الفصل الثالث

بعض توزيعات المعاينة الهامة
Some Important Sampling Distributions

مقدمة:

عند حساب إحصائيات (إنها نواتج عن العينات) بعض المؤشرات المجهولة لمجتمع إحصائي، كالمتوسط أو التباين مثلاً، نستخدم غالبا أسلوب المعاينة بدلاً من استخدام بيانات الحصر الشامل. فمثلاً لتقدير متوسط درجات مادة الإحصاء التطبيقي لطلاب السنة الثالثة بكلية الاقتصاد نأخذ عينة حجمها n ونستخدم متوسط العينة $\bar{x}$ كتقدير لها .

والسؤال المطروح هو: ما هو توزيع الإحصائية $\bar{x}$ (متوسط درجات مادة الإحصاء التطبيقي لطلاب السنة الثالثة بكلية الاقتصاد) عندما نأخذ جميع العينات الممكنة ذات الحجم المتساوي n من هذا المجتمع الإحصائي (جميع طلاب السنة الثالثة بكلية الاقتصاد) ؟ ، مثل هذا التوزيع يسمى عادة توزيع المعاينة للوسط الحسابي والسؤال الذي يطرح نفسه هو: ما هي خصائص هذا التوزيع كالتوقع والتباين ؟

– **تعريف العينة والمعاينة:** تعرف العينة على أنها جزء من المجتمع الإحصائي وتمثل هذا المجتمع تمثيلاً صحيحاً. أما المعاينة فهي الطريقة التي يتم بموجبها الحصول على هذه العينة.

– **توزيع المعاينة:** يدعى توزيع كل القيم الممكنة لبعض الإحصائيات، والمحسوبة من عينات عشوائية متساوية الحجم ومسحوبة من نفس المجتمع بتوزيع المعاينة لتلك الإحصاءات. ويمكن بناء توزيع المعاينة بشكل تجريبي عندما تكون المعاينة من مجتمع

متقطع منته بإتباع الخطوات التالية:

من مجتمع محدود حجمه N نسحب بشكل عشوائي كل العينات الممكنة بحجم n .

- نحسب الإحصائيات المفيدة والمطلوبة من كل عينة.
- نضع في العمود الأول كل القيم المختلفة لتلك الإحصائية.
- في العمود الثاني نضع التكرارات المقابلة لحدوث كل قيمة مشاهدة مختلفة تأخذها الإحصائية.
- وفي العمود الأخير نحسب التكرارات النسبية، ومن هذه البيانات يمكننا بناء المدرج والمنحنى التكراري للتوزيع والتعرف على التوزيع نفسه.

تستخدم المعاينة العشوائية البسيطة Simple Random Sampling عندما تعطى الفرصة نفسها لجميع مفردات المجتمع الإحصائي ليتم اختيارها ضمن مفردات العينة.

إن عدد العينات المختلفة بحجم n والتي يمكن سحبها من مجتمع منته ذي حجم N، عندما يكون السحب بدون إعادة، يكون مساويا إلى C_N^n عينة، فمثلاً يوجد 28 عينة مختلفة ذات حجم 2 من مجتمع يضم 8 مفردات، حيث: $C_N^n = \dfrac{N!}{(N-n)!\times n!}$

أما إذا كان السحب مع الإعادة فإن عدد العينات الممكنة هو N^n أي يوجد 64 عينة مختلفة[1] بحجم 2 يمكن سحبها من مجتمع يضم ٨ عناصر .

أما العينة العشوائية البسيطة فهي تلك العينة التي يتم اختيارها من مجتمع إحصائي منته بحيث يكون لكل عينة من العينات C_N^n المختلفة الاحتمال ($1/C_N^n$) نفسه في الاختيار عندما يكون السحب بدون إعادة.

ويكون لكل عينة من العينات المختلفة N^n الاحتمال نفسه في الاختيار ويساوي ($1/N^n$) عندما يكون السحب مع الإعادة.

---

[1] وتقول بعض المراجع أن هذه العينات غير مختلفة برأيهم لأن b , a وكذلك a , b هما نفس العينة ، كما أن العينة التي تضم فريد وأحمد هي نفسها العينة التي تضم احمد وفريد .

§1- توزيع المعاينة للوسط الحسابي Sampling Distribution of The Mean :

عملياً نسحب عينة عشوائية ونحسب منها إحصائيات العينة كالوسط الحسابي مثلاً ويكون مقداراً ثابتاً، أما إذا سحبنا عدة عينات بنفس الحجم من نفس المجتمع فمن المتوقع أن يأخذ الوسط الحسابي قيماً مختلفة في هذه العينات، وكذلك قد تكون مختلفة عن الوسط الحسابي للمجتمع، أما إذا قمنا بسحب كل العينات الممكنة والتي لها نفس الحجم n وحسبنا الأوساط الحسابية لهذه العينات، فان الوسط الحسابي العام $\bar{X}$ لمتوسطات هذه العينات يجب أن يساوي تماماً الوسط الحسابي للمجتمع μ، ويدعى توزيع الأوساط الحسابية بتوزيع المعاينة للوسط الحسابي.

مثال (3-1)

بفرض أن لدينا مجتمعاً حجمه $N = 5$ يتألف من درجات القسم العملي لمادة الإحصاء اللامعلمي لـ 5 طلاب من قسم الإحصاء والبرمجة بكلية الاقتصاد وكانت درجاتهم كما يلي: 6 ، 8 ، 10 ، 12 ، 14 سنة .

والمطلوب:

- حساب متوسط هذا المجتمع μ وانحرافه المعياري σ.
- كتابة جميع العينات الممكنة بحجم $n = 2$ في حالة السحب بدون إعادة وفي حالة السحب مع الإعادة .
- إيجاد تابع الكثافة الاحتمالي للمتغير العشوائي $\bar{X}$ أي $f(\bar{X})$.
- إيجاد توقع وتباين المتغير العشوائي $\bar{X}$ أي حساب متوسط توزيع المعاينة للوسط الحسابي $\mu_{\bar{X}}$، وتباين توزيع المعاينة للوسط الحسابي $\sigma^2_{\bar{X}}$.

الحل:

لحساب متوسط المجتمع وتباينه نستخدم العلاقات التالية:

$$\mu = \frac{\sum X_i}{N} = \frac{50}{5} = 10$$

$$\sigma^2 = \frac{\sum (X_i - \mu)^2}{N} = \frac{40}{5} = 8$$

ويمكن حساب مقياس آخر للتشتت يدعى تباين العينة:

$$S^2 = \frac{\sum (X_i - \bar{x})^2}{n-1} = \frac{40}{4} = 10$$

- إن جميع العينات الممكنة بحجم $n = 2$ والممكن سحبها من هذا المجتمع موضحة بالجدول رقم (3-1) .

جدول (3-1)

جميع العينات الممكنة بحجم n = 2

		السحب الثاني				
السحب الأول		**6**	**8**	**10**	**12**	**14**
	6	(6 , 6)	(6 ,8)	(6 , 10)	(6 , 12)	(6 , 14)
	8	(8 , 6)	(8 ,8)	(8 , 10)	(8 , 12)	(8 , 14)
	10	(10 , 6)	(10 ,8)	(10 , 10)	(10 , 12)	(10 , 14)
	12	(12 , 6)	(12 ,8)	(12 , 10)	(12 , 12)	(12 , 14)
	14	(14 , 6)	(14 ,8)	(14 , 10)	(14 , 12)	(14 , 14)

عندما يكون السحب مع الإعادة فإن عدد العينات الممكن سحبها سيكون $N^n = 25$ عينة ممكنة موضحة في الجدول رقم (3-1).

إن متوسطات هذه العينات نوردها في الجدول رقم (3-2)، حيث تم حساب متوسط كل عينة عن طريق جمع قيم مفرداتها (كل عينة تحوي مفردتين) مقسوماً على عددها.

جدول (3-2)

عدد العينات الممكنة في حالة السحب مع الإعادة[٢]

6	7	8	9	10
7	8	9	10	11
8	9	10	11	12
9	10	11	12	13
10	11	12	13	14

---

[٢] تبقى جميع العلاقات الرياضية المستخدمة صحيحة إذا استخدمنا C^n_{N+n-1} بدلاً من N^n في حالة السحب مع الإعادة، وبالتالي يمكن حذف جميع العناصر الواقعة تحت القطر الرئيسي للجدول.

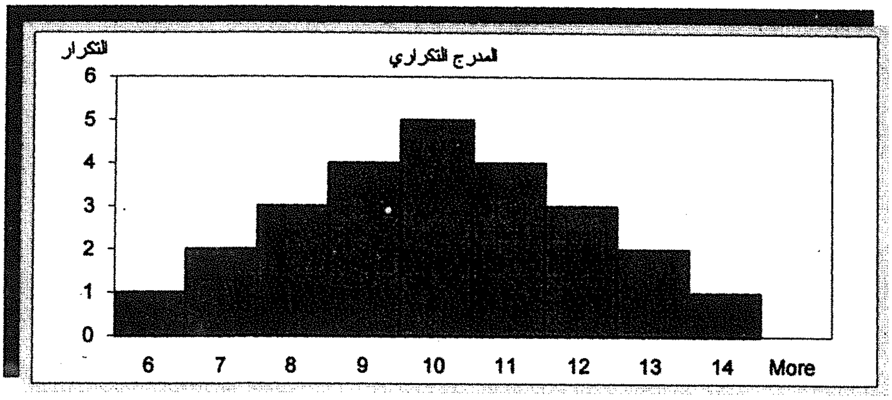
أما توزيع المعاينة للوسط الحسابي $\overline{X}$ فهو موضح بالجدول رقم (3-3)، لقد حاولنا بناء توزيع المعاينة لـ $\overline{X}$ عن طريق وضع القيم المختلفة لـ $\overline{X}$ في عمود واحد وتكراراتها المقابلة في عمود أخر.

جدول (3-3)

توزيع المعاينة للوسط الحسابي $\overline{X}$ في حالة السحب مع الإعادة

$\overline{X}$	التكرار	التكرار النسبي $f(\overline{X})$
6	1	1/25
7	2	2/25
8	3	3/25
9	4	4/25
10	5	5/25
11	4	4/25
12	3	3/25
13	2	2/25
14	1	1/25
المجموع		1

ويوضح الجدول أعلاه أن شرطي التابع الاحتمالي محققان حيث أن احتمال أي قيمة من قيم $\overline{X}$ أكبر من الصفر ومجموع الاحتمالات يساوي الواحد. والشكل التالي يوضح أن المتغير العشوائي $\overline{X}$ يتبع التوزيع الطبيعي.



شكل رقم (2) : المدرج التكراري للمتغير العشوائي $\overline{X}$

- إن توقع $\bar{X}$ أي متوسط توزيع المعاينة $\mu_{\bar{X}}$ يحسب كما يلي:

$$\mu_{\bar{X}} = \frac{\sum \bar{X}_i}{N^n} = \frac{6+7+8+....}{25} = \frac{250}{25} = 10$$

وتشير إلى أن:

- متوسط توزيع المعاينة $\mu_{\bar{X}}$ يساوي تماماً ودوماً متوسط المجتمع الإحصائي.

- أما تباين $\bar{X}$ أي تباين توزيع المعاينة $\sigma^2_{\bar{X}}$ فيحسب كما يلي:

$$\sigma^2_{\bar{X}} = \frac{\sum(\bar{X}_i - \mu_{\bar{X}})^2}{N^n}$$

$$= \frac{(6-10)^2 + (7-10)^2 + (8-10)^2}{25} = \frac{100}{25} = 4$$

هنا تباين توزيع المعاينة لا يساوي تباين المجتمع، وإنما يساوي تباين المجتمع مقسوماً على حجم العينة، في حين يدعى الجذر التربيعي لتباين توزيع المعاينة بالخطأ المعياري للمتوسط أي أن: تباين توزيع المعاينة يساوي:

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} = \frac{8}{2} = 4$$

والخطأ المعياري يساوي:

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

- عندما يكون السحب بدون إعادة فإن عدد العينات الممكنة (تجاهل الترتيب والتكرار) يساوي إلى 10 عينات مختلفة كما يلي:

$$C_N^n = \frac{N!}{n! \times (N-n)!} = \frac{5!}{2! \times (3!)} = 10$$

وهي موضحة تحت القطر الرئيسي في الجدول (3-1) أو فوق القطر الرئيسي.

- إن توقع $\bar{X}$ أي متوسط توزيع المعاينة $\mu_{\bar{X}}$ في حالة السحب بدون إعادة يحسب

كما يلي:

$$\mu_{\bar{x}} = \frac{\sum \bar{X}_i}{C_N^n} = \frac{7+8+9+....}{10} = \frac{100}{10} = 10 = \mu$$

إن متوسط توزيع المعاينة $\mu_{\bar{X}}$ في حالة السحب بدون إعادة يساوي تماماً ودوماً متوسط المجتمع الإحصائي.

- أما تباين $\bar{X}$ أي تباين توزيع المعاينة $\sigma^2_{\bar{X}}$ في حالة السحب بدون إعادة فيحسب كما يلي:

$$\sigma^2_{\bar{x}} = \frac{\sum(\bar{X}_i - \mu_x)^2}{C_N^n}$$

$$= \frac{(7-10)^2 + (8-10)^2 + (9-10)^2}{10} = \frac{30}{10} = 3$$

نلاحظ هنا في حالة السحب بدون إعادة، أن تباين توزيع المعاينة لا يساوي تباين المجتمع مقسوماً على حجم العينة لأن :

$$\sigma^2_{\bar{X}} = 3 \neq \frac{8}{2} = 4$$

بل تحقق علاقة أخرى وهي ضرب تباين توزيع المعاينة في حالة السحب مع الإعادة $\frac{\sigma^2}{n}$ بمعامل التصحيح $\frac{N-n}{N-1}$ أي :

ومنه:

$$\frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} = \frac{8}{2} \cdot \frac{5-2}{4} = 3$$

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

نظرية : توزيع المعاينة النظري للمتوسط $\overline{X}$ بالنسبة لعينات عشوائية ذات حجم n مأخوذة من مجتمع متوسطه μ وانحرافه المعياري σ يكون مساويا لمتوسط المجتمع μ (في حالتي السحب بدون إعادة والسحب مع الإعادة) ، أما انحرافه المعياري $\sigma_{\overline{X}}$ (ويسمى عادة الخطأ المعياري) فيكون: في حالة السحب بدون إعادة:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} . \sqrt{\frac{N-n}{N-1}} \qquad (3-1)$$

وفي حالة السحب مع الإعادة:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \qquad (3-2)$$

1-1 نظرية النهاية المركزية Central Limit Theorem :

بفرض أننا سحبنا عينة عشوائية حجمها n من مجتمع ما متوسطه μ وانحرافه المعياري σ معلوم فإن المتغير العشوائي :

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

يتبع تقريباً التوزيع الطبيعي المعياري إذا كانت n أكبر أو تساوي ٣٠ مهما كان توزيع المجتمع .

- يتبع التوزيع الطبيعي المعياري (مهما كان حجم العينة) إذا كان توزيع المجتمع المسحوبة منه تلك العينة طبيعياً (تباين المجتمع معلوم) .

مثال (3–1) :

لنفترض أن درجات العملي لمادة نظم التشغيل لطلاب السنة الثانية بكلية الاقتصاد تتبع تقريبا توزيعاً طبيعياً بمتوسط 18.56 درجة وبانحراف معياري 1.37 درجة ، سحبت

عينة عشوائية تضم عشرة أحداث n = 10 ، فما هو احتمال أن يكون متوسط درجات العملي لمادة نظم التشغيل لدى هذه العينة أكبر من 19 درجة ؟

الحل :

المطلوب هو حساب احتمال أن يكون متوسط درجات العملي أكبر من 19 درجة ، أي حساب احتمال المقدار $P(\overline{X} \geq 19)$.

بما أن توزيع المجتمع الإحصائي هو طبيعي تقريباً فالمتغير العشوائي $\overline{X}$ يتبع توزيعاً طبيعياً بمتوسط $\mu = 18.56$ ، وانحراف معياري (يسمى الخطأ المعياري) :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.27}{\sqrt{10}} = 0.402$$

وبعد تحديد متوسط وتباين توزيع المعاينة للوسط الحسابي يمكننا حساب الاحتمال المطلوب كما يلي:

$$P(\overline{X} \geq 19) = P(Z \geq \frac{\overline{X} - \mu_{\bar{x}}}{\sigma / \sqrt{n}})$$
$$= P(Z \geq \frac{19 - 18.56}{0.402} = 1.09) = 0.1379$$

وبالرجوع إلى جدول التوزيع الطبيعي نجد أن مساحة المنطقة الواقعة يمين النقطة 1.09 تساوي 0.1379 ، لذلك نقول بأن احتمال أن يكون متوسط عينة (بحجم n = 10) أكبر من 19 درجة يساوي إلى 0.1379.

مثال (3−2) :

إذا كان المتوسط والانحراف المعياري لدرجات الذكاء لطلاب السنة الثانية بكلية الاقتصاد هو 120 و 15 درجة على الترتيب. المطلوب حساب احتمال أن يكون متوسط درجات الذكاء لعينة عشوائية مؤلفة من 50 طالباً ، محصوراً بين 115 و 125 درجة ؟

الحل:

إن المتغير العشوائي $\bar{X}$ يتبع توزيعاً طبيعياً بمتوسط $\mu = 120$ درجة وانحراف معياري قدره : $\frac{15}{\sqrt{50}} = 2.12$ درجة ، أي:

$$\bar{X} \mapsto N(120, \frac{15}{\sqrt{50}})$$

وبناء عليه يمكن أن نكتب:

$$P(115 \le \bar{X} \le 125) = P(z_1 \le Z \le z_2)$$

نحسب القيم المعيارية الجديدة بموجب العلاقات التالية:

$$z_1 = \frac{115-120}{15/\sqrt{50}} = \frac{-5}{2.12132} = -2.36$$

$$z_2 = \frac{125-120}{15/\sqrt{50}} = \frac{+5}{2.12132} = +2.36$$

وبالتعويض يمكن أن نحسب الاحتمال المطلوب كالتالي:

$$\begin{aligned} P(115 \le \bar{X} \le 125) &= P(-2.36 \le Z \le 2.36) \\ &= F(2.36) - F(-2.36) = F(2.36) - 1 + F(2.36) \\ &= 2F(2.36) = 0.9818 \end{aligned}$$

أي إذا أخذنا عينة عشوائية حجمها 50 طالباً فهناك احتمال أكثر من 0.98 بأن يكون متوسط درجات الذكاء لديهم مختلفاً عن 120 درجة بحوالي 5 درجات.

