

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

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 $X_1, X_2, \dots, X_n$ حجمها n فإن الوسط الحسابي لهذه العينة هو الإحصاءة $\bar{X}$ ، حيث :

$$\bar{X} = \sum X_i / n \quad \dots\dots\dots(1)$$

بينما تباين العينة Sample variance هو الإحصاءة S^2 :

$$S^2 = \sum (x_i - \bar{X})^2 / n-1 \quad \dots\dots\dots(2)$$

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 المعادلة (2) بدلا من n وذلك لكي يكون التقدير الناتج تقديرا غير متحيزا (Unbiased Estimate) للمعلمة σ^2 ، أي يكون :

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

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: $(X_1, X_2, \dots, X_N)$ يكون :

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

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

مع العلم أن :

$$\sigma^2 = \Sigma(X_i - \mu)^2 / N \dots\dots\dots (5)$$

$$\mu = \Sigma X_i / N \dots\dots\dots(6)$$

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 معيناً ، 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 معيناً لهذه العينة وليكن على سبيل المثال الوسط الحسابي $\bar{X}$. 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 (X_i) ، هذه القيم تكون مجتمعا آخر عدد مفرداته أكبر 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 معيناً قد يختلف عن توزيع المجتمع الأصلي ، ويسمى هذا التوزيع

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

بتوزيع المعاينة لهذا المقياس سواء كان هذا المقياس هو الوسط الحسابي للعينة $\bar{X}$ أو الانحراف المعياري لها S .

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

مثال 1: 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 :

أ - احسب الوسط الحسابي (μ) والتباين (σ^2) .

ب- احسب الوسط الحسابي $\bar{X}$ لجميع العينات البسيطة الممكنة والتي حجم كل منها ثلاث وحدات، وكذلك حدد جدول التوزيع الاحتمالي للوسط الحسابي $\bar{X}$ ، ومنه احسب القيمة المتوقعة والتباين للمتغير $\bar{X}$.

ج - 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 :

أ- الوسط الحسابي

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

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

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

$$\sigma^2 = 4.5$$

ب- بما أن السحب تم بدون إرجاع فان العينات الممكنة عددها عشرة وذلك حسب :

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

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

والجدول الموالي يوضح مختلف هذه العينات وأوساطها الحسابية .

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

رقم العينة	القيم المختلفة للعينات X_i (x_1, x_2, x_3)	$\sum x_i$	$\bar{X}_i$
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

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

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

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

انطلاقاً من هذا الجدول فإن القيمة المتوقعة للوسط الحسابي ($\bar{X}$) هي:

$$E(\bar{X}) = \sum \bar{X}_i P(\bar{X}_i) \dots\dots\dots (7)$$

$$\begin{aligned} &= 3(1/10) + 3.5(1/10) + 4(2/10) + 4.5(2/10) + 5(2/10) + 5.5(1/10) \\ &\quad + 6(1/10) = 4.5 \end{aligned}$$

$$E(\bar{X}) = \mu = 4.5 \quad \text{مما سبق نستنتج أن :}$$

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

أما تباين الوسط الحسابي للعينة فيحسب بالعلاقة التالية :

$$\text{Var}(\bar{X}) = E(\bar{X})^2 - [E(\bar{X})]^2 \dots\dots\dots (8)$$

حيث:

$$E(\bar{X})^2 = \sum \bar{X}_i^2 \cdot P(\bar{X}_i)$$

$$= 3^2(1/10) + 3.5^2(1/10) + 4^2(2/10) + 4.5^2(2/10) + 5^2(2/10) +$$

$$(5.5)^2(1/10) + 6^2(1/10) = 21$$

$$\Rightarrow \text{Var}(\bar{X}) = 21 - (4.5)^2 = \mathbf{0.75}$$

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

$$S^2 = \sum (x_i - \bar{X})^2 / n - 1$$

$$= \sum (x_i - \bar{X})^2 / 3 - 1$$

ويمكن 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)			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

