن القيمة الحرجة هي:

$$z_{0.05} = 1.645$$

نرفض H_0 عند مستوى الدلالة 0.05 إذا كان:

$$Z > 1.645$$

والآن نقوم بحساب Z من خلال المعادلة التالية:

$$Z = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

فنجد أن:

$$Z = \frac{(57.5 - 54.4) - 0}{\sqrt{\frac{(6.2)^2}{50} + \frac{(10.6)^2}{60}}} = 1.908$$

نلاحظ أن $1.645 < 1.908$ ، أي أن قيمة Z تقع في منطقة رفض H_0 (المنطقة الحرجة)، لذلك نرفض H_0 ونقبل H_1 ؛ وعليه نستنتج أن هناك دليلا كافيا على أن متوسط المجتمع الأول أكبر من متوسط المجتمع الثاني عند مستوى المعنوية 0.05 .

3-4. عندما يكون تبايني المجتمعين مجهولين والعينتين مستقلتين وصغيرتا الحجم

وفي هذا العنصر لدينا حالتين:

- حالة تبايني المجتمعين مجهولين ومتساويين.
- حالة تبايني المجتمعين مجهولين وغير متساويين.

ويمكن توضيح كل هذا في النظريتين **السادسة** و**السابعة** التاليتين.

نظرية (6): إذا كان $\bar{X}_1$ هو الوسط الحسابي لعينة عشوائية حجمها n_1 **صغير** سحبت من **مجتمع طبيعي** وسطه μ_1 وتباينه σ_1^2، وكان $\bar{X}_2$ هو الوسط الحسابي لعينة عشوائية أخرى مستقلة عن الأولى حجمها n_2 **صغير** سحبت من **مجتمع طبيعي** آخر مستقل عن المجتمع

الأول وسطه μ_2 وتباينه σ_2^2، وكان **تبايني المجتمعين مجهولين ومتساويين**، فإن إحصاء الإختبار المناسب هو:

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{s_p^2(\frac{1}{n_1}+\frac{1}{n_2})}}$$

وتوزيعه الاحتمالي هو توزيع t بدرجة حرية $\upsilon = n_1 + n_2 - 2$

وأردنا إختبار الفرضية الصفرية: $H_0 : (\mu_1 - \mu_2) = d_0$ مقابل الفرضية البديلة:

1. $H_1 : (\mu_1 - \mu_2) \neq d_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$T < -t_{(\frac{\alpha}{2},\upsilon)} \quad \text{أو} \quad T > t_{(\frac{\alpha}{2},\upsilon)}$$

2. $H_1 : (\mu_1 - \mu_2) > d_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$T > t_{(\alpha,\upsilon)}$$

3. $H_1 : (\mu_1 - \mu_2) < d_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$T < -t_{(\alpha,\upsilon)}$$

مثال: قدمت مؤسستين من المؤسسات المنتجة للآلات عرضين، يتضمن كل عرض طريقة معينة للإنتاج، وانعكس ذلك في الزمن اللازم لإنتاج الوحدة من المنتج، ولقياس متوسط هذا الزمن للآلتين تم قياس الزمن لعدد 5 وحدات من الآلة الأولى، و6 وحدات من الآلة الثانية وكانت كما يلي:

الطريقة الأولى (الزمن بالدقيقة)	الطريقة الثانية (الزمن بالدقيقة)
2	3
4	7
9	5
3	8
2	4
/	3

المطلوب: إختبار الفرض القائل بعدم وجود فروق معنوية بين متوسطي الزمن للآلتين عند مستوى معنوية 0.01، مع العلم أن تبايني المجتمعين مجهولين ومتساويين.