§-2 توزيع المعاينة للفرق بين متوسطي عينتين

Sampling Distribution of the Difference between Two Sample Means

قد يرغب باحث في معرفة درجة الاختلاف بين متوسطي مجتمعين وذلك من خلال دراسة عينتين مسحوبتين من المجتمعين المذكورين ، فمثلاً قد يرغب فريق بحث اجتماعي اقتصادي معرفة ما إذا كان متوسط درجات سمة العصابية هو أعلى لدى

أطفال المرحلة الابتدائية الذين ينتمون لأسر تعاني من مشاكل اجتماعية واقتصادية (طلاق ، فقر) من مستواه لدى الأطفال الآخرين الذين ينتمون لأسر لا تعاني من أية مشاكل اجتماعية واقتصادية ، فإذا تمكن فريق البحث الاجتماعي الاقتصادي أن يستنتج بأن متوسطي المجتمعين هو مختلف ، فإنهم قد يرغبون بالوقوف على أسباب هذا الاختلاف أو في معرفة حجم هذا الاختلاف ، لتحديد ذلك فإن معرفة توزيع المعاينة للفرق بين متوسط عينتين سيكون مفيداً جداً ، إلا أننا لن ندخل في التفاصيل الرياضية لهذا التوزيع بل سنحاول التعرف على خصائصه وتوضيح ذلك بالمثال التالي.

مثال (3-3):

لنفترض أن لدينا مجتمعان إحصائييان، " مجتمع 1 " اختبرت فيه بعض الشروط المتعلقة بالتخلف العقلي " ومجتمع 2 " لم تختبر فيه هذه الشروط. ويُعتقد أن درجات الذكاء في كل من المجتمعين تتبع تقريبا توزيعاً طبيعياً بانحراف معياري يساوي 20 درجة، سحبنا عينة تضم 15 شخصاً من كل مجتمع وحسبنا متوسط درجات الذكاء لأفراد كل عينة على حدة فحصلنا على النتائج التالية: $\bar{X}_1 = 92$ ، $\bar{X}_2 = 105$

والمطلوب:

إذا لم يكن هناك فرق بين المتوسط الحقيقي لدرجات الذكاء في المجتمعين، ما هو احتمال وجود فرق كبير بين متوسط العينتين $(\bar{X}_1 - \bar{X}_2)$ ؟

الحل:

للإجابة عن هذا السؤال نحتاج لمعرفة طبيعة توزيع المعاينة للإحصائيات ذات الصلة، وهي الفرق بين متوسط العينتين، وللوصول لذلك نسحب من المجتمع الأول كل العينات الممكنة (عددها $C_{N_1}^{n_1}$ عينة) والتي حجم كل منها 15 ونحسب متوسط هذه العينات، وبشكل مشابه نسحب من المجتمع الثاني والذي حجمه يساوي 15، كل العينات الممكنة وعددها $C_{N_2}^{n_2}$ عينة حيث:

$$C_{N_2}^{n_2} = \frac{N_2!}{n_2!(N_2 - n_2)!}$$

نحسب متوسطات هذه العينات، ومن ثم نأخذ كل الأزواج الممكنة من متوسطات العينات واحدة من المجتمع الأول والأخرى من المجتمع الثاني ونأخذ فروقات هذه الأزواج ونرتبها في جدول يتضمن الفروقات بين الأوساط الحسابية وتكراراها المقابلة، ومنها نحسب التكرار النسبي ثم نرسم المدرج التكراري للفروقات نحصل على توزيع طبيعي متوسطه يساوي: $\mu_1 - \mu_2$ ويمثل الفرق بين متوسطي المجتمعين .

وتباينه يساوي:

$$(\sigma_1^2 / n_1) + (\sigma_2^2 / n_2)$$

أما الخطأ المعياري للفرق بين متوسطي العينتين فيساوي:

$$\sqrt{(\sigma_1^2 / n_1) + (\sigma_2^2 / n_2)}$$

نتيجة:

إذا كان لدينا مجتمعان طبيعيان متوسطهما μ_1, μ_2 وتباينهما σ_1^2, σ_2^2 على الترتيب، فإن توزيع المعاينة للفرق بين متوسطي عينتين $(\overline{X}_1 - \overline{X}_2)$ مستقلتين بحجم n_1, n_2 ومسحوبتين من هذين المجتمعين هو توزيع طبيعي متوسطه يساوي دوماً الفرق بين متوسطي المجتمعين أي يساوي: $\mu_1 - \mu_2$ وتباينه يساوي مجموع حاصل قسمة تباين المجتمع الأول على حجم العينة الأولى و حاصل قسمة تباين المجتمع الثاني على حجم العينة الثانية أي يساوي: $(\sigma_1^2 / n_1) + (\sigma_2^2 / n_2)$.

في مثالنا سيكون لدينا توزيع طبيعي متوسطه الصفر (إن لم يكن هناك فرق بين متوسطي المجتمعين الحقيقيين) وتباينه مساو إلى:

$$(\sigma_1^2 / n_1) + (\sigma_2^2 / n_2) = (400/15) + (400/15) = 53.33$$

ويمكن تحويل التوزيع الطبيعي الموصوف أعلاه إلى توزيع طبيعي معياري باستخدام الصيغة التالية:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2 / n_1) + (\sigma_2^2 / n_2)}} \qquad (3-2)$$

والمنطقة المحصورة تحت منحنى الفرق $(\overline{X}_1 - \overline{X}_2)$ تبين لنا الاحتمال المطلوب والمتمثل بالمساحة الواقعة إلى يسار المقدار:

$$(\overline{X}_1 - \overline{X}_2) = 92 - 105 = -13$$

والتي يقابلها قيمة معيارية z (بفرض عدم وجود فرق بين متوسط المجتمعين) تساوي:

$$Z = \frac{(92-105)-0}{\sqrt{53.33}} = \frac{-13}{7.3} = -1.78$$

ولحساب احتمال وجود فرق كبير بين متوسط العينتين نكتب:

$$P((\overline{X}_1 - \overline{X}_2) \geq -13) = P(z \geq -1.78)$$

وبالرجوع إلى جدول التوزيع الطبيعي المعياري نجد أن مساحة المنطقة الواقعة إلى يسار النقطة 1.78- تساوي 0.0375، أي أن احتمال وجود فرق كبير بين متوسط العينتين أكبر من 13 درجة يساوي 0.0375.

إن الإجراء الذي اتبعناه حتى الآن يكون صحيحاً حتى في حالة عدم تساوي حجمي العينتين المسحوبتين، وكذلك عندما يكون تباين المجتمعين مختلفين أي: $\sigma_1^2 \neq \sigma_2^2, \ldots\ldots, n_1 \neq n_2$.

1–2 المعاينة من مجتمعات غير طبيعية Sampling from nonnormal Population

في حالات كثيرة يواجه الباحث إحدى المشاكل التالية: ضرورة سحب عينات من مجتمعات غير طبيعية أو من مجتمعات توزيعاتها غير معروفة، والحل بهذه الحالة يكون بسحب عينات كبيرة الحجم أكبر من ٣٠ وبتطبيق نظرية النهاية المركزية يكون توزيع المعاينة للفرق بين متوسط العينتين طبيعياً تقريبا بمتوسط يساوي $\mu_1 - \mu_2$ وتباين يساوي:

$$(\sigma_1^2 / n_1) + (\sigma_2^2 / n_2)$$

مثال (3−4):

لنفترض أن متوسط طول الزيارة المترلية لمهندس معلوماتية يهتم بصيانة وإصلاح أجهزة الحاسوب داخل المنازل، لزبائن من النوع الأول هو 45 دقيقة بانحراف معياري قـــدره ١٥ دقيقة، ومتوسط طول الزيارة المترلية لزبائن من النوع الثاني هو 30 دقيقة بانحراف معياري قدره 20 دقيقة. إذا قام مهندس معلوماتية بإجراء زيارات عشـــوائية لــــ 35 عائلة من النوع الأول و لـــ 40 عائلة من النوع الثاني. فما هو احتمال أن يختلف متوسط طول الزيارة المترلية بين المجموعتين بمعدل 20 دقيقة أو أكثر ؟

الحل:

بما أن حجم العينة كبير ففي الحالتين يكون توزيع المعاينة للفرق بين متوسط العينـــتين طبيعياً تقريبا بمتوسط يساوي:

$$\mu_1 - \mu_2 = 45 - 30 = 15$$

أما تباين توزيع المعاينة فيساوي بهذه الحالة:

$$(\sigma_1^2 / n_1) + (\sigma_2^2 / n_2) = (15)^2 / 35 + (20)^2 / 40 = 16.4286$$

والمنطقة المحصورة تحت منحنى توزيع المعاينة للفرق: $(\overline{X}_1 - \overline{X}_2)$ تبين لنا الاحتمـــال المطلوب والمتمثل بالمساحة الواقعة إلى يمين المقدار 20 . والتي يقابلها قيمة معيارية z :

$$Z = \frac{20 - 15}{\sqrt{16.4286}} = \frac{5}{4.05} = +1.23$$

ولحساب احتمال أن يختلف متوسط طول الزيارة المترلية بين المجمـــوعتين بمعـــدل 20 دقيقة أو أكثر نكتب:

$$\begin{aligned} P((\overline{X}_1 - \overline{X}_2) \geq 20) &= P(z \geq +1.23) \\ &= 1 - P(z \leq 1.23) = 0.1093 \end{aligned}$$

ومن جدول التوزيع الطبيعي المعياري نجد أن المساحة الواقعة إلى يمين القيمة المعيارية $z = 1.23$ تساوي 0.1093.

§-3 توزيع المعاينة للنسبة في العينة

Sampling Distribution of the Sample Proportion

غالباً ما تواجهنا بعض المشاكل التي نحتاج فيها إلى معرفة نسبة مفردات المجتمع التي تتمتع بخاصية معينة مثل تحديد نسبة عملاء أحد المحلات التجارية الذين يشترون سلعة معينة، أو نسبة الأشخاص المؤمَنين صحياً والذين يرتادون الصيدليات النقابية وذلك من خلال سحب عينة من المجتمع الإحصائي.

يمكن بناء توزيع المعاينة للنسبة في العينة بشكل تجريبي وذلك بأن نسحب من مجتمع منته كل العينات الممكنة بأحجام محددة، ونحسب من أجل كل عينة نسبة العينة $\hat{p}$ ومن ثم نشكل جدول التوزيع التكراري لنسب العينة $\hat{p}$ الذي يتضمن القيم المختلفة للنسب وتكراراها المقابلة وكذلك تكراراها النسبية والذي يمثل توزيع المعاينة لـ $\hat{p}$.

- عندما يكون حجم العينة كبيرا فإن توزيع المعاينة لنسبة العينة سيكون تقريبا طبيعياً (نظرية النهاية المركزية) بمتوسط يساوي :

$\mu_{\hat{p}} = P$ متوسط نسب كل العينات الممكنة يساوي النسبة الحقيقية للمجتمع .

وتباين يساوي:

$$\sigma_{\hat{p}}^2 = P(1-P)/n$$

وللإجابة عن السؤال الاحتمالي المتعلق بالنسبة P نستخدم الصيغة التالية:

$$z = \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \qquad (3-3)$$

- والسؤال الذي يطرح نفسه، كم يجب أن يكون حجم العينة ليكون صحيحا استخدام التقريب الطبيعي ؟ إن المعيار الأكثر شيوعا هو أن يكون كل من n p و n(1-p) أكبر من 5 .

مثال (5-3)

لنفترض أنه في مجتمع العاملين بمجال الإدارة الحكومية هناك نسبة 8 % من العـــاملين يعانون من ظواهر إدارية مرضية قاتلة (تغليب الأنا على الصالح العام ، عدم الالتـــزام والانضباط ، عدم التقيد بالأنظمة والقوانين الإدارية المطبقة) ، فإذا اخترنـــا بشـــكل عشوائي 150 شخصاً من هذا المجتمع، فما هو احتمال أن تكون نسبة المقصرين وعدم الملتزمين في العينة أكبر أو تساوي 15 % ؟

الحل : إن حجم العينة المسحوبة كبير ، حيث أن المعيار هو أن يكون :

$n(1-p) > 5$, $n p > 5$ محقق .

$$n p = 150 (0.08) = 12$$

$$n (1-p) = 150 (0.92) = 138$$

يمكننا القول بأن $\hat{p}$ موزعة طبيعياً بشكل تقريبي بمتوسط يساوي $\mu_{\hat{p}} = P = 0.08$

وتباين يساوي :

$$\sigma^2_{\hat{p}} = P(1-P)/n = 0.08(0.92)/150 = 0.00049$$

نبحث عن احتمال $\hat{p}$ تحت المساحة المحصورة بين منحنى توزيع المعاينة للنسبة $\hat{p}$ وإلى يمين النسبة 15 % وهذه المساحة تساوي المساحة المحصورة تحت المـــنحنى الطبيعـــي المعياري إلى يمين القيمة المعيارية:

$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} = \frac{0.15 - 0.08}{\sqrt{0.00049}} = 3.15$$

ولحساب احتمال أن تكون نسبة المقصرين وعدم الملتزمين في العينة أكبر أو تساوي 15% نكتب:

$$P(\hat{p} \geq 0.15) = P(z \geq 3.15)$$
$$= 1 - P(z \leq 3.15) = 0.0008$$

ومن جدول التوزيع الطبيعي المعياري نجد أن المساحة الواقعة إلى يمين القيمة المعيارية $z = 3.15$ تساوي 0.0008 .

مثال (3–6) :

لنفترض أنه في مجتمع إحصائي يمثل الناجحين في المسابقة العامة لانتقاء أساتذة في الاقتصاد للمرحلة الثانوية، أن 90 % ممن يقع عليهم الاختيار يخضعون لدورة تدريبية مكثفة قبل الالتحاق بالمدارس. اخترنا عشوائيا 200 اقتصادي نجح في المسابقة (ينتمي للمجتمع الإحصائي المدروس) ، فما هو احتمال أن تكون نسبة الاقتصاديين الذين خضعوا لتلك الدورة في العينة أقل من 85 % ؟

الحل :

نفترض أن توزيع المعاينة للنسبة $\hat{p}$ هو طبيعي بشكل تقريبي بمتوسط $P = 0.90$ أما تباين توزيع المعاينة للنسبة $\hat{p}$ فيساوي إلى :

$$P(1-P)/n = 0.9(0.1)/200 = 0.0045$$

نحسب القيمة المعيارية:

$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} = \frac{0.85 - 0.90}{\sqrt{0.0004}} = -2.36$$

ولحساب احتمال أن تكون نسبة الاقتصاديين الذين خضعوا لتلك الدورة في العينة أقل من 85 % نكتب[3] :

$$P(\hat{p} \le 0.85) = P(z \le -2.36)$$
$$= P(z \ge 2.36) = 1 - P(z \le 2.36) = 0.0091$$

وهذا يعني بأن احتمال أن تكون نسبة المهندسين الذين خضعوا لتلك الدورة في العينة أقل من 85 % هي ضعيفة ولا تصل حتى إلى 1 % ، أي أن غالبية الناجحين في المسابقة يخضعون لدورة تدريبية مكثفة قبل الالتحاق بالمدارس .

---

[3] حسب الخاصة الرياضية : $F(-X) = 1 - F(X)$

§4- توزيع المعاينة للفرق بين نسبتي عينتين

Sampling Distribution of the Difference between Two Sample Proportions

إذا تم سحب عينتين عشوائيتين مستقلتين حجمهما على الترتيب n_1 , n_2 من مجتمعين حيث تمثل P_1 , P_2 على الترتيب نسبة الخاصية المدروسة في المجتمعين فإن توزيع المعاينة للفرق بين نسبة عينتين $(\hat{p}_1 - \hat{p}_2)$ هو طبيعي تقريبا بمتوسط يساوي:

$$\mu_{\hat{p}_1-\hat{p}_2} = P_1 - P_2$$

وتباين يساوي إلى:

$$\sigma^2{}_{\hat{p}_1-\hat{p}_2} = P_1(1-P_1)/n_1 + P_2(1-P_2)/n_2$$

وذلك عندما تكون كل من n_1 , n_2 كبيرة.

وللإجابة عن السؤال الاحتمالي المتعلق بالفرق بين نسبة العينتين $(\hat{p}_1 - \hat{p}_2)$، نستخدم الصيغة التالية:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}} \qquad (3-4)$$

مثال (3-5) :

لنفترض أن نسبة الطلبة الذين يستعملون منشطات للذاكرة قبل الامتحانات في مجتمع إحصائي أول هي 50 %، بينما تكون هذه النسبة في مجتمع إحصائي ثان (يضم طلبة جديين يعتمدون على أنفسهم ولا يستخدمون أدوية منشطة) هي 33 % . فإذا اخترنا بشكل عشوائي عينة حجمها 100 شخص من كل مجتمع، فما هو احتمال أن يكون الفرق بين نسبة العينتين $(\hat{p}_1 - \hat{p}_2)$ أكبر أو تساوي 30 % ؟

الحل:

سنفترض أن توزيع المعاينة للفرق بين نسبة العينتين هو طبيعي تقريباً بمتوسط يساوي:

$$\mu_{\hat{p}_1-\hat{p}_2} = 0.50 - 0.33 = 0.17$$

وتباين يساوي:

$$\sigma^2_{\hat{p}_1-\hat{p}_2} = (0.33)(0.67)/100 + (0.5)(0.5)/100 = 0.004711$$

إن المساحة الموافقة للاحتمال المطلوب هي المساحة المحصورة تحت منحني توزيع الفرق $(\hat{p}_1 - \hat{p}_2)$ إلى يمين النسبة 0.30 ويقابلها المساحة المحصورة تحت المنحنى الطبيعي المعياري إلى يمين القيمة المعيارية z :

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}} = \frac{0.30 - 0.17}{\sqrt{0.004711}} = 1.89$$

ولحساب احتمال أن يكون الفرق بين نسبة العينتين $(\hat{p}_1 - \hat{p}_2)$ أكبر أو تساوي 30% يمكننا أن نكتب:

$$P((\hat{p}_1 - \hat{p}_2) \geq 0.30) = P(z \geq 1.89)$$
$$= 1 - P(z \leq 1.89) = 0.0294$$

ومن جدول التوزيع الطبيعي المعياري نجد أن المساحة الواقعة إلى يمين النقطة ١,٨٩ تساوي 0.0294

مثال (3-6) :

لنفترض في مجتمع معين من المراهقين ، أن هناك نسبة 10 % من الشباب يشكون السمنة ، وإذا كانت نفس النسبة من البنات في المجتمع هن بدينات أيضاً . فإذا اخترنا بشكل عشوائي عينة تضم 250 شاب و 150 فتاة ، فما هو احتمال أن يكون الفرق بين نسبة العينتين $(\hat{p}_1 - \hat{p}_2) \geq 0.06$ ؟

الحل:

سنفترض أن توزيع المعاينة للفرق بين نسبة العينتين $(\hat{p}_1 - \hat{p}_2)$ هو طبيعي تقريباً متوسط يساوي الصفر (لأن نسبة من يشكون السمنة هي نفسها في المجتمعين) أي:

$$\mu_{\hat{p}_1-\hat{p}_2} = 0.10 - 0.10 = 0.00$$

وتباين يساوي :

$$\sigma^2_{\hat{p}_1-\hat{p}_2} = (0.1)(0.9)/250 + (0.1)(0.9)/200 = 0.00081$$

إن المساحة الموافقة للاحتمال المطلوب هي المساحة المحصورة تحت المنحنى الطبيعي المعياري إلى يمين القيمة المعيارية z :

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}} = \frac{0.06 - 0}{\sqrt{0.00081}} = 2.11$$

ولحساب احتمال أن يكون الفرق $(\hat{p}_1 - \hat{p}_2)$ أكبر أو تساوي ٦% يمكننا أن نكتب:

$$\begin{aligned} P((\hat{p}_1 - \hat{p}_2) \geq 0.06) &= P(z \geq 2.11) \\ &= 1 - P(z \leq 2.11) = 0.0174 \end{aligned}$$

ومن جدول التوزيع الطبيعي المعياري نجد أن المساحة الواقعة إلى يمين النقطة 2.11 تساوي 0.0174.