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

وانطلاقاً من هذا الجدول فانه 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 :

$$E(S^2) = \sum S^2 \cdot P(S^2) \dots \dots \dots (9)$$

$$= 2.25\left(\frac{3}{10}\right) + 5.25\left(\frac{4}{10}\right) + 9\left(\frac{1}{10}\right) + 9.75\left(\frac{2}{10}\right)$$
$$= 5.625$$

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

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

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

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

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

بفرض أنه أخذنا عينة عشوائية $(X_1, X_2, \dots, X_{ni})$ 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 :

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

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

إذا كان لدينا مجتمع مكون من عدد محدود من القيم $(X_1, X_2, \dots, X_N)$ فإن الوسط الحسابي والتباين لهذا المجتمع يكونان على الشكل التالي :

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

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

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

$$E(\bar{X}) = \mu_{\bar{X}} = \mu \quad \dots\dots\dots (10)$$

$$\text{Var}(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right) \quad \dots\dots\dots (11)$$

ومنه فالانحراف المعياري للمتغير $\bar{X}$ هو :

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

والمقدار $\left(\sqrt{\frac{N-n}{N-1}} \right)$ يسمى بمعامل تصحيح المجتمع المحدود، وهو دائما أقل من الواحد ، وكلما اقترب من الواحد فيمكن الاستغناء عنه. وعادة يستعمل معامل التصحيح إذا تحقق الشرط: $n \geq 5\% N$

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

إذا كانت $(X_1, X_2, \dots, X_n)$ هي مشاهدات عينة عشوائية مسحوبة من مجتمع غير محدود متوسطه μ وتباينه σ^2 ، فإن القيمة المتوقعة والتباين للمتغير $\bar{X}$ هما :

$$E(\bar{X}) = \mu_{\bar{X}} = \mu \quad \dots\dots\dots (12)$$

$$\text{Var}(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \quad \dots\dots\dots (13)$$

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

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

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

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 يتوقع أن تقترب $\bar{X}$ من μ كلما زاد حجم العينة n .

نظرية (1): إذا كانت $(X_1, X_2, \dots, X_n)$ هي مشاهدات عينة عشوائية مسحوبة من مجتمع طبيعي

وسطه μ وتباينه σ^2 ، فإن الوسط الحسابي $\bar{X}$ له توزيع طبيعي متوسطه μ ، وتباينه $\frac{\sigma^2}{n}$ ، أي أن

المتغير العشوائي: (14) $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ 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 600 غ.

المطلوب:

أ- أوجد الوسط الحسابي والتباين والانحراف المعياري للوسط الحسابي لأوزان الأطفال في العينة .

ب- أوجد احتمال أن الوسط الحسابي لأوزان الأطفال يزيد عن 3100 غ.

ج- أوجد احتمال أن الوسط الحسابي لأوزان الأطفال يقع ما بين 2700 غ و 3200 غ .

الحل:

نفرض أن $\bar{X}$ هو الوسط الحسابي لأوزان الأطفال في العينة.

$$\mu_{\bar{X}} = \mu = 2900g , \sigma^2 = (600)^2 = 360000 \quad \text{أ- لدينا:}$$

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} = \frac{360000}{9} = 40000g$$

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

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

ب- إيجاد الاحتمال التالي: $P(\bar{X} > 3100)$

نقوم أولاً بحساب المتغير العشوائي Z :

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

$$Z_{3100} = 1$$

ومنه نجد :

$$\begin{aligned} P(\bar{X} > 3100) &= P(Z > 1) = 0.5 - P(0 < Z \leq 1) \\ &= 0.5 - 0.3413 \end{aligned}$$

$$= 0.1587$$

ج- إيجاد الاحتمال التالي: $P(2700 < \bar{X} < 3200)$

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

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

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

$$= 0.4332 + 0.3413 = 0.7745$$

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

إذا كان لدينا مجتمع وسطه μ وتباينه σ^2 معلوم، وأخذت منه جميع العينات العشوائية ذات

الحجم n ، فما هو توزيع الوسط الحسابي $\bar{X}$ لهذه العينات حتى لم يكن توزيع المجتمع التوزيع الطبيعي ؟

للإجابة عن هذا السؤال يجب التطرق إلى النظرية التالية:

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

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

إذا كانت $(X_1, X_2, \dots, X_n)$ هي مشاهدات عينة عشوائية حجمها n أخذت من مجتمع إحصائي

وسطه μ وتباينه σ^2 ، فإن توزيع المعاينة للوسط الحسابي $\bar{X}$ يقترب من التوزيع الطبيعي بمتوسط μ وتباين $\sigma^2_{\bar{X}}$ كلما زاد حجم العينة تدريجياً ، ويكون ذلك أن يصل حجم العينة إلى 30 ، بعبارة أخرى

فإن المتغير (Z) حيث : $Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$ 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 : أ- أحسب احتمال أن يكون متوسط العينة أصغر من 78.

ب- أحسب احتمال أن لا يبعد متوسط العينة عن متوسط المجتمع بأكثر من 4 %.

الحل :

$$\begin{aligned}\mu_{\bar{X}} &= \mu = 75 \\ \sigma_{\bar{X}} &= \frac{\sigma}{\sqrt{n}} = \frac{13}{\sqrt{51}} = 1.82\end{aligned}$$

أ - حساب الاحتمال التالي : $P(\bar{X} < 78)$

نقوم أولاً بحساب المتغير العشوائي Z :

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

ومنه :

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

$$\begin{aligned}P(\bar{X} < 78) &= P(Z < 1.65) = 0.5 + P(0 < Z \leq 1.65) \\&= 0.5 + 0.4505 \\&= 0.9505\end{aligned}$$

ب - حساب احتمال أن لا يبعد متوسط العينة عن متوسط المجتمع بأكثر من 4 %.

$$\text{لدينا: } 4 \times 75 = 3 \%$$

إذن المراد حساب الاحتمال التالي:

$$\begin{aligned}P(72 < \bar{X} < 78) &= P\left(\frac{72-75}{1.82} < Z < \frac{78-75}{1.82}\right) \\&= P(-1.65 < Z < 1.65) \\&= 2P(0 < Z \leq 1.65) \\&= 0.901\end{aligned}$$

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

مقدمة لتوزيع T

في الكثير من التطبيقات المتعلقة باستعمال الوسط الحسابي $\bar{X}$ ، نجد أن الانحراف المعياري للمجتمع لا

يكون معلوماً ، 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

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

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

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

إن توزيع **T** يشبه التوزيع الطبيعي المعياري من حيث أنه متمائل ومركزة حول الصفر، ولكنه أكثر تشتتاً واختلافاً، 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 تقديراً دقيقاً جداً لـ σ

ويكون التشتت في **T** قليل جداً.

- أما إذا كانت n صغيرة جداً فإن S تصبح تقديراً غير دقيقاً لـ σ ، ومنه تظهر **T** تبايناً أكثر

وعليه فإن التشتت في توزيع **T** يعتمد على حجم العينة n .

فكلما زادت n فإن توزيع **T** يظهر تشتت أقل وأقل ويصبح مشابهاً أكثر فأكثر للتوزيع الطبيعي

المعياري ، وفي الحقيقة فإنها يصبحان متطابقين من الناحية النظرية كلما كبرت n واقتربت من المالا نهاية .

وهذا يعني أنه إذا كانت n كبيرة بدرجة كافية فإن التوزيع الطبيعي المعياري يعد تقريباً جيداً لتوزيع **T** ،

ويمكن أن يستخدم بدلاً منه إذا شئنا ذلك ، حيث نجد أن القاعدة المقبولة على نطاق واسع أن التقريب

يعد مقبولاً 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 $N(\mu, \sigma^2)$ وكان σ^2 غير

معلوم، وحجم العينة أقل من 30 ، وكان $\bar{X}$ هو الوسط الحسابي لهذه العينة و S هو انحرافها المعياري

فإن توزيع المعاينة للمتغير **T** ، حيث $T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$ هو توزيع **ستيودنت** بدرجات حرية $(n-1)$.

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

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: أ - مفترض أن معدل الوكالة (45 ميل / جالون) متحققا لهذه السيارة ، أوجد احتمال

أن متوسط الأميال / جالون في العينة يجب أن يكون 43.5 أو أقل ؟

ب - اعتمادا على بيانات العينة الحالية ، هل هناك سببا مقنعا للمنظمة لكي تشك في أن

معدل الوكالة متحققا لهذه السيارة.