الحل:

قبل اختبار الفرضية السابقة يجب أولا حساب ما يلي:

$$s_p^2 \text{ ، } s_2^2\text{ ، } s_1^2\text{ ، } \bar{x}_2 \text{ ، } \bar{x}_1$$

وبعد القيام بجميع العمليات الحسابية نجد أن:

$$\bar{x}_1 = 4 \quad , \bar{x}_2 = 5 \quad , \quad s_1^2 = 8.5 \quad , \quad s_2^2 = 4.4 \quad , \quad s_p^2 = 6.22$$

وعليه فإن اختبار الفرضية السابقة يتم وفقا للخطوات التالية:

$$H_0 : \mu_1 = \mu_2 \Rightarrow H_0 : (\mu_1 - \mu_2) = 0$$

$$H_1 : \mu_1 \neq \mu_2 \Rightarrow H_1 : (\mu_1 - \mu_2) \neq 0$$

$$\alpha = 0.01$$

بما أن الفرضية البديلة لا تحدد إتجاها واحدا، فإن الاختبار المناسب هو ذو طرفين، والقيم الحرجة تكون:

$$-t_{(\frac{\alpha}{2},\upsilon)} = -t_{(0.005,9)} = -3.250 \text{ ، } t_{(\frac{\alpha}{2},\upsilon)} = t_{(0.005,9)} = 3.250$$

نرفض H_0 عند مستوى الدلالة 0.01 إذا كان:

$$T < -3.250 \quad \text{أو} \quad T > 3.250$$

والآن نقوم بحساب T من خلال المعادلة التالية:

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

أي:

$$T = \frac{(4-5)-0}{\sqrt{6.22(\frac{1}{5}+\frac{1}{6})}} = -0.662$$

نلاحظ أن: $-0.662 > -3.250$؛ أي أن قيمة T تقع في منطقة قبول الفرضية الصفرية، لذلك نقبل H_0 ونرفض H_1 ، ومنه لا توجد فروق معنوية بين متوسطي الزمن للآلتين عند مستوى معنوية 0,01.

نظرية (7): إذا كان $\bar{x}_1$ هو الوسط الحسابي لعينة عشوائية حجمها n_1 **صغير** سحبت من **مجتمع طبيعي** وسطه μ_1 وتباينه σ_1^2، وكان $\bar{x}_2$ هو الوسط الحسابي لعينة عشوائية أخرى

مستقلة عن الأولى حجمها n_2 **صغير** سحبت من **مجتمع طبيعي** آخر مستقل عن المجتمع الأول وسطه μ_2 وتباينه σ_2^2، وكان **تبايني المجتمعين مجهولين وغير متساويين،** فإن إحصاء الاختبار المناسب هو:

$$T=\frac{\left(\bar{X}_1-\bar{X}_2\right)-d_0}{\sqrt{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)}}$$

وتوزيعه الاحتمالي هو توزيع t بدرجة حرية لها الصيغة المركبة التالية:

$$\upsilon=\left(\frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{(n_1-1)}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{(n_2-1)}}\right)$$

وأردنا اختبار الفرضية الصفرية: $H_0: (\mu_1-\mu_2)=d_0$ مقابل الفرضية البديلة:

1. $H_1:(\mu_1-\mu_2)\neq d_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$T<-t_{(\frac{\alpha}{2},\upsilon)} \quad \text{أو} \quad T>t_{(\frac{\alpha}{2},\upsilon)}$$

2. $H_1:(\mu_1-\mu_2)> d_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$T> t_{(\alpha,\upsilon)}$$

3. $H_1:(\mu_1-\mu_2)< d_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$T<-t_{(\alpha,\upsilon)}$$

مثال (07): لنفرض أننا سحبنا عينتين عشوائيتين مستقلتين من مجتمعين لكل منهما توزيع طبيعي فوجدنا النتائج الآتية:

$$\bar{X}_1=32 \quad , \quad \bar{X}_2=30.2$$

$$n_1=7 \quad , \quad n_2=6 \quad , \quad s_1^2=4.470 \quad , \quad s_2^2=0.652$$

المطلوب: إختبر الفرضية $H_0: (\mu_1-\mu_2)=0$ مقابل $H_1: (\mu_1-\mu_2)\neq 0$ عند مستوى المعنوية 0.05، وكان تبايني المجتمعين مجهولين وغير متساويين.