تمارين غير محلولة

١) ليكن معلوماً أن الأجرة الساعية لخريجي كلية الاقتصاد تخصص تسويق تتبع بشكل تقريبي التوزيع الطبيعي ، بمتوسط 450 ل.س وانحراف معياري 50 ل.س على الترتيب . إذا اخترنا عينة عشوائية بحجم 16 خريجاً متخصصاً في التسويق من هذا المجتمع . المطلوب إيجاد احتمال أن يكون متوسط الأجر لهذه العينة أكبر من 425 ل.س ؟ محصوراً بين 425 و 475 ل.س ؟ أكبر من 480 ل.س ؟

٢) ليكن معلوماً بأن متوسط الفترة اللازمة لمجموعة من الأشخاص المعوقين جسمياً لإنجاز مهام خاصة هو 25 ثانية بانحراف معياري 5 ثانية ، مفترضين أن الزمن اللازم لإنجاز المهام يتبع التوزيع الطبيعي . سحبت عينة عشوائية من هذا المجتمع بحجم 25 ، ما هو احتمال أن يكون متوسط العينة 28 ثانية فأكثر ؟ 26 ثانية أو اقل ؟

٣) إذا كانت درجات مادة أنظمة التشغيل لمجموعة من طلاب السنة الثانية بكلية الهندسة المعلوماتية تتبع التوزيع الطبيعي تقريبا بمتوسط وانحراف معياري 51 ، 10 درجة على الترتيب . اختيرت عينة عشوائية بحجم 9 طلاب من هذا المجتمع . المطلوب إيجاد احتمال أن يكون متوسط الدرجات لهذه العينة: أكبر من 60 درجة؟ محصوراً بين 50 درجة و 60 درجة ؟ .

٤) في دراسة للنفقات السنوية التي تخصصها الأسرة لتعليم أبنائها تقانات المعلوماتية ، تم إجراء مسح ميداني على مجتمعين مختلفين تباينهما معلوم $\sigma_2^2=3250, \sigma_1^2=2800$ ، سحبت عينة عشوائية من كل مجتمع فكانت النتائج: $n_1=40, \bar{x}_1=346$ $n_2=35, \bar{x}_2=300$ ، فما هو احتمال وجود فرق بين متوسط العينتين $(\bar{X}_1-\bar{X}_2)$ أكبر أو يساوي من الفرق المشاهد ، إذا علمنا بعدم وجود فرق بين متوسط المجتمعين الحقيقيين ؟

الفصل الرابع

التقدير المجالي Interval Estimate

مقدمة :

غالباً ما يحتوي المجتمع الإحصائي على مؤشرات مجهولة مثل المتوسط، التباين والنسبة، ويرغب الباحثون في تقدير هذه المؤشرات من البيانات المأخوذة من العينات العشوائية وذلك عن طريق حساب بعض الإحصاءات التي تعتمد على بيانات العينة واعتبارها كتقديرات لهذه المؤشرات. كأن يُستخدم متوسط العينة $\bar{X}$ كتقدير لمتوسط المجتمع μ، ويُستخدم الانحراف المعياري للعينة S كتقدير للانحراف المعياري للمجتمع σ. ولتقدير مؤشرات المجتمع المجهولة نستخدم التقدير النقطي Point Estimate ومن ثم التقدير المجالي. التقدير المجالي لإحدى مؤشرات المجتمع المجهولة μ ، σ ، P هو عبارة عن مجال يُحدد بقيمتين، قيمة دنيا وقيمة عليا ، تُحسب هذه القيم من مشاهدات العينة العشوائية المأخوذة من المجتمع المدروس.

من المتوقع أن يحتوي هذا المجال على قيمة المؤشر المجهول للمجتمع باحتمال ثقة معين وليكن (١ - α) حيث α مستوى الدلالة أو مستوى المعنوية وعادةً تأخذ قيماً صغيرة مثل: 0.01 , 0.05 , 0.10 .

تعتمد دقة التقدير على طول المجال الذي يحوي قيمة المؤشر، حيث كلما كان طول المجال صغيراً كلما زادت دقة التقدير ، لذلك سمي بمجال الثقة للتقدير .

§-1 الأخطاء المعيارية للإحصاءات ومجالات الثقة:

يزداد اقتراب إحصاءات العينات من معالم المجتمع الأصل، كلما زاد عدد مفردات أو عناصر تلك العينات، حتى تنطبق تلك الإحصاءات على المعالم عندما يصبح عدد مفردات العينة مساوياً لعدد أفراد المجتمع الأصل، أي عندما تصبح العينة

أصلاً. وتزداد ثقتنا في إحصاءات العينات كلما كان انحرافها أو تذبذبها عن معالم المجتمع الأصل صغيراً، ويقاس هذا التذبذب أو الانحراف بأهم مقاييس التشتت وهو الانحراف المعياري للمتوسطات والإحصاءات الأخرى المحسوبة من العينات، والتي نطلق عليها كما رأينا الخطأ المعياري Standard Error، ونستطيع أن نحدد المدى الذي تقع فيه تلك الإحصاءات اعتماداً على تلك الأخطاء المعيارية التي يمكن حسابها لكل إحصاءة، وذلك لتحدد مدى ثقتنا فيها.

فالمدى الذي يمتد من 1 ± انحراف معياري يختلف عن المدى الذي يمتد من 2 ± انحراف معياري و يختلف أيضاً عن المدى الذي يمتد من 3 ± انحراف معياري .. وهكذا نستطيع أن نتابع في تحديد هذا المدى إلى المستوى الذي يقرر الثقة في تلك الإحصاءات، ويسمى بمجال الثقة .

وبمعنى آخر لا يمكننا مطلقاً معرفة معالم المجتمع الأصل من معرفة إحصاءات العينة مهما كان اختيار تلك العينة دقيقاً، ولكن يستطيع الباحث أن يضع حدوداً أو مجالاً للقيمة المتوقعة أو مجالاً لمعالم المجتمع الأصل، ويقرن هذه الحدود بنسبة إحصائية معينة تسمى نسبة الثقة .

فإذا أردنا نسبة ثقة 68.27 % فإن لدينا 31.73 % شك ، وإذا أردنا نسبة ثقة 95.45 % فإن لدينا 4.55 % شك ، وإذا أردنا نسبة ثقة 99.73 % فإن لدينا 0.27 % شك ، وإذا أردنا نسبة ثقة 99 % فإن لدينا 1 % شك ، وإذا أردنا نسبة ثقة 95 % فإن لدينا 5 % شك .

ومثلما رأينا أن التوزيع التكراري للمتوسط يميل إلى أن يكون طبيعياً، فإن المساحة الطبيعية المحصورة بين 1± انحراف معياري أسفل هذا التوزيع الطبيعي تساوي 68.27 % وبذلك تصبح المساحة الطبيعية الباقية أسفل المنحنى 31.73 % ، أي أن نسبة احتمال وجود متوسط المجتمع الأصل في هذا المجال إلى احتمال عدم وجوده في هذا المجال تعادل أكثر من الضعف ، وبالإمكان أن نرتفع بحدود ثقتنا من 68.27 % إلى

95 % ثقة وبالتالي 5 % نسبة شك، أو نرفع حدود ثقتنا إلى 99 % مقابل نسبة 1 % شك. ومن المعلوم أن المساحة أسفل المنحنى الطبيعي والممتدة من 1.96- انحراف معياري إلى +1.96 انحراف معياري تعادل تقريباً 95 % ، كما أن المساحة أسفل المنحنى الطبيعي والممتدة من 2.58- انحراف معياري إلى 2.58+ انحراف معياري تعادل تقريباً 99 % .

وبالنتيجة نستطيع أن نرتفع بحدود الثقة من 95 % ثقة و 5ه % نسبة شك إذا ضربنا الخطأ المعياري بالرقم 1.96 لأن المساحة المعيارية التي تمتد من 1.96- درجة معيارية إلى 1.96+ درجة معيارية تعادل تقريباً 95 % من المساحة الكلية للمنحنى الطبيعي المعياري ، وهكذا نرى أن المدى الذي يمتد من :

متوسط المجتمع ± (١,٩٦) الخطأ المعياري

يجعل نسبة ثقتنا 95 % في وجود المتوسط في هذا المدى ودرجة الشك 5 %، وبشكل عام يمكن القول إن :

- الإحصائية ± (١) الخطأ المعياري ، تجعلنا أمام نسبة ثقة 68.27% من وقوع الإحصائية في هذا المدى و 31.73 % شك .
- الإحصائية ± (١,٩٦) الخطأ المعياري ، تجعلنا أمام نسبة ثقة 95 % من وقوع الإحصائية في هذا المدى و 5 % شك .
- الإحصائية ± (٢,٥٨) الخطأ المعياري ، تجعلنا أمام نسبة ثقة 99 % من وقوع الإحصائية في هذا المدى و 1 % شك .

وعلى هذا الأساس يستطيع الباحث أن يتنبأ بالحدين الأعلى والأدنى اللذين تقع بينهما المعلمة الحقيقية للمجتمع الأصل .

فإذا توصل باحث إلى أن العمر 55 سنة هو متوسط العمر التقاعدي لعينة عشوائية من المتقاعدين العاملين في مجال الإدارة لتسويقية، فلا شك أنه توصل إلى إحدى القيم المحتملة لمتوسط أعمار المتقاعدين في المجتمع الأصل ومن المحتمل أن يكون المتوسط قيمة

أخرى تختلف عن ذلك، ومدى بعد أو قرب القيم الأخرى عن المتوسط الحقيقي للمجتمع يتوقف على مدى الثقة التي يود الباحث أن يلتزم بها، فإذا قبل الباحث بنسبة خطأ قدرها 1 % من الفرص المحتملة لجميع القيم التي يأخذها المتوسط ، فإن المدى الذي يحدده للمتوسط بناءً على ما تقدم يساوي :

± 55 2.58 (الخطأ المعياري)

وإذا قبل أن يتسامح بنسبة 5 % من الفرص المحتملة فإن المدى الذي يحدده للمتوسط بناءً على ما تقدم يكون:

± 55 1.96 (الخطأ المعياري)

وهكذا نجد أنه كلما قبل الباحث نسبة أقل من الخطأ في الفرص المحتملة الحدوث حدد مدى أكثر اتساعاً، ولقد توصلت الدراسات الإحصائية لأشكال عدة ومختلفة للأخطاء المعيارية تختلف باختلاف الإحصائية المحسوبة من بيانات العينة العشوائية.

1-1 الخطأ المعياري للمتوسط

لقد رأينا بأن الخطأ المعياري للمتوسط يقاس اعتماداً على الانحراف المعياري للمجتمع وفي حالة عدم معلومية الانحراف المعياري للمجتمع يستبدل بالانحراف المعياري للعينة المختارة، بالإضافة إلى حجم العينة وذلك بموجب العلاقة التالية:

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

حيث: σ هي الانحراف المعياري للمجتمع .

أو

$$\sigma_{\bar{x}} = \frac{S}{\sqrt{n}}$$

حيث: S هي الانحراف المعياري للعينة .

§2-2 مجال الثقة لمتوسط المجتمع Confidence Interval for a Population Mean

نقصد بأسلوب التقدير المجالي لمتوسط المجتمع μ، تحديد مجال من القيم يحتوي على الوسط الحسابي للمجتمع باحتمال ثقة معين (α - ١) حيث تشير α إلى مستوى الدلالة وتأخذ القيم 0.01 , 0.05 , 0.10 ، وتدل على احتمال ألا يحتوي مجال الثقة على الوسط الحسابي للمجتمع ، في حين (α - ١) تشير لدرجة أو معامل الثقة Confidence Coefficient وتأخذ القيم المناظرة 0.95 , 0.99 , 0.90 على الترتيب ، وتشير إلى درجة ثقتنا بأن يحتوي المجال الذي حصلنا عليه على الوسط الحسابي μ .

نستخدم $\bar{X}$ متوسط العينة المسحوبة من المجتمع كتقدير نقطي لمتوسط المجتمع μ وبالتالي فإن خطأ التقدير يقاس ببعد $\bar{X}$ عن μ ونعبر عنه بالمقدار الموجب ε . وبناء عليه يمكن تحديد مجال ثقة للمؤشر المجهول μ باحتمال لا يقل عن (α - ١) بموجب العلاقة :

$$P(\bar{X} - \varepsilon \le \mu \le \bar{X} + \varepsilon) \ge 1 - \alpha \qquad (4-1)$$

واستنادا لنظرية النهاية المركزية يقترب توزيع المعاينة للمتوسط من التوزيع الطبيعي، ويمكن التأكيد باحتمال قدره (α - ١) بأن متوسط العينة $\bar{X}$ يختلف عن متوسط المجتمع μ بمقدار يقل عن $Z_{1-\alpha/2}$ مضروباً بالخطأ المعياري $\sigma_{\bar{X}}$ ويعبر عن ذلك بالعلاقة :

$$P(|\bar{X} - \mu| \le Z_{(1-\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}}) = 1 - \alpha \qquad (4-2)$$

وذلك في حالة كون حجم المجتمع الإحصائي كبيراً أو لانهائياً، ويمكننا أن نكتب باحتمال (α - 1) أن:

$$P\left[\bar{X} - Z_{(1-\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z_{(1-\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha \qquad (4-3)$$

أي أنه باحتمال قدره (α - 1) تقع μ داخل المجال:

$$\left[\bar{X} - Z_{(1-\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}} \;,\; \bar{X} + Z_{(1-\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

ويدعى ذلك بتقدير مجال الثقة عند مستوى معنوية α أو بدرجة ثقة $100(1-\alpha)\%$، كما يسمى المقدار $\left|\bar{X}-\mu\right|$ بخطأ التقدير.

1-2 توزيع t The t Distribution

تواجه الباحث الإحصائي مشكلة في بناء مجال الثقة عندما يكون كل من متوسط المجتمع وتباينه مجهولين ، وعلى الرغم من أن الإحصائية :

$$z = (\bar{x} - \mu) / \sigma / \sqrt{n}$$

تتبع توزيعاً طبيعياً عندما يكون المجتمع طبيعياً، أو تتبع على الأقل بشكل تقريبي التوزيع الطبيعي عندما يكون حجم العينة كبيراً بغض النظر عن شكل التوزيع الاحتمالي للمجتمع، إلا أننا لا نستطيع استخدام هذه الحقيقة بسبب كون الانحراف المعياري σ مجهولاً. والحل الأكثر منطقية هو استخدام الانحراف المعياري للعينة:

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

كتقريب للانحراف المعياري للمجتمع الأصل σ.

عندما يكون حجم العينة أكبر من 30 مفردة، يستخدم التوزيع الطبيعي في بناء مجال الثقة لمتوسط المجتمع. ولكن المشكلة هي في العينات الصغيرة، عندما يكون الانحراف المعياري للمجتمع مجهولاً حيث يستعاض عنه بالانحراف المعياري s المقدر من العينة ، إن توزيع المعاينة بهذه الحالة يقترب من توزيع آخر متماثل ، مختلف عن التوزيع الطبيعي ، وهو توزيع t لستودنت، اكتشفه العالم الألماني جوسيه Gosset

في أوائل القرن العشرين ونشره تحت اسم مستعار (توزيـــع t لطالـــب Student's distribution) ، ومن المعروف أن الإحصائية :

$$t = \frac{(\bar{x} - \mu)}{s / \sqrt{n}}$$

تتبع توزيع ستودنت الذي يتمتع بالخصائص التالية:

- متوسطه يساوي الصفر، متناظر حول الوسط، بشكل عام تباينه أكبر من الواحد ، إلا أن التباين يقترب من الواحد عندما يصبح حجم العينة كبيراً ، من أجل درجات حرية $df > 2$ فإن تباين هذا التوزيع t يساوي إلى $df/(df - 2)$ حيث ترمـــز df إلى درجات الحرية Degrees of freedom وتساوي $df = \nu = n - 1$.

يكتب مجال الثقة باحتمال ($1 - \alpha$) لمتوسط المجتمع μ في حالة العينات الصغيرة كما يلي :

$$\left[\bar{X} - t_{(1-\alpha/2)} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{(1-\alpha/2)} \cdot \frac{s}{\sqrt{n}}\right] \qquad (4-4)$$

حيث استبدلنا $Z_{1-\alpha/2}$ في العلاقة (٣-٤) بالقيمة $t_{1-\alpha/2}$ واستبدلنا الانحراف المعيـــاري σ للمجتمع بالانحراف المعياري للعينة S ، وتعطى $t_{\alpha/2}$ من جداول t حيث تعطى قيم t لكل درجة من درجات الحرية $\nu = n - 1$ وذلك لبعض قيم α .

ولبناء مجالات الثقة نستخدم دوما العلاقة :

التقدير النقطي ∓ (معامل الموثوقية) x (الخطأ المعياري)

أي أن :

$$\bar{X} \pm t_{(1-\alpha/2)} \cdot \frac{s}{\sqrt{n}}$$

والاختلاف الوحيد هو في مصدر معامل الموثوقية الذي نحصل عليه الآن من جـــدول توزيع t لستودنت.

نتيجة : عندما تكون المعاينة من توزيع طبيعي انحرافه المعياري σ غير معلوم ، فـإن مجال ثقة بدرجة $100(1-\alpha)\%$ لمتوسط المجتمع يعطى بالعلاقة :

$$\bar{X} \pm t_{(1-\alpha/2)} \cdot \frac{s}{\sqrt{n}}$$

نلاحظ أن هناك متطلباً يجب تحققه لنتمكن من استخدام توزيع t وهـو أن تكـون العينة مسحوبة من مجتمع طبيعي .