الحل:

لدينا : $\mu = 45$, $s = 2.5$, $\bar{x} = 43.5$, $n = 25$

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

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

أ - حساب الاحتمال التالي : $P(\bar{X} \leq 43.5)$

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 ستودنت
حيث :

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

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

$$P(\bar{X} \leq 43.5) = P(T \leq -3) = 0.005 \quad \text{ومنه :}$$

$$P(\bar{X} > 43.5) = P(T > -3) = 0.995 \quad \text{وبالتالي نجد :}$$

ب - 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 للفيوق بين طريقتي القياس للأعمال في المنظمتين (أي وكالة لحماية البيئة ومنظمة المستهلكين).

2-3 . توزيع المعاينة للفرق بين متوسطي عينتين مستقلتين :

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

نظرية (4): إذا كان $\bar{X}_1, \bar{X}_2$ هما متوسطا عينتين عشوائيتين مستقلتين حجمها على التوالي هو 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:

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

$$V(\bar{X}_1 - \bar{X}_2) = \sigma_{\bar{X}_1 - \bar{X}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \dots \dots \dots (17)$$

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 :

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

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

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

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

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

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

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 : $P(\bar{X}_1 - \bar{X}_2 \geq 8000)$

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 :

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

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

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، فإن توزيع المعاينة 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 1500 دج وتباينه 600 دج ، وكان الإنفاق الاستهلاكي العائلي اليومي في ولاية بسكرة وسطه الحسابي 1000 دج 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 .

والمطلوب هو إيجاد الاحتمال التالي: $P(\bar{X}_1 - \bar{X}_2 > 510)$

بتطبيق النظرية (5) نجد :

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

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

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

$$Z = 3.43$$

$$P(\bar{X}_1 - \bar{X}_2 > 510) = P(Z > 3.43) = 0.5 - 0.4997 = 0.0003$$

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 **U**.

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 $\sigma_1^2 = \sigma_2^2$ ، فأن 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:

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

وبما أن التباين المشترك σ^2 مجهول القيمة فإننا نستطيع تقديره باستخدام تبايني العينتين S_1^2 و 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 $(n_1 - 1)$ ، $(n_2 - 1)$ كترجيحات بدلا من استخدام العينيتين: n_1 ، n_2 بشكل مباشر .

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

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

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

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:

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

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:

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

يخضع لتوزيع ستودنت (**t**) بدرجة حرية: $v = n_1 + n_2 - 2$

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 فأوجد احتمال أن الفرق بين متوسطي العينتين يكون

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

أقل من 3 .

الحل:

$$n_1=16, n_2=25, s_1^2=4, s_2^2=7$$

لدينا:

$$\mu_1=30, \mu_2=28$$

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

لدينا:

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

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 :

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

$$= P(T < 1.30)$$

$$n_1 + n_2 - 2 = 16 + 25 - 2 = 39$$

عند درجات الحرية :

$$P(\bar{X}_1 - \bar{X}_2 < 3) = 0.9$$

نجد :

2- حالة تبايني المجتمعين غير متساويين ومجهولين:

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

نظرية (7): إذا كان $\bar{X}_1$ و $\bar{X}_2$ هو 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 $(s_1^2$ و $s_2^2)$ وذلك كما يلي:

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

وعليه فإن المتغير:

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

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

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

مثال(08): إذا كانت نقاط طلبة السنة الثانية (ل.م.د) إدارة أعمال في مقياس الإحصاء لديها وسط حسابي قدره 15، وكانت نقاط طلبة السنة الثانية (ل.م.د) محاسبة في نفس المقياس لديها وسط حسابي قدره 10. وقمنا بسحب عينة من طلبة السنة الثانية (ل.م.د) إدارة أعمال حجمها 25 طالب ، وقمنا أيضا بسحب عينة من طلبة السنة الثانية (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:

$$n_1 = 25 , n_2 = 20 , s_1^2 = 6 , s_2^2 = 4 \quad \text{لدينا:}$$

$$\mu_1 = 15 , \mu_2 = 10$$

بما أن تبايني المجتمعين مجهولين وغير متساويين فإننا نستخدم توزيع **t** بدرجة حرية لها الصيغة المركبة