الحل:

$$H_0 : (\mu_1 - \mu_2) = 0$$

$$H_1 : (\mu_1 - \mu_2) \neq 0$$

$$\alpha = 0.05$$

نلاحظ أن الفرضية البديلة لا تحدد اتجاها واحدا، ومنه فإن الإختبار المناسب هو ذو طرفين، ومن أجل إيجاد القيم الحرجة، يجب أولا تحديد قيمة درجة الحرية، حيث:

$$\upsilon = \left(\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{(n_1 - 1)} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{(n_2 - 1)}} \right) = \left(\frac{\left(\frac{4.470}{7} + \frac{0.652}{6}\right)^2}{\frac{\left(\frac{4.470}{7}\right)^2}{(6)} + \frac{\left(\frac{0.652}{6}\right)^2}{(5)}} \right) = 8$$

ومنه تكون القيم الحرجة كما يلي:

$$-t_{(\frac{\alpha}{2},\upsilon)} = -t_{(0.025,8)} = -2.306 \quad ، \quad t_{(\frac{\alpha}{2},\upsilon)} = t_{(0.025,8)} = 2.306$$

نرفض H_0 عند مستوى الدلالة 0.05 إذا كان:

$$T < -2.306 \quad \text{أو} \quad T > 2.306$$

والآن نقوم بحساب T من خلال المعادلة التالية:

$$T = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - d_0}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

أي :

$$T = \frac{(32 - 30.2) - 0}{\sqrt{\left(\frac{4.470}{7} + \frac{0.652}{6}\right)}} = 2.08$$

نلاحظ أن: $2.08 < 2.306$ ؛ أي أن قيمة T تقع في منطقة قبول الفرضية الصفرية، لذلك نقبل الفرضية الصفرية H_0 ونرفض الفرضية البديلة H_1، ومن هنا نستنتج أنه لا توجد فروق معنوية بين متوسطي المجتمعين عند مستوى المعنوية 0,05.

المحاضرة التاسعة / الفصل الثاني
" إختبار الفرضيات"
يتبع ....

5- اختبار الفرضيات حول تباين المجتمع

في الكثير من الأحيان يكون من الأهمية بما كان معرفة تباين مجتمع من المجتمعات. ففي عملية التصنيع يلزمنا معرفة التباين في أوزان المنتجات أو في أحجامها، لأن هذا التباين يعكس مستوى الدقة ومستوى الجودة في عملية التصنيع. ولهذا كثيرا ما يقوم مهندسوا الإنتاج بقياس التباين في المنتجات من حين لآخر ويحاولون أن يبقوا ذلك التباين ضمن حدود معينة.

نظرية (9): إذا كان لدينا مجتمع يتوزع توزيعا طبيعيا تباينه σ^2، وأردنا إجراء إختبارات حول المعلمة σ^2، فإن إحصاء الإختبار المناسب في هذه الحالة هو الإحصاء الذي يعتمد على أفضل مقدر للمعلمة σ^2، وهو تباين العينة S^2، وهذا الإحصاء هو المتغير العشوائي χ^2 ، حيث:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

والتوزيع الاحتمالي لهذا المتغير يسمى **توزيع كاي تربيع** بدرجة حرية:

$$\upsilon = n-1$$

وبالتالي أردنا اختبار الفرضية الصفرية $H_0:\ \sigma^2 = \sigma_0^2$ مقابل الفرضية البديلة:

1. $H_1:\ \sigma^2 \neq \sigma_0^2$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$\chi^2 > \chi^2_{(\frac{\alpha}{2},\upsilon)} \quad \text{أو} \quad \chi^2 < \chi^2_{(1-\frac{\alpha}{2},\upsilon)}$$

2. $H_1:\ \sigma^2 > \sigma_0^2$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$\chi^2 > \chi^2_{(\alpha,\upsilon)}$$

3. $H_1:\ \sigma^2 < \sigma_0^2$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$\chi^2 < \chi^2_{(1-\alpha,\upsilon)}$$

مثال: يدعى مدير أحد المصانع أن منتجاته لا يزيد الإنحراف المعياري لأطوالها عن 0.2سم. أراد أحد الزبائن أن يتأكد من هذا الإدعاء، فأخذ عينة عشوائية مكونة من 10 منتجات فوجد أن الانحراف المعياري فيها كان 0.4سم.