مثال (4-1)

باحث اقتصادي مهتم بالحصول على تقدير متوسط استهلاك الفرد السـنوي من المعجنات بكافة أشكالها (المعكرونة ، السباجيتي ، الشعيرية) ، سـحبت عينـة عشوائية مؤلفة من 10 أشخاص يمثلون ربات المنازل في مدينة حلب ، لمعرفة مسـتوى الاستهلاك لدى كل منهم فكان متوسط الاستهلاك لدى أفراد العينة بالكغ يسـاوي $\bar{X} = 22$ ، إذا كان المجتمع الإحصائي يتبع تقريباً التوزيع الطبيعي بتباين مقـداره 45 كغ ، أوجد باحتمال 95 % تقدير مجال الثقة للمتوسط الحقيقي للمجتمع μ .

الحل :

إن تقدير مجال الثقة المطلوب يحسب بموجب العلاقة التالية :

$$\bar{X} - Z_{(1-\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{(1-\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow 22 - 1.96\,\frac{\sqrt{45}}{\sqrt{10}} \leq \mu \leq 22 + 1.96\,\frac{\sqrt{45}}{\sqrt{10}}$$

$$22 - 1.96(2.12132) \leq \mu \leq 22 + 1.96(2.12132)$$

$$\Leftrightarrow \quad 17.842 \leq \mu \leq 26.158$$

أي باحتمال قدره 95 % يقع تقدير متوسط الاستهلاك مـن المعجنـات في المجتمـع المدروس بين حد أدنى قدره 17.842 كغ وحد أعلى قدره 26.158 كغ .

مثال (4–2) :

أعطي اختبار قدرات لعينة عشوائية تضم 36 عاملاً من العاملين الجدد بإحدى الشركات الوطنية فكان متوسط الدرجات هو 108 درجة (العلامة من 150 درجة) ، فإذا كان معلوماً أن تباين درجات اختبار القدرات يساوي 324٣٢٤ درجة .

المطلوب:

أوجد بدرجة ثقة 95 % تقدير مجال الثقة لمتوسط درجات اختبار القدرات μ .

الحل :

إن تقدير مجال الثقة المطلوب يحسب بموجب العلاقة التالية :

$$\overline{X} - Z_{(1-\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{(1-\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow 108 - 1.96\frac{\sqrt{324}}{\sqrt{36}} \leq \mu \leq 108 + 1.96\frac{\sqrt{324}}{\sqrt{36}}$$

وبالحساب نحصل على مجال الثقة المطلوب كما يلي :

$$102.12 \leq \mu \leq 113.88$$

أي سنكون واثقين باحتمال قدره 95 % بأن متوسط اختبار القدرات في مجتمع العاملين الجدد سيتراوح بين حد أدنى قدره 102.12 درجة وحد أعلى قدره 113.88 درجة .

مثال (4–3) :

يرغب معالج فيزيائي في تقدير متوسط القوة العظمى لعضلة معينة في جسم مجموعة من الرياضيين لاعبي رفع الأثقال من قسم الإحصاء والبرمجة بكلية الاقتصاد، مفترضين أن معدلات قوة العضلة تتبع تقريباً توزيعاً طبيعياً تباينه معلوم ويساوي 144 ، تم سحب عينة عشوائية مؤلفة من 15 رياضياً فكانت $\overline{X} = 84.3$.

والمطلوب أوجد باحتمال 99 % تقدير مجال الثقة للمتوسط μ .

الحل :

إن تقدير مجال الثقة المطلوب يحسب بموجب العلاقة التالية:

$$\overline{X} - Z_{(1-\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{(1-\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow 84.3 - 2.58 \frac{\sqrt{144}}{\sqrt{15}} \leq \mu \leq 84.3 + 2.58 \frac{\sqrt{144}}{\sqrt{15}}$$

وبالحساب نحصل على مجال الثقة المطلوب كما يلي:

$$76.3 \leq \mu \leq 92.3$$

أي باحتمال قدره 99 % يقع تقدير متوسط القوة العظمى لهذه العضلة في المجتمع المدروس بين حد أدنى قدره 76.3 وحد أعلى قدره 92.3 .

ملاحظة :

نلاحظ في كل من المثالين السابقين أن مجال الثقة كبير نوعاً ما، والسبب في ذلك يعود إلى صغر حجم العينة، حيث تزداد دقة التقدير كلما كبر حجم العينة.

مثال (4−4):

نرغب في تقدير متوسط درجات الذكاء لطلاب المرحلة ما قبل الابتدائية في المدارس التابعة لوزارة الشؤون الاجتماعية والعمل في سورية، لذلك سحبت عينة عشوائية من الطلبة عددها 15 طالباً، تم إخضاعهم لاختبارات تخصصية بإشراف بعض المشرفين الاجتماعيين المتخصصين فكانت النتائج كالتالي: متوسط درجات الذكاء لأفراد العينة 96 درجة وانحرافها المعياري 35 درجة، مع العلم أن تباين المجتمع الإحصائي غير معلوم ؟

الحل:

بما أن الانحراف المعياري للمجتمع σ مجهول، نفترض بأن درجات الذكاء لطلاب المرحلة الابتدائية (المجتمع الإحصائي) تتبع تقريباً توزيعاً طبيعياً.

إن تقدير مجال الثقة المطلوب يحسب بموجب العلاقة التالية:

$$\overline{X} - t_{(1-\alpha/2)} \cdot \frac{s}{\sqrt{n}} \le \mu \le \overline{X} + t_{(1-\alpha/2)} \cdot \frac{s}{\sqrt{n}}$$

$$\Leftrightarrow 96 - 2.1448 \ \frac{35}{\sqrt{15}} \le \mu \le 96 + 2.1448 \ \frac{35}{\sqrt{15}}$$

وبالحساب نحصل على مجال الثقة المطلوب كما يلي:

$$76.6 \le \mu \le 115.4$$

إن معامل الموثوقية يساوي بهذه الحالة 2.1448 ويمثل قيمة t الحرجة المرتبطة بمجال ثقة 95 % ، ودرجات حرية تساوي :

$$\nu = n - 1 = 14$$

وهذه القيمة نجدها في جدول توزيع t والعمود المعنون $t_{0.975}$ في السطر 14 .

أي باحتمال قدره 95 % يقع متوسط درجات الذكاء لطلاب المرحلة الابتدائية بين حد أدنى قدره 76.6درجة وحد أعلى قدره 115.4.

§-3 مجال الثقة للفرق بين متوسطي مجتمعين

Confidence Interval for the Difference between Two Population Means

كثيرا ما تواجه الباحثين مواقف تتطلب منهم إيجاد مجال ثقة للفرق بين متوسطين مثل الفرق بين متوسط كمية النيكوتين في نوعين من السجائر، أو الفرق بين متوسط إنتاجية محصول القطن في قطعتي أرض متماثلتين تختلفان فقط بنوعية البذور المزروعة، وكذلك الفرق بين متوسط درجات سمة العصابية لمجموعتين متماثلتين من الأطفال ولكن تختلفان فقط بانتمائهم الاجتماعي، ولذلك يجب سحب عينة عشوائية مستقلة من كل مجتمع ومن بياناتهما تحسب متوسطات العينتين $\overline{x}_2 , \overline{x}_1$ على الترتيب، وكما نعلم بأن المقدَر $(\overline{x}_1 - \overline{x}_2)$ يعتبر تقديراً غير متحيز[1] للفرق بين متوسط المجتمعين $(\mu_1 - \mu_2)$، كما أن توزيع المعاينة للفرق بين متوسط العينتين هو طبيعي تقريباً بمتوسط : $\mu_1 - \mu_2$.

---

[1] نقول إن $(\overline{x}_1 - \overline{x}_2)$ هو تقدير غير متحيز لـ $(\mu_1 - \mu_2)$ إذا تحقق الشرط $E(\overline{x}_1 - \overline{x}_2) = \mu_1 - \mu_2$

وتباين:

$$(\sigma_1^2 / n_1) + (\sigma_2^2 / n_2)$$

إن مجال الثقة للفرق بين متوسط المجتمعين $(\mu_1 - \mu_2)$ عندما يكون تباين المجتمع σ^2 معلوماً، وبدرجة ثقة $100(1-\alpha)\%$ يعطى بموجب العلاقة التالية:

$$(\bar{x}_1 - \bar{x}_2) - z_{(1-\alpha/2)} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + z_{(1-\alpha/2)} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad (4-5)$$

التقدير النقطي $\mp$ (معامل الموثوقية) x (الخطأ المعياري)

حيث تمثل:

$(\bar{x}_1 - \bar{x}_2)$: التقدير النقطي للفرق بين متوسط المجتمعين.

$\sqrt{(\sigma_1^2 / n_1) + (\sigma_2^2 / n_2)}$: الخطأ المعياري.

$Z_{(1-\alpha/2)}$: معامل الموثوقية

لنوضح من خلال بعض الأمثلة الحالة التي تكون فيها المعاينة من توزيع طبيعي وكذلك الحالة التي تكون فيها فرضية توزيع المجتمع طبيعياً غير محققة.

مثال (4-5):

لمعرفة الفرق بين متوسط مقاومة الشد لسلكين يتم إنتاجهما من معملين مختلفين A و B ، سحبت عينة عشوائية من الأسلاك من المعمل A حجمها 12 سلك فكان المتوسط $\bar{x}_1 = 4.5\ kg/mm^2$ ، ومن المعمل B سحبت عينة عشوائية تضم 15 سلكاً فكان متوسط العينة $\bar{x}_2 = 3.4\ kg/mm^2$ ، نفترض بأن القيم في المجتمعين موزعة طبيعياً بتباين مساو إلى الواحد .

والمطلوب:

أوجد مجال ثقة بدرجة 95 % للفرق بين متوسط المجتمعين $(\mu_1 - \mu_2)$ ؟ .

الحل:

كتقدير نقطي للفرق بين متوسط المجتمعين $(\mu_1 - \mu_2)$ نستخدم:

$$(\bar{x}_1 - \bar{x}_2) = 4.5 - 3.4 = 1.1$$

– معامل الموثوقية الموافق لدرجة ثقة 95 % يساوي 1.96 ، الخطأ المعياري يساوي :

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^{\ 2}}{n_2}} = \sqrt{\frac{1}{12} + \frac{1}{15}} = 0.39$$

إن مجال الثقة المطلوب يمكن حسابه بوساطة العلاقة:

$$1.1 - 1.96.(0.39) \le \mu \le 1.1 - 1.96.(0.39)$$
$$\Leftrightarrow 0.3 \le \mu \le 1.9$$

أي أننا واثقون ٩٥ % بأن الفرق الحقيقي بين $(\mu_1 - \mu_2)$ سيقع بين 0.3 و 1.9 .

3–1 المعاينة من مجتمعات غير طبيعية

Sampling from nonnormal Population

لبناء مجال ثقة للفرق بين متوسط مجتمعين عندما تكون المعاينة من مجتمعات غير طبيعية ، نطبق نظرية النهاية المركزية إذا كان مجموع حجم العينتين n_1 , n_2 أكبر من 30 .

مثال (4–6):

لمقارنة الحالة الاقتصادية والاجتماعية لمجموعتين من الأسـر (الأولى تنتمي لطبقة اجتماعية متوسطة من ذوي المهن الحرة، والثانية تنتمي لطبقة اجتماعية مختلفة وبسيطة من ذوي الدخل المحدود يقضون عطلة الصيف في منتجعات سياحية، سحبت عينة عشوائية تضم 75 شخصاً من المجموعة الأولى فكان متوسط دخل الأسرة لأفراد هذه العينة $\bar{x}_1 = 68000$ ل .س ، في حين كان متوسط دخل الأسرة سنوياً لأفراد العينة المسحوبة من المجموعة الثانية وعددها 80 شخصاً هو $\bar{x}_2 = 44500$ ل.س وكان

الانحراف المعياري للمجتمعين هو $\sigma_1 = 6000 , \sigma_2 = 5000$ ل.س .

المطلوب:

إيجاد مجال ثقة باحتمال 99 % للفرق بين المتوسطين ؟

الحل :

كتقدير نقطي للفرق بين متوسط المجتمعين $(\mu_1 - \mu_2)$ نستخدم:

$$(\bar{x}_1 - \bar{x}_2) = 68000 - 44500 = 23500$$

- معامل الموثوقية الموافق لدرجة ثقة 99 % نجده من جدول التوزيع الطبيعي المعياري مساوياً إلى 2.58.

- أما الخطأ المعياري للفرق بين متوسط مجتمعين يساوي:

$$\sigma_{\bar{x}-\bar{x}_2} = \sqrt{\frac{(6000)^2}{75} + \frac{(5000)^2}{80}} = 890$$

إن مجال الثقة المطلوب يمكن حسابه بوساطة العلاقة:

التقدير النقطي ∓(معامل الموثوقية) x (الخطأ المعياري)

$$23500 - 2.58(890) \le \mu_1 - \mu_2 \le 23500 + 2.58(890)$$

$$21204 \le \mu_1 - \mu_2 \le 25796$$

2–3 توزيع t والفرق بين متوسطين

The t Distribution and the Difference between Means

عندما يكون تباين المجتمعين مجهولاً ونرغب في تقدير الفرق بين متوسط المجتمعين ، نستخدم توزيع t كمصدر لعامل الموثوقية إذا تحققت افتراضات معينة ، مفترضين أن توزيع المجتمعين طبيعي، وفيما يتعلق بتباين المجتمعين سنميز بين :

1- حالة تساوي التباين في المجتمعين $\sigma_1^2 = \sigma_2^2$ Population Variances Equal

إذا كانت فرضية تساوي تباين المجتمعين محققة، فإن تباين العينتين المحسوبتين من هذين المجتمعين المستقلين يعتبر كتقدير مجمع Pooled estimate للتباين المشترك

The Common Variance ، و نحصل عليه عن طريق حساب الوسط المرجح لتباين العينتين حيث يرجح تباين كل عينة بدرجات حريتها كما يلي:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \qquad (4-6)$$

إن الخطأ المعياري للتقدير يعطى بموجب العلاقة :

$$\sqrt{S_p^2 / n_1 + S_p^2 / n_2}$$

كما أن $100(1-\alpha)$ % مجال ثقة للفرق بين متوسط المجتمعين $(\mu_1 - \mu_2)$ ، يعطى بالعلاقة :

$$(\bar{x}_1 - \bar{x}_2) \pm t_{(1-\alpha/2)} \cdot \sqrt{\frac{S_{1p}^2}{n_1} + \frac{S_p^2}{n_2}} \qquad (4-7)$$

إن عدد درجات الحرية المستخدمة في تحديد قيمة t لبناء مجال الثقة يساوي $v = n_1 + n_2 - 2$.

مثال (4–7) :

لمقارنة مستوى الذكاء[2] في مجتمعين الأول طلابه يعيشون في المناطق الحضرية ويتابعون دراستهم في أفضل المدارس والمجتمع الثاني طلابه يعيشون في المناطق الريفية ويتابعون دراستهم في مدارس لا تتوفر فيه نفس الإمكانيات البشرية والمادية من حيث المستلزمات والتجهيزات التعليمية ، تم سحب عينتين عشوائيتين الأولى من المجتمع الأول تضم 15 شخصاً فكان متوسط العينة $\bar{x}_1 = 120$ درجة ، وانحرافها المعياري 35 درجة والثانية من المجتمع الثاني وتضم 22 طالباً فكان متوسط العينة $\bar{x}_2 = 96$ ، درجة ، وانحرافها المعياري 40 درجة ، أوجد مجال ثقة بدرجة 95 % للفرق بين المتوسطين $(\mu_1 - \mu_2)$ ، إذا علمت أن تبايني المجتمعين متساويان ومجهولان ؟

---

[2] تم التعبير عن مستوى الذكاء في هذا المثال بوساطة الدرجات التي يحصل عليها الطلبة في الاختبارات التحريرية

الحل :

المجتمعان يتبعان التوزيع الطبيعي ، تباينهيما متساويان ولكن مجهولان ، لتقدير التباين المشترك نستخدم العلاقة التالية :

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 1450$$

- نستخدم الفرق بين متوسط العينتين:

$$(\bar{x}_1 - \bar{x}_2) = 120 - 96 = 24$$

كتقدير نقطي للفرق بين متوسط المجتمعين $(\mu_1 - \mu_2)$.

- معامل الموثوقية الموافق لدرجة ثقة 95 % يمكن أن نجدها في جـــدول توزيـــع t لستودنت وذلك عند درجات حرية $v = n_1 + n_2 - 2 = 35$ ، مساوية إلى 2.0301

- الخطأ المعياري يساوي :

$$\sqrt{S_p^2 / n_1 + S_p^2 / n_2} = 12.75$$

إن مجال الثقة المطلوب يمكن حسابه بوساطة العلاقة:

التقدير النقطي ∓(معامل الموثوقية) x (الخطأ المعياري)

$$(\bar{x}_1 - \bar{x}_2) \pm t_{(1-\alpha/2)} \cdot \sqrt{\frac{S_{1p}^2}{n_1} + \frac{S_p^2}{n_2}}$$

$$24 \pm 2.0301(12.75) \Leftrightarrow -2 \leq (\mu_1 - \mu_2) \leq 50$$

§-4 تقدير فترة الثقة للنسبة P في المجتمع:

Confidence Interval for a Population Proportion

إذا كانت الإحصائية P هي نسبة الشباب الذين لا يتقيدون بأنظمة السير (لايضعون حزام الآمان، يتجاوزون الحدود المسموح بها في الســـرعة)، أو نســـبة الأشـــخاص

المعرضين للإصابة بمرض معين في مجتمع محدد ، أو نسبة الحصول على عبوة دوائية معيبة عند سحب عينة حجمها n عبوة من إنتاج إحدى الشركات الوطنية للأدوية ، أو نسبة الشباب الذين يخالفون النظام العام أو قانون السير أو نسبة الطلبة الذين لا يحترمون الأنظمة الامتحانية في جامعة حلب ؟

لتقدير هذه النسبة نتبع نفس الإجراءات المتبعة في تقدير متوسط المجتمع ، نسحب عينة من المجتمع المدروس نحسب $\hat{p}$ نسبة الخاصية المدروسة في العينة والتي ستستخدم كتقدير نقطي للنسبة P ، وبتطبيق الصيغة :

المقدَر ± عامل الموثوقية x (الخطأ المعياري)

نحصل على تقدير مجال الثقة للنسبة P .

وكما نعلم سابقاً عندما تكون كل من np و $n(1-p)$ أكبر من ٥ ، نفترض أن توزيع المعاينة للنسبة $\hat{p}$ يقترب من التوزيع الطبيعي ، وفي هذه الحالة فإن معامل الموثوقية z يتبع التوزيع الطبيعي المعياري .