الموضحة في العلاقة (26) وذلك لإيجاد الاحتمال التالي :

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

ومنه يكون :

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

$$= P(T > 1.51)$$

عند درجات الحرية :

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

نجد:

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

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 تضاهي 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:

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

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

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

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

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

أما الانحراف المعياري لتباين العينة فيعطى بالصيغة التالية:

$$\sigma_{S^2} = \sigma^2 \cdot \sqrt{\frac{2}{n-1}} \dots\dots\dots (27)$$

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:

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

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

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

نظرية (8): إذا سحبت عينة عشوائية $(X_1, X_2, \dots, X_n)$ من مجتمع له توزيع طبيعي (هذا شرط أساسي) وسطه μ وتباينه σ^2 ، وكانت S^2 تمثل تباين العينة، فإن المتغير:

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

له توزيع كأي تربيع (Chi-square statistic) بدرجات حرية: $\nu = n-1$.

المتغير السابق (χ^2) هو دالة في S^2 . والرمز χ هو حرف يوناني كأي. أما درجات الحرية

$(n-1)$ المقترنة بالإحصاء كأي تربيع تعكس حقيقة أن هناك $(n-1)$ من درجات الحرية مقترنة بتباين العينة S^2 .

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

$$E(\chi^2) = n - 1 \dots\dots\dots (29)$$

وبخصوص منحني توزيع χ^2 فإنه غير متماثل، وهو منحني ملتو إلى اليمين (موجب الالتواء) ويقول هذا الالتواء كلما زادت قيمة درجة الحرية، والشكل (2) يوضح ذلك.

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

نستنتج أنه 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.

تسحب قيمة المتغير χ^2 لأي 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 يرمز لقيمة χ^2 التي تكون المساحة الواقعة على يمينها مساوية α والمناظرة لدرجات حرية ν بالرمز: $\chi^2_{(\alpha, \nu)}$

انظر الشكل (3) الموالي، والجدول الموضح في الملحق رقم (3) من ملاحق الجداول الإحصائية.

مثال(10): إذا كانت S^2 هو تباين عينة عشوائية ذات الحجم 4 وحدات مسحوبة من مجتمع له توزيع طبيعي تباينه 25.

المطلوب: أ- أوجد احتمال أن تباين العينة سوف يكون 2.5 أو أقل.

ب- أوجد احتمال أن تباين العينة سوف يتعدى 66.

الحل:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{3S^2}{25} \quad \text{أ- بتطبيق الصيغة (28) نجد :}$$

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

$$P(S^2 \leq 2.5) = 1 - P(S^2 > 2.5) \quad \text{ومنه يكون :}$$

$$\begin{aligned} \Rightarrow P(S^2 > 2.5) &= P\left(\frac{(n-1)S^2}{\sigma^2} > \frac{3(2.5)}{25}\right) \\ &= P(\chi^2 > 0.3) \end{aligned}$$

وباستخدام جدول توزيع كآي تربيع بالملحق رقم (3) نجد: $P(\chi^2 > 0.3) \cong 0.95$ أي أن :

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

$$P(S^2 \leq 2.5) = 1 - P(S^2 > 2.5) \quad \text{وفي الأخير نجد :}$$

$$= 1 - 0.95 = 0.05$$

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

ب- 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 : $v = 3$ ، نجد: $P(\chi^2 > 7.92) \cong 0.05$
أي أن :

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

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. و سنتطرق 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 ، n_2 مسحوبتين من مجتمعين لهما توزيعين طبيعيين ذو التباينين σ_1^2 ، σ_2^2 على الترتيب. فإن المتغير:

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

يكون له توزيع فيشر (F) بدرجات حرية (v_1, v_2) أي $(n_1 - 1, n_2 - 1)$.
و ما 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 ، ويرمز لها بـ: $v_1 = n_1 - 1$