المطلوب: هل تعطينا هذه النتيجة مبررا قويا لرفض إدعاء البائع مستخدما في ذلك مستوى الدلالة 0,05.

الحل:

$$H_0:\sigma^2 = 0.04$$

$$H_1:\sigma^2 > 0.04$$

$$\alpha = 0.05$$

نلاحظ أن الفرضية البديلة ذات طرف واحد، إذن القيمة الحرجة يتم إيجادها كما يلي:

$$\chi^2{}_{(\alpha,\upsilon)} = \chi^2{}_{(0.05,9)} = 16.919$$

نرفض H_0 عند مستوى الدلالة 0.05 إذا كان:

$$\chi^2 > 16.919$$

الآن نقوم بحساب إحصاء الاختبار:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(10-1)0.16}{0.04} = 36$$

بما أن: $36 > 16.919$ ، فإننا نرفض الفرضية الصفرية H_0 ونقبل الفرضية البديلة عند مستوى الدلالة 0.05؛

أي أن هذه النتيجة تعطينا مبررا قويا لرفض إدعاء البائع، ومن هنا يمكننا القول أن الإنحراف المعياري يزيد عن 0.2 سم.

6 . اختبار الفروض حول النسبة بين تبايني مجتمعين $\frac{\sigma_1^2}{\sigma_2^2}$

نظرية (10): إذا كان لدينا مجتمعان يتوزعان طبيعيا، وسحبنا من المجتمع الأول الذي تباينه σ_1^2 عينة عشوائية حجمها n_1، وكان تباينها s_1^2 ، ثم سحبنا من المجتمع الثاني الذي تباينه σ_2^2 عينة عشوائية أخرى مستقلة عن الأولى حجمها n_2 ، وكان تبايـنها s_2^2 ، وأردنا إجراء إختبار خاص بمقارنة σ_1^2 و σ_2^2، فإن إحصاء الإختبار المناسب هو المتغير العشوائي F ، حيث:

$$\mathrm{F} = \frac{s_1^2}{s_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2}$$

والتوزيع الاحتمالي لهذا المتغير يسمى **توزيع فيشر** بدرجتي حرية:

$\upsilon_1 = n_1 - 1$ (درجة حرية البسط).

$\upsilon_2 = n_2 - 1$ (درجة حرية المقـــــام).

وعليه أردنا اختبار الفرضية الصفرية:

$$\mathrm{H_0}:\sigma_1^2 = \sigma_2^2 \quad \text{أو} \quad \mathrm{H_0}:\frac{\sigma_1^2}{\sigma_2^2} = 1$$

مقابل الفرضية البديلة:

1. $\mathrm{H_1}:\ \sigma_1^2 \neq \sigma_2^2$ ؛ فإننا نرفض $\mathrm{H_0}$ عند مستوى الدلالة (α) إذا كانت:

$$\mathrm{F} < F_{(1-\frac{\alpha}{2},\upsilon_1,\upsilon_2)} \quad \text{أو} \quad \mathrm{F} > F_{(\frac{\alpha}{2},\upsilon_1,\upsilon_2)}$$

2. $\mathrm{H_1}:\ \sigma_1^2 > \sigma_2^2$ ؛ فإننا نرفض $\mathrm{H_0}$ عند مستوى الدلالة (α) إذا كانت:

$$\mathrm{F} > F_{(\alpha,\upsilon_1,\upsilon_2)}$$

3. $\mathrm{H_1}:\ \sigma_1^2 < \sigma_2^2$ ؛ فإننا نرفض $\mathrm{H_0}$ عند مستوى الدلالة (α) إذا كانت:

$$\mathrm{F} < F_{(1-\alpha,\upsilon_1,\upsilon_2)}$$

وما يجب الإشارة إليه هو أن:

$\boldsymbol{F}_{(\frac{\alpha}{2},\upsilon_1,\upsilon_2)}$: هي نقطة على المحور الأفقي لتوزيع فيشـــر والتي يقع إلـى يمينـها مساحة $(\frac{\alpha}{2})$.