أما الخطأ المعياري للنسبة فسيكون:

$$\sigma_{\hat{p}} = \sqrt{P(1-P)/n}$$

وبما أن P هو المؤشر المجهول الذي نحاول تقديره ، لذلك نستخدم $\hat{p}$ كتقدير له ولهذا السبب نحسب الخطأ المعياري بالعلاقة :

$$\sigma_{\hat{p}} = \sqrt{\hat{P}(1-\hat{P})/n}$$

كما أن 95 % مجال ثقة للنسبة P يعطى بالعلاقة التالية :

$$\hat{p} - z_{(1-\alpha/2)} \cdot \sqrt{\hat{p}(1-\hat{p})/n} \le P \le \hat{p} + z_{(1-\alpha/2)} \cdot \sqrt{\hat{p}(1-\hat{p})/n} \qquad (4-7)$$

مثال (4−8) :

أجري مسح إحصائي لدراسة الممارسات الصحية السنية لدى مجتمع الشباب في مدينة حلب ، ولدى مقابلة 300 شاب تم انتقاؤهم عشوائيا ، أجـــاب 123 منـــهم بـــأنهم يخضعون لفحص (يراجعون طبيب أسنان) دوري منتظم مرتين في السنة .

المطلوب:

بناء مجال ثقة باحتمال 95 % للنسبة P في المجتمع المدروس ؟

الحل :

إن التقدير النقطي لنسبة المجتمع هو :

$$\hat{p} = 123/300 = 0.41$$

وبما أن حجم العينة كبير، نستخدم التوزيع الطبيعي المعياري في بناء مجال الثقة ، عامل الموثوقية الموافق لدرجة ثقة 95٪، % يساوي **1.96** .

والخطأ المعياري المقدر يساوي:

$$\sigma_{\hat{p}} = \sqrt{\hat{P}(1-\hat{P})/n} = 0.028$$

ومجال الثقة المطلوب هو :

$$0.41 - 1.96(0.028) \le P \le 0.41 + 1.96(0.028)$$

$$0.36 \le P \le 0.46$$

مثال (4−9) :

لدراسة اختبار صلاحية إحدى الطرق الحديثة المترافقة مع تقانات المعلوماتية، سحبت عينة عشوائية تضم 400 طالب ، أبدى 136 طالباً منهم عدم شعورهم بالراحة مـــن التأقلم والاستفادة من هذه التقانات التي لا تتناسب مع ميولهم ورغبـــاتهم ومســـتوى تأهيلهم السابق .

المطلوب:

بناء مجال ثقة باحتمال 95 % للنسبة P في المجتمع ؟

الحل :

نستخدم القيمة: $\hat{p} = 136/400 = 0.34$ كتقدير نقطي لنسبة المجتمع .

وبما أن حجم العينة كبير، فإن عامل الموثوقية الموافق لدرجة ثقة 95% يساوي 1.96

والخطأ المعياري المقدر يساوي :

$$\sigma_{\hat{p}} = \sqrt{\hat{P}(1-\hat{P})/n} = 0.0237$$

ومجال الثقة المطلوب يحسب كما يلي:

$$0.34 - 1.96(0.0237) \leq P \leq 0.34 + 1.96(0.0237)$$

$$0.294 \leq P \leq 0.386$$

مثال (4–9)

بفرض أن الإدارة التسويقية في إحدى المنشآت الوطنية العاملة في مجال توزيع المنظفات ترغب في تقدير النسبة الحقيقية لربات البيوت اللاتي يفضلن استخدام نـــوع معين من مساحيق الغسيل . سحبت عينة عشوائية تضم 100. من ربات البيوت فتبين أن 64 منهن يفضلن استخدام هذا النوع .

المطلوب :

بناء مجال ثقة باحتمال 99 % للنسبة الحقيقية P لربات البيـــوت الـــلاتي يفضـــلن استخدام هذا النوع من مساحيق الغسيل ؟

الحل :

نستخدم القيمة : $\hat{p} = 64/100 = 0.64$ كتقدير نقطي لنسبة المجتمع ، كما أن نسبة ربات البيوت اللاتي لا يفضلن استخدام هذا النوع من مساحيق الغسيل تعادل

$$1 - \hat{p} = 1 - \frac{64}{100} = 0.36$$

وبما أن حجم العينة كبير، فإن عامل الموثوقية الموافق لدرجة ثقة 99 % يساوي 2.575

والخطأ المعياري المقدر يساوي :

$$\sigma_{\hat{p}} = \sqrt{\hat{P}(1-\hat{P})/n} = \sqrt{\frac{0.64(0.36)}{100}}$$

ومجال الثقة المطلوب يحسب كما يلي:

$$0.64 - 2.575\sqrt{\frac{0.64(0.36)}{100}} \leq P \leq 0.64 - 2.575\sqrt{\frac{0.64(0.36)}{100}}$$

$$0.5164 \leq P \leq 0.7636$$

أي أن الإدارة التسويقية ستكون واثقة باحتمال 0.99 بأن نسبة ربات البيوت اللاتي تستخدمن هذا المسحوق في المجتمع الأصل ستتراوح ما بين 51.6 % على الأقل إلى 76.4 % كحد أقصى .

5-§ تقدير مجال الثقة للفرق بين نسبتي مجتمعين :

Confidence Interval for the Difference between Two Population Proportions

إن الاهتمام بالفرق بين نسبتي مجتمعين تزداد ما دامت الحاجة للمقارنة في ازدياد مستمر ، فمثلاً المقارنة بين نسبة البطالة بين الرجال ونسبتها بين النساء في مجتمعين ، ونسبة فعالية أحد الأدوية مع نسبة فعالية دواء أخر مماثل ، ونسبة المعارضين لعملية الإجهاض بين الرجال ونسبتها بين النساء ، ونسبة المؤيدين لعملية الإصلاح الاقتصادي والإداري ومحاربة الفساد بين الموظفين الإداريين ونسبتها بين العمال ....

ومثلما رأينا في حالة مجال الثقة للفرق بين متوسط المجتمعين، يستخدم الفرق بين نسبة العينتين $(\hat{p}_1 - \hat{p}_2)$ كتقدير نقطي غير متحيز للفرق بين نسبة المجتمعين $P_1 - P_2$ ، عندما يكون حجم العينتين n_2 , n_1 كبيراً ولا تكون النسب P_2 , P_1 قريبة جداً من الصفر أو من الواحد ، نطبق نظرية النهاية المركزية ونستخدم التوزيع الطبيعي المعياري في بناء مجال الثقة المطلوب .

وبما أن النسب P_2 , P_1 في المجتمعين مجهولة فان الخطأ المعياري للتقدير يجب أن يقدَّر بالعلاقة التالية :

$$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\hat{P}_1(1-\hat{P}_1)/n_1 + \hat{P}_2(1-\hat{P}_2)/n_2} \qquad (4-8)$$

إن مجال الثقة للفرق بين نسبتي المجتمعين $P_1 - P_2$ باحتمال قدره 95 % يعطى بالعلاقة التالية :

$$(\hat{p}_1 - \hat{p}_2) \pm z_{(1-\alpha/2)}\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2} \qquad (4-9)$$

مثال (4-9) :

لمقارنة وجهة نظر كل من الرجال والنساء حول موضوع الإجهاض بين مؤيد ومعارض، تم توزيع استبيانات على 100 رجل و 100 امرأة تم اختيارهم عشوائياً وهما عبارة عن مجموعتين متساويتين في العدد، من المجموعة الأولى التي تمثل الرجال كان عدد المعارضين لعملية الإجهاض 78 رجلاً، وفي المجموعة الثانية التي تمثل النساء كان عدد المعارضات لعملية الإجهاض 90 امرأة .

المطلوب:

تقدير الفرق الحقيقي بين نسبتي المجتمعين (الذين يعارضون عملية الإجهاض من الرجال والنساء) باحتمال 95 % .

الحل :

نستخدم الفرق بين نسبتي العينتين:

$$(\hat{p}_1 - \hat{p}_2) = 0.78 - 0.90 = -0.12$$

كتقدير نقطي غير متحيز للفرق بين نسبتي المجتمعين .

إن مجال الثقة للفرق $P_1 - P_2$ يمكن حسابه بوساطة العلاقة :

التقدير النقطي ∓ (معامل الموثوقية) x (الخطأ المعياري)

$$-0.12 \pm 1.96(0.05) \Leftrightarrow -0.22 \le (P_1 - P_2) \le -0.02$$

ملاحظة : إن الإشارات السالبة تعكس فقط حقيقة كون أن الفرق هو $(\hat{p}_1 - \hat{p}_2)$ وليس $(\hat{p}_2 - \hat{p}_1)$ حيث كان من الممكن بناء مجال الثقة باستخدام الفرق $(\hat{p}_2 - \hat{p}_1)$.

6-§ تحديد حجم العينة لتقدير المتوسطات:

DETERMINATION OF SAMPLE SIZE FOR ESTIMATING MEANS

إن السؤال المطروح وباستمرار عند التخطيط لإجراء تجارب عشوائية أو القيام ببحوث اقتصادية ميدانية أو إجراء مسح إحصائي هو: كم يجب أن يكون حجم العينة ؟ لأن سحب عينة بحجم كبير بهدف الحصول على نتائج جيدة، يكون مكلفاً زمنياً ومادياً، كما أن العينات صغيرة الحجم تعطي عادة نتائج لا يمكن استخدامها بشكل عملي .

إن الهدف من التقدير الحالي هو الحصول على مجالات ضيقة بموثوقية عالية، حيث أن طول المجال يحدده حجم المقدار:

(معامل الموثوقية) x (الخطأ المعياري)

وعلى الرغم من أن طول المجال الكلي قد يكون ضعف هذا المقدار، من أجل خطأ معياري معطى، فإن زيادة الموثوقية تعني معامل موثوقية أكبر وهذا يعطي مجالاً أوسع في حال ثبات الخطأ المعياري، ومن جهة أخرى إذا ثبتنا معامل الموثوقية فإن الطريقة الوحيدة لتضييق مجال الثقة هو تصغير الخطأ المعياري، وبما أن الخطأ المعياري يساوي $\sigma/\sqrt{n}$ و σ ثابت ، فان الطريقة الوحيدة للحصول على خطأ معياري صغير هو أخذ عينة كبيرة الحجم .

ولكن إلى أي حد يجب أن تكون العينة كبيرة ؟ فإن ذلك يعتمد على حجم الانحراف المعياري للمجتمع σ وعلى درجة الموثوقية المطلوبة وعلى الطول المرغوب لمجال الثقة. لنفترض أننا نرغب في بناء مجال ثقة يمتد E وحدة على طرفي المقدَر (الخطأ الاعظمي للتقدير)، يمكننا أن نكتب:

E = (معامل الموثوقية) x (الخطأ المعياري)

فإذا كانت المعاينة مع الإعادة فإن:

$$E = Z_{(1-\alpha/2)}.\sigma/\sqrt{n} \qquad (4-10)$$

وبحل هذه المعادلة من أجل n نجد أن حجم العينة يساوي إلى:

$$n = \left(\frac{Z_{\alpha/2}.\sigma}{E}\right)^2 \qquad (4-11)$$

أما إذا كانت المعاينة بدون إعادة من مجتمع صغير منته فإن مقدار الخطأ الأعظمي للتقدير يصبح كما يلي:

$$E = z.\frac{\sigma}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}} \qquad (4-12)$$

وبحل هذه المعادلة من أجل n نجد أن حجم العينة يساوي إلى:

$$n = \frac{N.z^2.\sigma^2}{E^2(N-1)+z^2.\sigma^2} \qquad (4-13)$$

وفي حال تجاهل عامل تصحيح المجتمع المنتهي The finite Population Correction ، $\sqrt{(N-n)/(N-1}$ ، فإن حجم العينة يحسب بنفس الطريقة سواء كانت المعاينة مع الإعادة أو بدون إعادة .

لحساب حجم العينة يلزمنا معلومات عن تباين المجتمع σ^2 الذي لا يكون معلوماً ويجب تقديره بواحدة من الطرق التالية:

- يمكن سحب عينة أولية من المجتمع وحساب تباينها واستخدامها كتقدير لتباين المجتمع σ^2 ، والمشاهدات المستخدمة في هذه العينة يجب أن تعتبر جزءاً من العينة النهائية .
- يمكن الاستفادة من الدراسات السابقة المماثلة والاعتماد على نتائجها في تقدير تباين المجتمع σ^2 .
- إذا كان المجتمع المعاين (الذي سيتم سحب العينة منه) يتبع تقريباً التوزيع الطبيعي ، يمكننا استخدام الحقيقة التالية : المدى العام يساوي تقريباً ٦ أمثال الانحراف المعياري أي : $\sigma \approx R/6$ وهذه الطريقة تتطلب معرفة أصغر قيمة وأكبر قيمة يأخذها المتغير المدروس في المجتمع .

مثال (4–10) :

يرغب قسم الإحصاء والبرمجة بكلية الاقتصاد / جامعة حلب بإجراء مسح على مجموعة من أرباب الأسر العاملين بجامعة حلب لتحديد متوسط كمية اللحوم اللازمة للأسرة يومياً، لهذا انحصر هدف الفريق الباحث في تحديد حجم العينة الواجب دراستها. وللوصول لذلك كان من الضروري تزويد الفريق الباحث بثلاثة معلومات أساسية وهي : طول مجال الثقة المرغوب ، درجة الثقة المرغوبة ، مقدار تباين المجتمع ؟

الحل :

لنفترض بأن رغبة قسم الإحصاء والبرمجة بكلية الاقتصاد هي أن يكون طول المجال 100 غرام بمعنى أن يقع التقدير على مقربة من القيمة الحقيقية بـ 50 غرام من كل طرف ، كما أن درجة الثقة المطلوبة هي 95 % ، ومن التجارب السابقة يشعر الفريق الباحث بأن الانحراف المعياري للمجتمع (الكمية المستهلكة من اللحوم) بحدود 200 غرام أي :

$$z = 1.96 \,,\ \sigma = 200 \,,\ E = 50$$

وبفرض أن المجتمع المدروس كبير جداً بشكل يمكن معه تجاهل معامل التصحيح، فإن حجم العينة المطلوبة سيكون :

$$n = \frac{(1.96)^2 (200)^2}{(50)^2} = 61.47 \approx 62 \quad \text{شخص}$$

§–7 تحديد حجم العينة لتقدير النسب

DETERMINATION OF SAMPLE SIZE FOR ESTIMATING PROPORTIONS

إن طريقة تحديد حجم العينة عند تقدير النسبة في المجتمع هي نفسها الطريقة الموصوفة من أجل تقدير متوسط المجتمع. فإذا كانت المعاينة مع الإعادة من مجتمع كبير جداً فإن حجم العينة يساوي إلى:

$$n = \frac{z^2 p.q}{E^2} \,,\ p = 1 - q \qquad (4-14)$$

أما في حالة المعاينة بدون إعادة من مجتمع صغير منته فان حجم العينة يساوي إلى:

$$n = \frac{N.z^2.p.q}{E^2(N-1) + z^2.p.q} \qquad (4-15)$$

عندما يكون حجم المجتمع كبيراً مقارنة بحجم العينة $n/N \leq 0.05$ يمكن تجاهل معامل التصحيح وتستخدم المعادلة الأولى بتحديد حجم العينة.

ولحساب حجم العينة يلزمنا معلومات عن النسبة P ، التي تكون غير معلومة والتي نقدرها بإحدى الطرق التالية :

- يمكن سحب عينة أولية من المجتمع وحساب $\hat{p}$ لاستخدامها كتقدير لـ P في الصيغة المستخدمة لتحديد حجم العينة .
- في بعض الأحيان قد يمتلك الباحث معلومات عن الحد الأعلى للنسبة المستخدمة بالصيغة، فمثلاً قد يكون من المرغوب تقدير النسبة في مجتمع تقيده بعض الشروط، كأن نشعر بأن النسبة الحقيقية يجب ألا تتجاوز 0.30 ، عندها نعوض في الصيغة المستخدمة لحساب حجم العينة $p = 0.30$.
- إذا كان من المتعذر إعطاء أفضل تقدير للنسبة p ، يمكننا وضع $p = 0.50$ في الصيغة المستخدمة لتحديد حجم العينة وهذه ستعطي قيمة أعظمية لحجم العينة ، تستخدم هذه الطريقة فقط في حالة عدم التمكن من إعطاء تقدير أفضل لـ p .

مثال (4-11) :

تخطط مديرية صحة حلب لإجراء مسح ميداني لتحديد نسبة العائلات التي تعاني من نقص وعوَز في الرعاية الطبية في منطقة معينة، تعتقد المديرية بأن النسبة لن تكون أكبر من 0.35 وفي الوقت نفسه ترغب بإنشاء مجال ثقة للنسبة باحتمال 95 % مع ارتكاب خطأ أعظمي E = 0.05 .

والمطلوب:

ما هو حجم العينة الواجب اختيارها من العائلات ؟

الحل :

إذا تجاهلنا معامل تصحيح المجتمع المنتهي ، سيحسب حجم العينة المطلوبة بموجب العلاقة التالية:

$$n = \frac{z^2 p.q}{E^2} = \frac{(1.96)^2 (0.35)^2 (0.65)^2}{(0.05)^2} = 349.6 \approx 350$$

§-8 تقدير مجال الثقة لتباين مجتمع موزع طبيعياً :

Confidence Interval for the Variance of a Normally Distributed Population

إن بناء مجال ثقة لتباين المجتمع σ^2 يعتمد غالباً على توزيع المعاينة لـ $(n-1).s^2/\sigma^2$ ، فإذا سحبنا عينات بحجم n من مجتمع طبيعي فإن المقدار :

$$(n-1).s^2/\sigma^2$$

يتبع توزيع كاي تربيع Chi-square distribution مع n − 1 درجة حرية ، وللحصول على مجال ثقة للتباين σ^2 نستخدم العلاقة التالية :

$$\frac{(n-1).s^2}{\chi^2_{1-\alpha/2}} \le \sigma^2 \le \frac{(n-1).s^2}{\chi^2_{\alpha/2}} \qquad (4-16)$$

ولحساب ذلك نختار قيمتين:

$$\chi^2_{\alpha/2} \text{ ، } \chi^2_{1-\alpha/2}$$

من جدول كاي تربيع بطريقة يتم فيها تقسيم α بالتساوي بين طرفي التوزيع .

مثال (4−12) :

للتعرف على مستوى درجات الذكاء لأطفال الحضانة التابعة للاتحاد العام النسائي، سحبت عينة عشوائية تضم 15 طفلاً فكان متوسط درجات الذكاء لدى أفراد العينة مساوياً إلى 96 درجة، وكان الانحراف المعياري 35 درجة، فإذا علمت أن درجات الذكاء لأطفال الحضانة التابعة للاتحاد العام النسائي تتبع توزيعاً طبيعياً.

المطلوب:

- أوجد مجال ثقة باحتمال 95 % لتباين المجتمع σ^2 ؟
- أوجد مجال ثقة باحتمال 95 % للانحراف المعياري للمجتمع ؟

الحل:

كما نعلم بأن تباين العينة هو $s^2 = 1225$.