- درجات حرية مقترنة بتباين العينة S_2^2 في المقام ، ويرمز لها بـ: $v_2 = n_2 - 1$

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

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

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

و 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 درجات الحرية. إن توزيع F مثل توزيع Z و توزيع T و توزيع χ^2 ، نجده متواجدا في معظم البرامج الإحصائية الجاهزة ، و لكي 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 $(v_1 = n_1 - 1)$ ، و نبحث عن درجات الحرية الخاصة بالمقام $(v_2 = n_2 - 1)$ ، و منه قيمة F التي تنتج من تقاطع درجتَي حرية البسط و المقام 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 احتمال أن يأخذ المتغير F قيمة لا تزيد عن : (أ) 0.36 ، (ب) 2.23

الحل:

لدينا : $v_1 = n_1 - 1 = 16 - 1 = 15$ ، $v_2 = n_2 - 1 = 20 - 1 = 19$

أ - باستخدام جدول توزيع F الموضح في الملحق رقم (4) يكون:

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

$$P(F_{(15,19)} > 0.36) = 0.975 \quad \text{وبالمقابل يكون :}$$

$$P(F_{(15,19)} \leq 2.23) = 0.95 \quad \text{ب - بنفس الطريقة نجد :}$$

$$P(F_{(15,19)} > 2.23) = 0.05 \quad \text{وبالمقابل يكون :}$$

3-5. توزيع المعاينة لنسبة العينة p -

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 :

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

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

$$q = 1 - P$$

وتعتبر 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 p وتحسب بالعلاقة التالية :

$$\bar{p} = \frac{\text{عددالعقيدات العينة التي تتوافر فيها الظاهرة المدروسة}}{\text{العدد الكلي لمفردات}} = \frac{X}{n}$$

نسبة العينة $\bar{p}$ ، كأى 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 μ_p والرمز σ_p^2 ، وبين نسبة المجتمع ، فنجد أن :

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

وعندما يكون المجتمع لا نهائيا أو مجتمعا غير محدودا أو كانت عملية السحب تتم مع الإرجاع

(أي $n \leq 0.05N$) فإن :

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

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

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

$$\sigma_p^2 = \frac{Pq}{n} \left(\frac{N-n}{N-1} \right) \dots\dots\dots(33)$$

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 ، لان 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 توزيع ذات الحدين (Binomial Distribution) بمعلمتين p , n ، في حالة سحب مفردات العينة مع الإرجاع.

ونعلم انه 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 : $np \geq 5$, $nq \geq 5$ ، فان المتغير العشوائي Z ، حيث :

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

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 $n = 40$

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

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

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 $\bar{p}$ سيكون قريباً من التوزيع الطبيعي بمتوسط وتباين قدرهما على التوالي كما يلي :

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

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

نستطيع التعبير عن الاحتمال المطلوب كما يلي :

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

حيث :

$$z_1 = \frac{\bar{p}_1 - P}{\sqrt{\sigma_{\bar{p}}^2}} = \frac{0.55 - 0.5827}{\sqrt{0.0061}} = -0.42$$

$$z_2 = \frac{\bar{p}_2 - P}{\sqrt{\sigma_{\bar{p}}^2}} = \frac{0.70 - 0.5827}{\sqrt{0.0061}} = 1.50$$

إذن :

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

$$= 0.1628 + 0.4332 = 0.5960$$

3-6. توزيع المعاينة للفرق بين نسبتين $(p_1 - p_2)$:

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 ، و P_2 ترمز لنسبة الظاهرة في المجتمع الثاني ، وعند عدم توافر بيانات عن مفردات كل من المجتمع الأول والمجتمع الثاني ، نقوم بالاستدلال

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

على المعلمة $(P_1 - P_2)$ أي استنتاجها باستخدام الفرق بين نسبي العينتين العشوائيتين المسحوبتين من هذين المجتمعين ، أي باستخدام الإحصائية $(p_1 - p_2)$ ، حيث p_1 هي نسبة الظاهرة في العينة العشوائية المسحوبة من المجتمع الأول ، و p_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 نسبي عينتين".