$F_{(1-\frac{\alpha}{2},\upsilon_1,\upsilon_2)}$: هي نقطة على المحور الأفقي لتوزيع فيشر والتي يقع إلى يمينها مساحة $(1-\frac{\alpha}{2})$.

$F_{(\alpha,\upsilon_1,\upsilon_2)}$: هي نقطة على المحور الأفقي لتوزيع فيشر والتي يقع إلى يمينها مساحة (α).

$F_{(1-\alpha,\upsilon_1,\upsilon_2)}$: هي نقطة على المحور الأفقي لتوزيع فيشر والتي يقع إلى يمينها مساحة $(1-\alpha)$.

7. اختبار الفرضيات حول النسبة P في المجتمع

نظرية (11): إذا كانت $(x_1,x_2......,x_n)$ عينة عشوائية مسحوبة من مجتمع يخضع لتوزيع ذي الحدين، وكان حجم هذه العينة **كبير**، وأردنا إختبار الفرضية الصفرية: $H_0:P=P_0$ ، مقابل الفرضية البديلة:

1. $H_1:\ P\neq P_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$Z>z_{\alpha/2} \quad \text{أو} \quad Z<-z_{\alpha/2}$$

2. $H_1:\ P>P_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$Z>z_{\alpha}$$

3. $H_1:\ P<P_0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$Z<-z_{\alpha}$$

مع العلم أن إحصاء الاختبار Z في هذه الحالة نجده يساوي:

$$Z=\frac{\bar{P}-P_0}{\sqrt{\frac{P_0q_0}{n}}}$$

8. اختبار الفرضيات حول الفرق بين نسبتي المجتمعين

نظرية (11): إذا أخذت عينتين عشوائيتين مستقلتين من توزيع ذي الحدين وكان حجمهما n_1 و n_2 كبير بدرجة كافية. فتكون الفرضية الصفرية التالية:

$H_0:(P_1-P_2)=d_0$. وبالتالي فإن إحصاء الاختبار المناسب هو:

$$Z = [(\bar{p}_1 - \bar{p}_2) - d_0]/\sqrt{\frac{\bar{p}_1 \bar{q}_1}{n_1} + \frac{\bar{p}_2 \bar{q}_2}{n_2}}$$

وما يجب الإشارة إليه هو أنه إذا كانت $d_0 = 0$ في الفرضية الصفرية السابقة، فهذا يعني أن $P_1 = P_2$. فعندئذ يكون إحصاء الاختبار كما يلي:

$$Z = [(\bar{p}_1 - \bar{p}_2)]/\sqrt{\frac{\bar{P}\bar{q}}{n_1} + \frac{\bar{P}\bar{q}}{n_2}}$$

حيث:

$$\bar{P} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} = \frac{n_1\frac{x_1}{n_1} + n_2\frac{x_2}{n_2}}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

وتشير $\bar{P}$ الى **المتوسط المرجح** لنسبتي العينتيّن $\bar{p}_1$ و p_2 ، ويطلق عليها أيضا **بالنسبة التجميعية**.

ومنه أردنا اختبار الفرضية الصفرية: $H_0 : (P_1-P_2) = 0$ مقابل الفرضية البديلة:

1. $H_1 :(P_1-P_2) \neq 0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$Z > z_{\alpha/2} \quad \text{أو} \quad Z < -z_{\alpha/2}$$

2. $H_1 :(P_1-P_2) > 0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$Z > z_{\alpha}$$

3. $H_1 :(P_1-P_2) < 0$ ؛ فإننا نرفض H_0 عند مستوى الدلالة (α) إذا كانت:

$$Z < -z_{\alpha}$$