ودرجات الحرية هي : $n - 1 = 14$.

من جدول توزيع كاي تربيع (انظر الملحق رقم 1) ، يمكن أن نستخرج قيمتين لكاي تربيع الأولى صغرى والثانية كبرى وهما :

$$\chi^2_{\alpha/2} = 5.629$$
$$\chi^2_{1-\alpha/2} = 26.119$$

ومجال الثقة المطلوب هو:

$$\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{\alpha/2}}$$

$$\Leftrightarrow \frac{14(1225)}{26.119} \le \sigma^2 \le \frac{14(1225)}{5.629} \Leftrightarrow 656.6 \le \sigma^2 \le 3046.7$$

وبأخذ جذر الطرفين نحصل على مجال الثقة المطلوب للانحراف المعياري كما يلي:

$$25.62 \le \sigma \le 55.20$$

ملاحظة:

إن فرضية كون المجتمع المدروس يتبع التوزيع الطبيعي أساسية وبدونها ستكون النتيجة مضللة، إن المقدّر لا يقع في منتصف المجال كما كان الحال مع المتوسط والنسبة والسبب أن توزيع كاي تربيع لا يشبه التوزيع الطبيعي وليس متناظرا ، مما يترتب عليه أن الطريقة المستخدمة في بناء مجال الثقة لتباين المجتمع لا تعطي أضيق مجال ثقة ممكن.

9-§ تقدير مجال ثقة لنسبة تبايني مجتمعين موزعين طبيعيا :

Confidence Interval for the Ratio of the Variances of Two Normally Distributed Populations

في كثير من المواقف نحتاج إلى مقارنة تبايني مجتمعين، وإحدى طرق المقارنة هي تشكيل نسبة التباينين σ_1^2 / σ_2^2 فإذا كان التباينان متساويان فإن نسبتهم ستساوي الواحد، وغالبا ما يكون تباين المجتمعات المدروسة مجهولاً، وبالتالي فان أية مقارنة سوف تعتمد على تباين العينات العشوائية المسحوبة من تلك المجتمعات .

لتقدير نسبة تبايني مجتمعين لا بد من الاعتماد على توزيع للمعاينة، والتوزيع المستخدم في هذه الحالة هو توزيع المقدار:

$$(S_1^2 / \sigma_1^2)/(S_2^2 / \sigma_2^2)$$

الذي يتطلب استخدامه تحقق بعض الشروط مثل :

- أن يتم حساب S_1^2 , S_2^2 من عينات عشوائية مستقلة بأحجام n_1 , n_2 على الترتيب مسحوبة من مجتمعين موزعين طبيعيا.

وفي حال تحقق هذه الشروط فإن الاحصائية:

$$(S_1^2 / \sigma_1^2)/(S_2^2 / \sigma_2^2)$$

تتبع توزيع F لفيشر F distribution الذي يعتمد قيمتين لدرجات الحرية ، الأولى تتعلق بقيمة $n_1 - 1$ المستخدمة في حساب S_1^2 والثانية تتعلق بقيمة $n_2 - 1$ المستخدمة في حساب S_2^2 ، والتي يشار إليهما غالباً بدرجات الحرية للصورة numerator ودرجات الحرية للمخرج denominator.

إن جدول توزيع F (من اجل تراكيب محددة من درجات الحرية وقيم α) يتضمن قيم F حيث على يسارها يقع $\alpha/2$ من المساحة تحت منحنى توزيع F .

لإيجاد $100(1-\alpha)\%$ مجال ثقة للنسبة σ_1^2 / σ_2^2 نبدأ مع الصيغة:

$$F_{\alpha/2} < (S_1^2 / \sigma_1^2)/(S_2^2 / \sigma_2^2) < F_{(1-\alpha/2)}$$

حيث $F_{(1-\alpha/2)}$, $F_{\alpha/2}$ هي القيم المستخرجة من جدول F ويقع على يســـار الأولى ويمين الثانية $\alpha/2$ من مساحة المنحني.

وبإعادة ترتيب العلاقة السابقة نحصل على مجال الثقة المطلوب كما يلي:

$$\frac{S_1^2/S_2^2}{F_{(1-\alpha/2)}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2/S_2^2}{F_{\alpha/2}} \qquad (4-17)$$

مثال (4–13) :

بهدف دراسة زمن الاستجابة (ردة الفعل) لمنبه خاص، اختيرت عينتان عشـــوائيتان الأولى بحجم ٢١ شخص من مجتمع أصحاء جسمياً ، والثانية بحجم 16 شخصـــاً مـــن مجتمع من المرضى يعانون من مرض الشلل الارتعاشي التصلبي فكان تباين العينة الأولى مساويا إلى 1600 وتباين العينة الثانية 1225 .

المطلوب:

إيجاد مجال ثقة باحتمال 95 % لنسبة التباينين σ_1^2/σ_2^2 .

الحل:

لدينا تباين العينتين، وبالاعتماد على درجات الحرية ومستوى المعنوية نستخرج القـــيم الجدولية :

$$\nu_1 = 21-1 = 20 \quad , \quad \nu_2 = 16-1 = 15 \quad ,, S_1^2 = 1600, S_2^2 = 1225$$

$$\alpha = 0.05 \quad , F_{0.025} = 0.389, F_{0.975} = 2.76$$

كيفية الحصول على القيم الحرجة (الجدولية) :

إن القيمة الجدولية $F_{0.975} = 2.76$ نجدها عند تقاطع العمود المخصـــص لـــدرجات حرية البسط 20 مع السطر المخصص لـــدرجات حريـــة المقـــام 15 وفي الصـــفحة المخصصة لدرجة الثقة 0.975 .

ولإيجاد القيمة الجدولية $F_{0.025} = 0.389$ نستخدم العلاقة التالية :

$$F_{(1-\alpha,\nu_1,\nu_2)} = \frac{1}{F_{(\alpha,\nu_2,\nu_1)}}$$

نبدل بين مواقع درجات حرية البسط والمقام ونضع القيمة المناسبة لـــ F .

في مثالنا وفي الصفحة المخصصة لدرجة الثقة 0.975 نجد القيمة 2.57 عنـــد تقـــاطع العمود 15 والسطر 20 وبأخذ مقلوب هذه القيمة نجدها مساوية إلى :

$1 / 2.57 = 0.389$.

إن 95 % مجال ثقة لنسبة التباينين σ_1^2 / σ_2^2 يكون على النحو التالي :

$$S_1^2 / S_2^2 . \frac{1}{F_{\alpha/2,\nu_1,\nu_2}} < \frac{\sigma_1^2}{\sigma_2^2} < S_1^2 / S_2^2 . \frac{1}{1 / F_{1-\alpha/2,\nu_2,\nu_1}}$$

$$\frac{1600/1225}{2.76} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{1600/1225}{0>389} \Leftrightarrow 0.473 < \frac{\sigma_1^2}{\sigma_2^2} < 3.36$$

تمارين غير محلوله

١- أجريت تجربة على 15 شخصاً من رواد الفضاء، في مجال يحاكي مجال انعدام الوزن وجد أن متوسط زيادة ضربات القلب لديهم 25 ضربة / دقيقة ، بانحراف معياري قدره 5 ضربة / الدقيقة ، احسب أقصى خطأ في تقدير المتوسط عند مستوى معنوية $\alpha = 0.01$ ؟

٢- لتحديد نسبة ذوي الضغط العالي من بين الأشخاص البالغين ، كان المطلوب هو تحديد حجم العينة n حتى يمكننا التأكد باحتمال 0.99 من أن الخطأ لا يتجاوز 0.05 وذلك إذا علمنا أن قيمة النسبة P في المجتمع تساوي 0.20

٣- في دراسة لعينة مكونة من 150 مريضا متقدماً بالسن في أحد مراكز الرعاية الاجتماعية الخاصة لمعرفة رأيهم في سبب تفضيلهم لهذا المركز على غيره ، أجاب 108 منهم بأن سبب التفضيل هو الرعاية الصحية والمعاملة اللطيفة وقيمة الفاتورة ، احسب مقدار القيمة العظمى للخطأ في التقدير عند درجة ثقة 0.99 .

٤- ترغب إحدى الشركات في تقدير نسبة المستهلكين المتوقع أن يفضلوا شراء سلعة جديدة تنوي طرحها في الأسواق . تشعر الشركة أن موقفها سيكون طيباً إذا كانت نسبة المستهلكين اللذين يفضلون شراء المنتج الجديد هي 0.20 . ولهذا السبب طلبت إدارة الشركة من الإدارة التسويقية إجراء مسح ميداني على عينة من المستهلكين وموافاتها بحجم العينة المطلوب لتقدير نسبة المستهلكين بحيث لا يزيد خطأ التقدير عن 0.04 وذلك باحتمال قدره 99 % .

٥- في دراسة للنفقات السنوية التي تخصصها الأسرة للرعاية الصحية العامة ، تم إجراء مسح ميداني على مجتمعين مختلفين تباينهما معلوم $\sigma_2^2 = 3250, \sigma_1^2 = 2800$ ، سحبت عينة عشوائية من كل مجتمع فكانت النتائج التالية :

$n_2 = 35\,, \bar{x}_2 = 300$ $\quad$ $n_1 = 40\,, \bar{x}_1 = 346$ ، فما هو احتمال وجود فرق بـــين متوسطي العينتين $(\bar{X}_1 - \bar{X}_2)$ أكبر أو يساوي من الفرق المشاهد ، إذا علمنا بعـــدم وجود فرق بين متوسطي المجتمعين الحقيقيين ؟

٦- في دراسة لعينة مكونة من 150 مصطافاً في إحدى المنتجعات السياحية الخاصـــة لمعرفة رأيهم في سبب تفضيلهم لهذا المنتجع عن غيره ، أجاب عدد قدره 108 منـــهم بأن سبب التفضيل هو توفر المرافق الخدمية بمختلف أنواعها، والمعاملة اللطيفة وقيمـــة الفاتورة ، احسب مقدار القيمة العظمى للخطأ في التقدير عند درجة ثقة 0.99 .

الفصل الخامس

اختبار الفرضيات الإحصائية
Statistical Hypothesis Testing

مقدمة:

يعتبر اختبار الفرضيات الإحصائية أسلوباً لاتخاذ قرار بقبول أو رفض صياغة مبدئية حول إحدى مؤشرات المجتمع المجهولة (متوسط ، نسبة ، انحراف معياري) ، وذلك بناء على بيانات عينة عشوائية اختيرت من المجتمع نفسه (وذلك لعدم معرفتنا الكاملة بقيمة هذه المؤشرات أو طبيعتها) .

فمثلاً قد تختبر إدارة الكلية فيما إذا كان متوسط فترة التدريب لدورة متخصصة في المعلوماتية هي 15 يوماً أم لا، أو قد يختبر خبير اقتصادي فيما إذا كان من شأن برنامج إصلاح اقتصادي معين (إحداث هيئة مكافحة البطالة في سورية) تخفيض نسبة البطالة في سوريا إلى أقل من 10 % أم لا، أو قد تختبر مشرفة اجتماعية فيما إذا كان من شأن برنامج تعليمي خاص أن يؤدي إلى تحسين عملية الاتصال ما بين المعلمات والأطفال في دور الحضانة التابعة لنقابة المعلمين بجامعة حلب ؟ أو كأن يختبر طبيب في العيادة الاجتماعية فعالية طريقة جديدة تساعد المدخنين للإقلاع عنه ، أو أن يختبر معالج فيزيائي فيما إذا كانت طريقة معينة في العلاج ستؤدي إلى التأثير الإيجابي لأكثر من 90 % من الحالات أم لا ...الخ .

وللوصول إلى هذه القرارات الإحصائية نقوم عادة بوضع فروض عن مؤشرات المجتمع الإحصائي مثل المتوسط، النسبة ، والانحراف المعياري .... ثم نختبر هذه الفروض بناء على عينة عشوائية نختارها من المجتمع .

هذه الفروض ، قد تكون صحيحة أو غير صحيحة ، تسمى بالفروض الإحصـــائية وتنقسم إلى قسمين :

أ- فروض عن مؤشرات المجتمع Parametric .

- فروض عن شكل تابع التوزيع Non Parametric .

وسنقتصر هنا على دراسة الفروض عن مؤشرات المجتمع فقط.

§1- الخطوات الأساسية في اختبار الفرضيات :

أولاً : الفرضيات Hypothesis :

هناك نوعان من الفرضيات الإحصائية، يشار عـــادة إلى الفرضـــية الأولى الواجـــب اختبارها بفرضية العدم Null Hypothesis وتسمى كذلك لأن الغرض منـــها هـــو الشك فيها ونرمز لها بـــ H_0، ويسمى الافتراض الثاني الذي يختلف عن H_0 بالفرضية البديلة Alternative Hypothesis ويرمز لها بـــ H_1.

مثلاً إذا افترض أحد الباحثين الإحصائيين أن متوسط دخل مجموعة مـــن الأشـــخاص العاملين في مجال الإحصاء والبرمجة هو 7500 ليرة سورية شهرياً، واختار عينة عشوائية مكونة من 50 خريجاً مختصاً بالإحصاء والبرمجة، وحسب متوسط الـــدخل الشـــهري الذي يحصلون عليه فكانت النتيجة هي 6000 ليرة سورية شهرياً وكـــان الانحـــراف المعياري لهذا المجتمع الإحصائي معروفا ويساوي 3600 ليرة سورية شـــهرياً فتكـــون فرضية العدم بهذه الحالة:

$$H_0 : \mu = 7500$$

حيث نفترض عدم وجود فروق حقيقية بين متوسط المجتمـــع والقيمـــة المفروضـــة. والفروق المشاهدة إنما تعزى للصدفة، ويقابل الفرضية الأولية فرضية أخـــرى تســـمى بالفرضية البديلة ويرمز لها H_1 وفي مثالنا يمكن أن تأخذ إحدى الحالات التالية :

- متوسط دخل الأشخاص العاملين بمجال الإحصاء والبرمجة بكلية الاقتصاد لا يساوي

7500 ل.س شهرياً ؟

$$H_1 : \mu \neq 7500$$

- متوسط دخل الأشخاص العاملين في مجال الإحصاء والبرمجة بكلية الاقتصاد هو أكبر من 7500 ل .س ؟

$$H_1 : \mu > 7500$$

- متوسط دخل الأشخاص العاملين في مجال الإحصاء والبرمجة بكلية الاقتصاد هو أقل من 7500 ل .س ؟

$$H_1 : \mu < 7500$$

- عادة نضع ما نرغبه أو نتوقع أننا قادرون على أن نستنتجه من الاختبار كفرضية بديلة، كما أن طبيعة الاختبار والمسألة المعالجة تحددان نوعية الفرضية البديلة.
- الفرضية الأولية والفرضية البديلة تكملان بعضهما البعض بشكل تشملان كل الإمكانيات المتعلقة بالقيم التي يأخذها المؤشر موضع الاختبار.

ثانياً : حساب إحصائية الاختبار Calculation of the Test Statistic :

تحسب إحصائية الاختبار من مشاهدات العينة العشوائية ومعطيات فرضية العدم H_0 ، وتستخدم كوسيلة لاتخاذ قرار (كمتخذ قرار) Decision Maker ، طالما أن قرار رفض R eject أو عدم رفض H_0 يعتمد على قيمة إحصائية الاختبار ، حيث تقارن قيمة إحصائية الاختبار مع القيم الحرجة أو الجدولية المستخرجة من جداول معدة خصيصاً لكل توزيع احتمالي وهذه القيم تحدد بدورها منطقتي القبول والرفض الواقعتين تحت منحنى توزيع إحصائية الاختبار.

كمثال على إحصائية الاختبار نأخذ المقدار:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

حيث :

μ_0 : هي القيمة المختبرة لمتوسط المجتمع[1].

$\sigma / \sqrt{n}$: الانحراف المعياري للمجتمع.

$\sigma / \sqrt{n}$: هي الخطأ المعياري لمتوسط العينة.

ثالثاً: تحديد توزيع إحصائية الاختبار Distribution of the Test Statistic :

إن إحصائية الاختبار:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

تتبع التوزيع الطبيعي المعياري إذا كانت فرضية العدم H_0 صحيحة.

رابعاً: قاعدة القرار واتخاذ القرار الإحصائي:

إن كل القيم الممكنة لإحصائية الاختبار تتوضع على المحور الأفقي لمنحنى توزيع إحصائية الاختبار، وتنقسم هذه القيم إلى مجموعتين الأولى تشكل ما يسمى بمنطقة الرفض Rejection Region وهي التي يكون فيها احتمال حدوث قيم الإحصائية α صغيراً عندما تكون الفرضية H_0 صحيحة، والثانية تحدد منطقة القبول acceptance Region وهي التي يكون فيها احتمال حدوث قيم الإحصائية (١ − α) كبيراً عندما تكون الفرضية H_0 صحيحة. تمكننا قاعدة القرار Decision Rule من اتخاذ القرار برفض فرضية العدم إذا كانت قيمة إحصائية الاختبار المحسوبة من العينة العشوائية واقعة في منطقة الرفض أو بعدم رفض فرضية العدم إذا كانت قيمة إحصائية الاختبار المحسوبة من العينة العشوائية واقعة في منطقة القبول .

إن قرار وقوع قيم الإحصائية بمنطقة الرفض أو القبول يعتمد على مستوى المعنوية Level of Significance المرغوب[2] α، حيث α تحدد المساحة الواقعة تحت منحنى توزيع إحصائية الاختبار وتمثل تماماً احتمال رفض فرضية العدم الصحيحة ، وبما

---

[1] حيث تأخذ فرضية العدم الشكل التالي: $H_0 : \mu = \mu_0$

أن رفض فرضية العدم الصحيحة يشكل خطأ فمن المنطقي أن نحاول جعل هذا الاحتمال صغيراً، وفي الواقع العملي نفعل ذلك حيث نختار قيمة صغيرة جداً لمستوى المعنوية α بهدف تصغير احتمال رفض فرضية العدم الصحيحة، ومن القيم الشائعة لمستوى المعنوية α نذكر: 0.01 ، 0.05 ، 0.10 .

– الخطأ من النوع الأول والخطأ من النوع الثاني :

يسمى الخطأ المرتكب عند رفض فرضية العدم الصحيحة بخطأ من النوع الأول Type 1 error ، ويسمى الخطأ المرتكب عند قبول فرضية العدم الخاطئة بخطأ من النوع الثاني Type 2 error .