فإذا سحبنا من المجتمع الأول كل العينات العشوائية ذات الحجم n_1 ، وحسبنا نسبة الظاهرة المدروسة $\bar{p}_1$ لكل عينة ، وسحبنا من المجتمع الثاني كل العينات العشوائية ذات الحجم n_2 ، وحسبنا نسبة الظاهرة المدروسة $\bar{p}_2$ لكل عينة ، فإذا كانت العينات المسحوبة من المجتمع الأول مستقلة عن العينات المسحوبة من المجتمع الثاني ، و حسبنا كل الفروق بين نسب عينات المجتمع الأول ونسب عينات المجتمع الثاني ، أي حسبنا كل قيم $(\bar{p}_1 - p_2)$ ، فسنحصل على توزيع المعاينة للفرق بين نسبي عينتين $(p_1 - p_2)$ ، وإذا حسبنا الوسط الحسابي $\mu_{p_1-p_2}$ ، والتباين $\sigma_{\bar{p}_1-\bar{p}_2}^2$ لهذا التوزيع ، فنجد أن هناك علاقات تربط هذين المقياسين مع نسبة المجتمع الأول ونسبة المجتمع الثاني ، وذلك كما يلي:

$$\mu_{\bar{p}_1-\bar{p}_2} = P_1 - P_2 \dots\dots\dots (35)$$

فإذا كان المجتمع غير محدود أو كان السحب مع الإرجاع، و n_1 / N_1 ، n_2 / N_2 كليهما أقل من أو يساوي 0.05 فإن:

$$\sigma_{\bar{p}_1-\bar{p}_2}^2 = \frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2} \dots\dots\dots (36)$$

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

$$\sigma_{\bar{p}_1-\bar{p}_2}^2 = \frac{P_1 q_1}{n_1} \left(\frac{N_1 - n_1}{N_1 - 1} \right) + \frac{P_2 q_2}{n_2} \left(\frac{N_2 - n_2}{N_2 - 1} \right) \dots\dots\dots (37)$$

ومن هنا نصل إلى النظرية التالية :

المحور الأول.....توزيعات المعاينة (The Sampling Distributions)

نظرية (11): إذا كان لدينا عينتان مستقلتان كبيرتا الحجم تم سحبهما من مجتمعين، فوفقا لنظرية النهاية المركزية، يكون توزيع المعاينة للفرق بين نسبي 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:

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

سيتم 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 و 30 سنة، موزعين حسب الجنس كما يلي: 2157136 ذكرا منهم 221914 يحملون شهادة جامعية ، و 2067508 أنثى منهم 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 : اوجد احتمال أن يكون الفرق بين نسبي العينتين أكبر أو يساوي 5 %.

الحل:

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

$$P_1 \text{ تمثل نسبة الذين يحملون شهادة جامعية في المجتمع الأول } = \frac{221914}{2157136} = 0.10$$

$$P_2 \text{ تمثل نسبة الذين يحملون شهادة جامعية في المجتمع الثاني } = \frac{144423}{2067508} = 0.07$$

n_1 حجم العينة الأولى = 2000 ذكر.

n_2 حجم العينة الثانية = 1500 أنثى.

والاحتمال المطلوب هو : $P [(\bar{p}_1 - \bar{p}_2) \geq 0.05] = ?$

بما أن n_1 و n_2 كبيرتان ، فإن توزيع المعاينة للإحصائية ($p_1 - p_2$) سيكون قريبا من التوزيع الطبيعي ، وبالتالي فإن الاحتمال المطلوب يتم حسابه كما يلي :

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

$$Z = \frac{(\bar{p}_1 - \bar{p}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}} = \frac{(0.05) - (0.10 - 0.07)}{\sqrt{\frac{0.10 \times 0.90}{2000} + \frac{0.07 \times 0.93}{1500}}} \quad \text{حيث أن:}$$

$$= 2.13$$

إذن:

$$\begin{aligned} P[(\bar{p}_1 - \bar{p}_2) \geq 0.05] &= P(Z \geq 2.13) \\ &= 0.5 - 0.4834 \\ &= 0.0166 \end{aligned}$$