نرمز لاحتمال ارتكاب خطأ من النوع الثاني بـ γ، ولاحتمال ارتكاب خطأ من النوع الأول α. وذلك كما هو موضح بالجدول

البيـــــان	H_0 صحيحة	H_0 خاطئة
رفض الفرضية H_0	خطأ من النوع الأول α	قرار صحيح واحتماله $\beta = 1-\alpha$
عدم رفض الفرضية H_0	قرار صحيح واحتماله $1-\gamma$	خطأ من النوع الثاني γ

– الخلاصة Conclusion:

إذا تم رفض العدم H_0 نستنتج بأن الفرضية البديلة H_1 مقبولة، إذا تم قبول H_0 نستنتج بأن الفرضية H_0 يمكن أن تكون صحيحة. نؤكد بأنه عند عدم رفض H_0 يجب آلا نقول بأن H_0 مقبولة بل نقول H_0 " غير مرفوضة not rejected " أي نتجنب استخدام كلمة " مقبولة accept " في هذه الحالة بسبب إمكانية وقوعنا بخطأ من النوع الثاني واحتماله عادة كبير. إن رفض الفرضية H_0 لا يعني أنها خاطئة، بل لا نملك أية معلومات أخرى تساعدنا على قبولها .

[٢] إذا كان مستوى الدلالة ٥ % ، والاختبار من طرفين فإن منطقة الرفض ستكون واقعة خارج المجال $[-1.96 , +1.96]$

§- 2 اختبار الفرضيات لمتوسط المجتمع μ

Hypothesis Testing: A single Population Mean

سنحاول اختبار الفرضية حول متوسط المجتمع ضمن ثلاثة شروط:

أولاً: عندما تكون المعاينة من مجتمع طبيعي تباينه معلوم.

ثانياً: المعاينة من مجتمع طبيعي تباينه غير معلوم.

ثالثاً: المعاينة من مجتمع غير طبيعي .

2–1 المعاينة من مجتمع طبيعي تباينه معلوم :

عندما يكون الانحراف المعياري σ للمجتمع معلوماً فإن الإحصائية:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \qquad (5-1)$$

تتبع توزيع طبيعي معياري مهما كان حجم العينة المسحوبة من هذا المجتمع.

مثال (5–1)

لمعرفة متوسط الدخل الشهري للأشخاص العاملين في مجال التسويق، سحبت عينة عشوائية تضم 9 خريجين من كلية الاقتصاد من مجتمع طبيعي تباينه معلوم $\sigma^2 = 810000$ ، فكانت النتيجة أن متوسط الدخل في العينة هو : $\bar{x} = 7000$.

المطلوب:

هل يمكننا أن نستنتج بأن متوسط الدخل الشهري للأشخاص العاملين في مجال التسويق (المجتمع الإحصائي) هو مختلف عن 7500 ل.س شهرياً وذلك عند مستوى معنوية 0.05 ؟

الحل :

نستنتج بأن متوسط الدخل الشهري للأشخاص العاملين في مجال التسويق مختلف عن 7500 ل.س إذا استطعنا رفض فرضية العدم القائلة بأن متوسط الدخل يساوي 7500 ل.س، البيانات المتاحة هي نتائج العينة المسحوبة $\bar{x} = 7000$ ، العينة مسحوبة من

مجتمع (العاملين في مجال التسويق) موزع طبيعياً تباينه معلوم $\sigma^2 = 810000$.

– **الفرضيات:**

إن فرضية العدم تفترض عدم وجود فروق حقيقية بين متوسط المجتمع والقيمة المفروضة وتأخذ الشكل التالي:

$$H_0 : \mu = 7500$$

مقابل الفرضية البديلة:

$$H_1 : \mu \neq 7500$$

حيث الفرض البديل من طرفين .

–**إحصائية الاختبار:** إن إحصائية الاختبار :

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{7000 - 7500}{900 / \sqrt{9}} = -1.667$$

تتبع توزيعاً طبيعياً معيارياً (N(0,1 ، إذا كانت H_0 صحيحة .

– **قاعدة القرار:**

رفض فرضية العدم H_0 إذا كانت القيمة المحسوبة لإحصائية الاختبار z واقعة في منطقة الرفض ، وعدم رفض H_0 إذا وقعت قيمة z في منطقة القبول، علماً أن تحديد منطقتي القبول والرفض تعتمد على مستوى المعنوية α .

– عند مستوى معنوية $\alpha = 0.05$ تكون قيمة z المعيارية الموافقة لها هي ± 1.96.
أي أن منطقة الرفض ستتألف من كل قيم z التي هي أكبر من 1.96 أي على يمينها أو أصغر من 1.96- أي جميع القيم التي تقع على يسارها.

– **القرار الإحصائي:**

لا يمكننا رفض فرضية العدم طالما أن قيمة z غير واقعة ضمن منطقة الرفض ، ويمكننا القول بأن القيمة المحسوبة لإحصائية الاختبار ليست معنوية عند مستوى معنوية

$\alpha = 0.05$. ونستنتج مما تقدم بأن متوسط المجتمع الإحصائي يمكن أن يكون مساويا إلى $\mu = 7500$.

ملاحظات :

1- في مثالنا السابق اختبرنا الفرضية $H_0 : \mu = 7500$ ، مقابل الفرضية البديلة $H_1 : \mu \neq 7500$ ، ولم نتمكن من رفض H_0 بسبب وقوع قيمة إحصائية الاختبار في منطقة القبول، بإمكاننا الوصول لنفس النتيجة باستخدام مجال ثقة 95 % للمتوسط μ الذي يعطى بوساطة العلاقة :

$$7000 \pm 1.96 \sqrt{810000 / 9}$$
$$\Leftrightarrow 6412 \leq \mu \leq 7588$$

وبما أن مجال الثقة يتضمن القيمة 7500 نقول بأن هذه القيمة مرشحة لأن تكون المتوسط الذي نقوم بتقديره، وبالتالي يمكن أن تكون $\mu = 7500$ و H_0 غير مرفوضة، وهي نفس النتيجة التي توصلنا إليها سابقاً.

نستنتج مما تقدم: عند اختبار الفرضية H_0 بوساطة مجال الثقة، فإننا نرفض فرضية العدم H_0 عند مستوى معنوية α إذا كان المؤشر المختبر لا يقع ضمن $100(1-\alpha)\%$ مجال ثقة، وإذا وقع المؤشر المختبر ضمن المجال فإننا لا نستطيع رفض H_0 عند مستوى معنوية α .

2- إن اختبار الفرضيات الموضح سابقاً هو اختبار من طرفين Two-Sided Test ولذلك قسمت منطقة الرفض إلى منطقتين أو وزعت على طرفي توزيع إحصائية الاختبار ، إلا أن طبيعة السؤال الذي يطرحه الباحث قد يجعل اختبار الفرضيات من طرف واحد One-Sided Hypothesis Tests وفي هذه الحالة تقع منطقة الرفض إما على الطرف الأيمن وإما على الطرف الأيسر للتوزيع .

3- إن قيمة P **(P - Value)** في اختبار الفرضيات تعني الحصول ، عندما

تكون H_0 صحيحة، على قيمة حدية لإحصائية الاختبار المحسوبة فعلياً . ويمكن تعريفها أيضاً بأنها أصغر قيمة لـ α والتي من أجلها يمكن رفض H_0 ، فإذا كانت $P \le \alpha$ نرفض الفرضية H_0 ، وإذا كانت $P > \alpha$ فإننا لا نرفض H_0.

4- يمكن تلخيص بعض القيم الحرجة لـ Z المستخرجة من جدول التوزيع الطبيعي المعياري عند مستويات معنوية مختلفة في الجدول التالي:

مستوى المعنوية α	0.05	0.01
قيم Z : الفرض البديل من طرف واحد	±1.645	±2.33
قيم Z : الفرض البديل من طرفين	±1.96	±1.96

مثال (5-2) :

لنفترض في مثالنا السابق أن السؤال المطروح هو: هل يمكننا أن نستنتج بأن متوسط الدخل الشهري للعاملين بمجال التسويق هو أقل من 7500 ليرة سورية أي: .

$H_1 : \mu < 7500$ بدلاً من الفرضية: $H_1 : \mu \neq 7500$.

نستطيع أن نستنتج صحة الفرض البديل إذا تمكنا من رفض فرضية العدم القائلة بأن متوسط الدخل الشهري للعاملين بمجال التسويق يساوي 7500 ؟

الحل:

الفرضيات الواجب اختبارها هي:

- الفرضية الأولية (العدم): $H_0 : \mu = 7500$
- الفرضية البديلة: $H_1 : \mu < 7500$

حيث الفرض البديل في هذه الحالة من طرف واحد هو الطرف الأيسر .

- إن إحصائية الاختبار Z تتبع التوزيع الطبيعي المعياري وتحسب بموجب العلاقة :

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{7000 - 7500}{900 / \sqrt{9}} = -1.667$$

- عند مستوى معنوية $\alpha = 0.05$ تكون قيمة z المعيارية الموافقة لها هي -1.645 .

أي أن منطقة الرفض ستتألف من كل قيم z التي هي أصغر من -1.645 أي جميع القيم التي تقع على يسارها .

- القرار الإحصائي:

لا يمكننا قبول فرضية العدم طالما أن قيمة z غير واقعة ضمن منطقة القبول ، ويمكننا القول بأن القيمة المحسوبة لإحصائية الاختبار هي معنوية عند مستوى $\alpha = 0.05$.

اتخاذ القرار:

نرفض فرضية العدم H_0 القائلة بأن متوسط الدخل الشهري للأشخاص العاملين في مجال المعلوماتية يساوي 7500 ل.س شهرياً، و نستنتج بأن متوسط الدخل الشهري للأشخاص العاملين في مجال المعلوماتية يمكن أن يكون أقل من 7500 ل.س شهرياً .

مثال (5-3) :

تعاقدت إحدى الشركات على شراء كمية من الأنابيب الحديدية متوسط قطرها 10سم وانحرافها المعياري 0.5 سم ، قامت إدارة الشركة بسحب عينة عشوائية من الشحنة تضم 64 أنبوباً ، فكان متوسط قطرها 10.1 سم .

المطلوب:

هل تقبل المؤسسة هذه الشحنة أم ترفضها ؟ بمعنى أخر هل الفرق بين الوسط الحسابي للعينة والوسط الحسابي للمجتمع جوهري أم لا ؟ إذا علمت أن مستوى المعنوية هو 0.05 .

الحل :

الفرضيات الواجب اختبارها هي:

- الفرضية الأولية القائلة بعدم وجود فرق جوهري بين متوسط العينة ومتوسط المجتمع

$$H_0 : \mu = \mu = 10$$

- الفرضية البديلة القائلة بوجود فرق جوهري بين متوسط العينة ومتوسط المجتمع :

$$H_1 : \mu \neq \mu_0 = 10$$

حيث الفرض البديل في هذه الحالة من طرفين لأن الوسط الحسابي للشحنة يجب آلا يزيد ولا ينقص بشكل جوهري عن 10 سم .

- **إحصائية الاختبار Z** تتبع توزيع طبيعي معياري وتحسب بموجب العلاقة:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{10.1 - 10}{0.5 / \sqrt{64}} = \frac{0.8}{0.5} = +1.6$$

- عند مستوى معنوية $\alpha = 0.05$ والاختبار من طرفين تكون قيمة z الحرجة أو الجدولية الموافقة لها هي ± 1.96 .

اتخاذ القرار:

بما أن قيمة إحصائية الاختبار المحسوبة هي أصغر من قيمة z الحرجة ، نقبل فرضية العدم H_0 القائلة بأن الفرق بين الوسط الحسابي للعينة والوسط الحسابي للمجتمع ليس جوهرياً، أي نقبل الشحنة.

حساب قيمة α و γ في المثال السابق :

لنفترض أن المجتمع الإحصائي طبيعي (الأنابيب الحديدية التي تنتجها الشركة) انحرافه المعياري معلوم $\sigma = 0.5$ ونرغب في اختبار الفرضيتين التاليتين :

$$H_0 : \mu = \mu = 10$$

$$H_1 : \mu \neq \mu_1 = 11 > 10$$

سحبنا عينة عشوائية من هذا المجتمع حجمها $n = 64$ وبعد حساب الوسط الحسابي لهذه للعينة قررنا تطبيق القاعدة القرارية التالية:

- إذا كان $\bar{x} < 10.5$ فإننا نقرر الفرضية الأولية : H_0 .

- إذا كان $\bar{x} \geq 10.5$ فإننا نقرر الفرضية البديلة : H_1 .

المطلوب:

حساب قيمة الخطأ المرتكب من النوع الأول α (رفض الفرضية H_0 وهي صحيحة) وكذلك قيمة الخطأ المرتكب من النوع الثاني γ (قبول الفرضية H_0 وهي خاطئة)

الحل:

إن احتمال ارتكاب خطأ من النوع الأول α يساوي إلى احتمال أن نقبل الفرضية البديلة H_1 في الوقت الذي تكون فيه فرضية العدم صحيحة أي :

$$\alpha = P(H_1 / H_0 \quad Correct \ )$$
$$= P(\bar{x} \geq 10.5 / \mu = \mu_0 = 10)$$

وللانتقال إلى التوزيع الطبيعي المعياري نكتب:

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{\bar{x} - 10}{0.0625} \Rightarrow 0.0625Z + 10 = \bar{x}$$

وبتعويض $\bar{x}$ بما تساويها في العلاقة السابقة نجد ما يلي :

$$\alpha = P(0.0625Z + 10 \geq 10.5) = P(Z \geq \frac{0.5}{0.0625})$$
$$= P(Z \geq 8) = 1 - P(Z \leq 8) = 1 - 0.999999 = 0.0000001$$

إن احتمال ارتكاب خطأ من النوع الأول يساوي $\alpha = 0.0000001$.

إن احتمال ارتكاب خطأ من النوع الثاني γ يساوي إلى احتمال أن نقبل فرضية العدم H_0 في الوقت الذي تكون فيه الفرضية البديلة H_1 صحيحة أي :

$$\gamma = P(H_0 / H_1 \quad Correct \ )$$
$$= P(\bar{x} < 10.5 / \mu = \mu_1 = 11)$$

وللانتقال إلى التوزيع الطبيعي المعياري نكتب:

$$Z = \frac{\bar{x} - \mu_1}{\sigma / \sqrt{n}} = \frac{\bar{x} - 11}{0.0625} \Rightarrow 0.0625Z + 11 = \bar{x}$$

وبتعويض $\bar{x}$ بما تساويها في العلاقة السابقة نجد ما يلي:

$$\gamma = P(0.625Z + 11 < 10.5) = P(Z < \frac{-0.5}{0.0625})$$
$$= P(Z < -8) = 1 - P(Z \le 8) = 1 - 0.0000001 = 0.0000001$$

أي أن احتمال ارتكاب خطأ من النوع الثاني يساوي $\gamma = 0.0000001$ وهو نفس الاحتمال السابق ويسمى هذا الاحتمال بقوة الاختبار .

مثال (5-4):

ليكن معلوماً أن متوسط الوزن للعجول الجاهزة للبيع في مزرعة متخصصة بالتسمين بمنطقة المسلمية في محافظة حلب هو 200 كغ ، بانحراف معياري معلوم ويساوي 18 كغ ، تم سحب عينة عشوائية تضم 36 عجلاً من المزرعة فتبين أن متوسط الوزن لهذه العجول هو 190 كغ .

المطلوب اختبار معنوية الفرق بين هذا المتوسط ومتوسط المجتمع الإحصائي (أوزان العجول في المزرعة) وذلك عند مستوى معنوية 0.05؟

الحل :

الفرضيات الواجب اختبارها هي:

- الفرضية الأولية القائلة بعدم وجود فرق معنوي بين متوسط العينة ومتوسط المجتمع: $H_0 : \mu = 200$
- الفرضية البديلة القائلة بوجود فرق معنوي بين متوسط العينة ومتوسط المجتمع : $H_1 : \mu \neq 200$

حيث الفرض البديل في هذه الحالة من طرفين.

- إحصائية الاختبار Z تتبع توزيع طبيعي معياري وتحسب بموجب العلاقة :

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{190 - 200}{18 / \sqrt{36}} = -3.33$$

- عند مستوى معنوية $\alpha = 0.05$ والاختبار من طرفين تكون قيمة z المعيارية الموافقة لها هي ±1.96 .

- **القرار الإحصائي:**

لا يمكننا قبول فرضية العدم طالما أن قيمة z غير واقعة ضمن منطقة القبول ، ويمكننـــا القول بأن القيمة المحسوبة لإحصائية الاختبار هي معنوية عند مستوى $\alpha = 0.05$.

- **اتخاذ القرار:**

نرفض فرضية العدم H_0 القائلة بعدم وجود فرق معنوي بين متوسط العينة ومتوسط المجتمع ، أي أن هناك فرق معنوي بين المتوسطين وبالتالي يمكن القول بـــأن العجـــول بحاجة لفترة إضافية لكي تكون جاهزة للبيع .

2-2 المعاينة من مجتمع طبيعي تباينه غير معلوم :

نواجه غالبا مواقف تتطلب الاستدلال حول متوسط المجتمع، وعندما تكـــون المعاينة من مجتمع طبيعي تباينه غير معلوم فإن إحصائية الاختبار المســـتخدمة لاختبـــار الفرضية:

$$H_0 : \mu = \mu_0$$

تكون t بدلاً من Z وتأخذ الصيغة التالية :

$$t = \frac{\overline{x} - \mu_0}{s / \sqrt{n}} \qquad (5-2)$$

والتي تتبع ، عندما تكون H_0 صحيحة ، توزيع t لستودنت مع n-1 درجة حرية .

مثال (5-5) :

في دراسة هدفها التعرف على متوسط النفقات اليومية لطلاب كلية الاقتصـــاد قســـم الاحصاء والبرمجة، تم سحب عينة عشوائية تضم 15 طالباً فكانـــت النتـــائج التاليـــة ($\overline{x} = 96 , s = 35$) ل.س.

المطلوب:

ضمن مستوى معنوية $\alpha = 0.05$ معرفة فيما إذا كان متوسط المجتمع الإحصائي الذي سحبت منه العينة يختلف عن 120 ل.س ؟

الحل:

البيانات عبارة عن النفقات اليومية الخاصة بعينة عشوائية تضم 15 طالباً من كلية الاقتصاد قسم الاحصاء والبرمجة ($\bar{x} = 96 , s = 35$) مسحوبة من مجتمع طبيعي تباينه غير معلوم .

الفرضيات الواجب اختبارها هي:

- الفرضية الاولية (العدم): $H_0 : \mu = 120$
- الفرضية البديلة : $H_1 : \mu \neq 120$

حيث الفرض البديل من طرفين .

- إحصائية الاختبار:

بما أن تباين المجتمع غير معلوم فان إحصائية الاختبار t تتبع توزيع ستودنت مع n-1 درجة حرية، عندما تكون H_0 صحيحة، وبحسابها نجد:

$$t = \frac{96 - 120}{35/\sqrt{15}} = -2.65$$

عند مستوى معنوية $\alpha = 0.05$ و $\nu = 14$ درجة حرية فإن القيمة الحرجة لإحصائية الاختبار هي:

$$t_{1-\alpha/2} = t_{0.975} = \pm 2.1448$$

وعند مستوى معنوية $\alpha = 0.01$ و $\nu = 14$ درجة فإن القيمة الحرجة لإحصائية الاختبار هي :

$$. \ t_{1-\alpha/2} = t_{0.995} = \pm 2.9768$$

اتخاذ القرار:

عند مستوى معنوية $\alpha = 0.05$ نرفض فرضية العدم H_0 القائلة بأن متوسط النفقات

اليومية للطلبة يساوي 120 ل.س لأن قيمة إحصائية الاختبار بالقيمة المطلقة أكبر من القيمة الحرجة لـ t ، أي أن :

$$t = -2.65 \leq t_{1-\alpha/2} = -2.1448$$

تقع في منطقة الرفض ، وبالتالي نستنتج ، بأن متوسط المجتمع الذي سحبت منه العينة لا يساوي 120 .

وعند مستوى معنوية $\alpha = 0.01$ نقبل فرضية العدم H_0 القائلة بأن متوسط النفقات اليومية للطلبة يساوي 120 ل.س لأن قيمة إحصائية الاختبار بالقيمة المطلقة أصغر من القيمة الحرجة لـ t التي تساوي ± 2.9768 أي أن متوسط المجتمع الذي سحبت منه العينة يساوي 120 ل.س .

نلاحظ بأننا رفضنا فرضية العدم عند مستوى معنوية 0.05 وقبلناها عند مستوى معنوية 0.01 عندما زادت درجة الثقة لدينا إلى 0.99 .

ملاحظة هامة جداً :

إذا كانت العينة التي نعتمد عليها في اختبار الفرض حول متوسط مجتمع مسحوبة من مجتمع غير طبيعي، يمكننا إذا كانت العينة كبيرة $n \geq 30$ أن نطبق نظرية النهاية المركزية ونستخدم z كإحصائية اختبار، وإذا كان الانحراف المعياري للمجتمع مجهولاً نستخدم عادة الانحراف المعياري للعينة كتقدير له.

وتصبح إحصائية الاختبار بهذه الحالة:

$$z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

والتي تتبع ، عندما تكون H_0 صحيحة ، تقريبا التوزيع الطبيعي المعياري إذا كانت $n \geq 30$.

مثال (5-6) :

في مسح اجتماعي صحي لمجتمع المعلمين في جامعة حلب قام بإجرائه صندوق التكافل

الاجتماعي بنقابة المعلمين، تم مقابلة عينة عشوائية تضم 150 شخصا ، وواحدة من المعلومات التي تم الحصول عليها هو عدد الوصفات الطبية التي صرفها الشخص الواحد خلال العام الماضي ، فكان متوسط عدد الوصفات للعينة 5.8 وصفة بانحراف معياري 3.1 وصفة.

المطلوب:

هل هذه البيانات كافية لتشير بأن متوسط المجتمع (عدد الوصفات المصروفة سنوياً للمعلم) هو أكبر من 5 وصفات ؟

الحل:

البيانات عبارة عن عدد الوصفات التي قام بصرفها أفراد عينة تضم 150 شخصاً يعمل في مجال نقابة المعلمين بالجامعة (بمتوسط 5.8 وانحراف معياري 3.1 وصفة) مسحوبة من مجتمع قد لا يكون طبيعياً .

- الفرضيات هي: الفرضية الأولية (العدم) القائلة بأن متوسط عدد الوصفات المصروفة يساوي 5 أي[3] : $H_0 : \mu = 5$ ، مقابل الفرضية البديلة القائلة بأن متوسط عدد الوصفات المصروفة أكبر من 5 : $H_1 : \mu > 5$ ، والفرض البديل من طرف واحد أيمن .

- إحصائية الاختبار بهذه الحالة هي:

$$z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

تتبع تقريبا توزيع طبيعي معياري مع $\mu = 0$ إذا كانت H_0 صحيحة ، مع العلم أن الانحراف المعياري للمجتمع الإحصائي σ مجهولة[4] ، وبحسابها نجد :

$$z = \frac{5.8 - 5}{3.1 / \sqrt{150}} = 0.8 / 0.25 = 3.2$$

[3] الفرضية H_0 تعبر عن المواصفات القياسية المعتمدة أي القيمة الافتراضية للموشر المدروس ، في حين تعبر الفرضية البديلة إلى حد ما عن ما تشير إليه نتائج المعاينة من معلومات .

[4] على الرغم من عدم معلومية تباين المجتمع استخدمنا إحصائية الاختبار Z وذلك بسبب كون حجم العينة كبيراً .

– عند مستوى معنوية $\alpha = 0.05$ فإن القيمة الحرجة لـــ z تساوي 1،645 .

اتخاذ القرار:

بما أن قيمة إحصائية الاختبار أكبر من القيمة الحرجة أو الجدولية أي : $z = 3.2 > z'_\alpha = 1.645$ فإننا لا نستطيع قبول فرضية العدم H_0 ، ونستنتج بأن متوسط عدد الوصفات التي يصرفها المعلمون في جامعة حلب سنويا هو أكبر من 5 .

مثال (5–7) :

يقال بأن متوسط سعر الكغ الواحد من البندورة قد بلغ في شهر نيسان بمدينة حلب خلال العام الماضي 12 ل.س . تم أخذ مشاهدات متتالية بأيام مختلفة ولفترة 10 أيام للتعرف على متوسط أسعار البيع في المدينة فكانت النتائج التالية :

14 – 12 – 13 – 15 – 14 – 13 – 12 – 13 – 12 – 13

المطلوب:

اختبار فيما إذا كان متوسط سعر البندورة هو أعلى من 12 ليرة سورية وذلك عند مستوى معنوية 0.05 وأيضاً 0.01 ؟

الحل :

البيانات عبارة عن أسعار البندورة التي قامت برصدها لجنة مختصة خلال 10 أيام اعتبرت بمثابة عينة عشوائية تضم 10 مشاهدات (بمتوسط 13.1 ل.س وانحراف معياري 0.994 ل.س).

الوسط الحسابي لأسعار العينة:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{131}{10} = 13.1$$

الانحراف المعياري لأسعار العينة:

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{8.9}{9}} = 0.99 = 0.994$$

الفرضيات هي:

- الفرضية الأولية القائلة بأن متوسط سعر الكغ الواحد من البندورة في مدينة حلب في شهر نيسان هو 12 ل.س ، أي $H_0 : \mu = 12$

- الفرضية البديلة القائلة بأن متوسط السعر هو أعلى من 12 ل.س ، أي $H_1 : \mu > 12$

حيث الفرض البديل من طرف واحد أيمن .

- إحصائية الاختبار هي:

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

تتبع تقريبا توزيع ستودنت، لأن الانحراف المعياري للمجتمع σ مجهول وحجم العينة صغير جداً ، وبحسابها نجد :

$$t = \frac{13.1 - 12}{0.994 / \sqrt{10}} = 3.5$$

- عند مستوى معنوية $\alpha = 0.01$ و 9 درجات حرية $\nu = n - 1 = 10 - 1 = 9$ فإن القيمة الحرجة لـــ t تساوي 3.250 .

اتخاذ القرار:

بما أن : $t = 3.5 > t_\alpha = 3.250$ فإننا لا نستطيع قبول فرضية العــدم H_0 ، أي أن فرضية العدم مرفوضة ، ونستنتج بأن متوسط أسعار البندورة في مدينة حلب هو أكبر من 12 ل.س في شهر نيسان .

مثال (5–8) :

تنتج آلة قطع معدنية على شكل سلسلة ، وقد تم ضبط الآلة ليكون قطر القطع المنتجة 24 مم ، وللتأكد من ذلك سحبت عينة عشوائية مؤلفة من 49 قطعة فوجد أن متوسط قطر القطع $\bar{x} = 24.2\ mm$ وتباينها $s^2 = 0.09\ mm$.

المطلوب:

اختبار فيما إذا كان ضبط الآلة يعتبر دقيقاً أم لا ؟ وذلك عند مستوى معنوية 0.01 .

الحل :

– الفرضية الأولية القائلة بأن متوسط قطر القطع المعدنية هو 24 مم أي :

$$H_0 : \mu = 24$$

– الفرضية البديلة القائلة بأن متوسط قطر القطع المعدنية هو مختلف عن 24 مـــم ، أي $H_1 : \mu \neq 24$ حيث الفرض البديل من طرفين .

– إحصائية:

$$Z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

تتبع تقريبا توزيع طبيعي معياري ، الانحراف المعياري للمجتمع σ مجهول ويســتبدل بالانحراف المعياري للعينة .

$$Z = \frac{24.2 - 241 - 12}{0.3 / \sqrt{49}} = 4.667$$

– عند مستوى معنوية $\alpha = 0.01$ و اختبار من طرفين فإن قيمة Z الحرجة أو الجدولية تساوي ± 1.96

اتخاذ القرار:

بما أن قيمة إحصائية الاختبار المحسوبة أكبر من القيمة الحرجة أو الجدولية لإحصـــائية الاختبار ، فإننا لا نستطيع قبول فرضية العدم H_0 ، أي أن فرضية العدم مرفوضة ، ونستنتج بأن ضبط الآلة غير دقيق .

3-§ اختبار الفرضيات للفرق بين متوسطي المجتمعين :

في الحياة العملية ، غالبا ما تواجهنا مواقف تتطلب منا أن نستنتج فيما إذا كان هناك فرق أم لا بين متوسطي المجتمعين المدروسين ، وفي مثل هذه الحالات يمكن صياغة واحدة من الفرضيات التالية :

أ- اختبار من طرفين :

$$H_0 = \mu_1 - \mu_2 = 0$$
$$H_1 = \mu_1 - \mu_2 \neq 0$$

ب- اختبار من طرف واحد أيسر :

$$H_0 = \mu_1 - \mu_2 = 0$$
$$H_1 = \mu_1 - \mu_2 < 0$$

ج- اختبار من طرف واحد أيمن :

$$H_0 = \mu_1 - \mu_2 = 0$$
$$H = \mu_1 - \mu_2 > 0$$

يمكن مناقشة اختبار الفرضيات للفرق بين متوسطي المجتمعين في ثلاث حالات :

3-1 المعاينة من مجتمع طبيعي تباينه معلوم :

عندما تكون العينتان العشوائيتان مسحوبتين من مجتمعين طبيعيين تباينهما معلوم ، فإن إحصائية الاختبار المستخدمة لاختبار فرضية العدم القائلة بتساوي متوسطي المجتمعين :

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{(\sigma_1^2 / n_1) + (\sigma_2^2 / n_2)}} \qquad (5-3)$$

تتبع التوزيع الطبيعي المعياري عندما تكون H_0 صحيحة .

مثال (5-9) :

يرغب فريق بحث اجتماعي في معرفة ما إذا كانت البيانات التي تم جمعها من قبلهم كافية لتشير إلى وجود فرق بين متوسط درجات الثقة بالنفس بين أشخاص طبيعيين

وأشخاص ينتمون لأسر تعاني من تفكك اجتماعي . البيانات عبارة عن نتائج اختبار متخصص (تم إجراءه بإشراف مختصين نفسيين واجتماعيين خصص له 5 درجات كحد أعلى) لدرجات الثقة الخاصة بعينتين عشوائيتين مسحوبتين من مجتمعين تباينهما يساوي الواحد ، تضم الأولى 12 شخصاً ينتمون لأسر تعاني من مشاكل اجتماعية ، والثانية تضم 15 شخصاً ينتمون لأسر طبيعية ، متوسط العينة الأولى 4.5 درجة ومتوسط العينة الثانية 3.5درجة .

الحل :

البيانات هي عبارة عن قراءات عينتين عشوائيتين مسحوبتين من مجتمعين طبيعيين تباينهما يساوي الواحد الصحيح .

- الفرضيات هي :

$$H_0 = \mu_1 - \mu_2 = 0$$
$$H_A = \mu_1 - \mu_2 \neq 0$$

حيث الفرض البديل من طرفين ، ويمكن وضع الفرضية بطريقة أخرى :

$$H_0 = \mu_1 = \mu_2$$
$$H_A = \mu_1 \neq \mu_2$$

– إحصائية الاختبار:

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

تتبع توزيع طبيعي معياري إذا كانت H_0 صحيحة وبحسابها نجد:

$$Z = \frac{(4.5 - 3.4) - 0}{\sqrt{(1/12) + (1/15)}} = 1.1/0.39 = 2.82$$

بفرض أن $\alpha = 0.05$ فان القيمة الحرجة لإحصائية الاختبار z تساوي ± 1.96 .

القرار هو رفض H_0 لأن $z_{computed} = 2.82 > 1.96$ (قيمة إحصائية الاختبار

المحسوبة) تقع في منطقة الرفض . وبالتالي نستنتج ، بأن متوسطي المجتمعين غير متساويين ، أي هناك فرق بين متوسط درجات الثقة بالنفس بين الأشخاص الذين ينتمون لأسر طبيعية وأشخاص ينتمون لأسر تعاني من تفكك اجتماعي .

وعند مستوى معنوية $\alpha = 0.01$ فإن القيمة الحرجة لـ z تساوي ± 2.58. والقرار هو رفض H_0 أيضاً لأن $z_{computed} = 2.82 > 2.58$ تقع في منطقة الرفض .

وبالتالي نستنتج ، بأن متوسط المجتمعين غير متساويين ، أي هناك فرق بين متوسط درجات الثقة بالنفس بين الأشخاص الذين ينتمون لأسر طبيعية وأشخاص ينتمون لأسر تعاني من تفكك اجتماعي .

2-3 المعاينة من مجتمع طبيعي تباينه غير معلوم :

عندما تكون العينتان العشوائيتان مسحوبتين من مجتمعين طبيعيين تباينهما غير معلوم فهناك إمكانيتان الأولى تساوي التباينين والثانية عدم تساويهما.

- **حالة تساوي تبايني المجتمعين:**

عندما تكون العينتين العشوائيتين مسحوبتين من مجتمعين طبيعيين تباينهما متساو وغير معلوم فإننا نستخدم تباين العينة المرجح :

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \qquad (5-4)$$

وإحصائية الاختبار t تتبع ، عندما تكون H_0 صحيحة ، توزيع ستودنت مع : $v = n_1 + n_2 - 2$ درجة حرية ، وتحسب بموجب العلاقة التالية :

$$t = \frac{(\overline{x}_1 - \overline{x}_2)}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}} = \frac{(\overline{x}_1 - \overline{x}_2)}{S_p \cdot \sqrt{\frac{n_1 + n_2}{n_1 . n_2}}}. \qquad (5-5)$$

مثال (5–10) :

لمعرفة الفرق بين المستوى الاقتصادي – الاجتماعي (المهني) للأمريكيين الصينيين و اليابانيين مقارنة بالأمريكيين البيض من أصول غير أسبانية ، تم اختيار عينة عشوائية من الأشخاص (من حملة الشهادة الجامعية) والذين شملهم التعداد السكاني لعام 1980م فكانت النتائج كما هي موضحة بالجدول التالي الذي يوضح الوسط الحسابي والانحراف المعياري لدخل حملة الشهادة الجامعية بولاية كاليفورنيا بالدولار .

البيض من غير الأسبان	اليابانيون	الصينيون	
24891	22907	21439	$\bar{X}$ الوسط الحسابي
14225	11120	10289	الانحراف المعياري S
14225	11120	10289	حجم العينة

يتضح من الجدول بأن متوسط الدخل والانحراف المعياري للدخل بالنسبة للأمريكيين صيني أو ياباني الأصل مقارنة بمتوسط الدخل والانحراف المعياري للدخل بالنسبة للبيض من أصول غير أسبانية ، الذين يقيمون في كاليفورنيا وممن يحملون الشهادة الجامعية ليس هو نفسه على الرغم من تماثل الاعتمادية (الدرجة الجامعية) وكون الوظيفة هي نفسها وبنفس الولاية كاليفورنيا .

إن متوسط الدخل للبيض يعادل 24891 دولار وهو أعلى من متوسط الدخل للأمريكيين من أصل ياباني الذي يعادل 22907 دولار ، وأعلى من متوسط الدخل للأمريكيين من أصل صيني الذي يعادل 21439 دولاراً .

والمطلوب:

هل يعتبر الفرق بين متوسط الدخل للأمريكيين البيض ومتوسط الدخل للأمريكيين من أصل ياباني ذي دلالة إحصائية أم لا :

$$\mu_1 - \mu_2 = 24891 - 22907 = 1984\ \$$$

وكذلك هل يعتبر الفرق بين متوسط الدخل للأمريكيين البيض من غير الأسبان ومتوسط الدخل للأمريكيين من أصل صيني ذي دلالة إحصائية أم لا:

$$\mu_1 - \mu_3 = 24891 - 21439 = 3452\ \$$$

وذلك عند مستوى معنوية 0.05 في الحالة الأولى ، و 0.01 في الحالة الثانية .

<u>الحل في الحالة الأولى:</u>

– **الفرضية الأولية (العدم):** نقول بأنه لا توجد فروق ذات دلالة إحصائية ما بين متوسط الدخل للأمريكيين البيض ومتوسط الدخل للأمريكيين من أصل ياباني ، مقابل الفرضية البديلة القائلة بوجود فرق معنوي بين هذين المتوسطين، أي نكتب:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

– **حساب الخطأ المعياري المقدر Calculating the Estimated Standard Error** : لحساب إحصائية الاختبار t لا بد من حساب الخطأ المعياري المقدر مفترضين أن تباين المجتمعين متساويين ، وبهذه الحالة ندمج المعلومات الخاصة بتبايني العينتين لحساب الخطأ المعياري المقدر لتوزيع المعاينة بموجب العلاقة التالية :

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \cdot \frac{n_1 + n_2}{n_1 . n_2}}$$

$$S_p = \sqrt{\frac{(2122)(14225)^2 + (757)(11120)^2}{2123 + 758 - 2} \cdot \frac{2123 + 758}{2123(758)}} = 570.3$$

– درجات الحرية في مثالنا تساوي:

$$v = n_1 + n_2 - 2 = 2123 + 758 - 2 = 2881$$

- **حساب إحصائية الاختبار:**

إن إحصائية الاختبار t تتبع ، عندما تكون H_0 صحيحة ، توزيع t لستودنت مع $\nu = n_1 + n_2 - 2$ درجة حرية ، وتحسب بموجب العلاقة التالية :

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p} = \frac{24891 - 22907}{570.3} = 3.48$$

وبمقارنة قيمة إحصائية الاختبار المحسوبة مع قيمة t الحرجة والمستخرجة من جدول توزيع ستودنت عند ٢٨٨١ درجة حرية نلاحظ من العمود الأول من الجدول المخصص لدرجات الحرية عدم وجود الرقم ١٤٤٤ ولهذا السبب يستخدم السطر الأخير في الجدول من أجل درجات حرية تساوي لا نهاية $df = \infty$.

عند مستوى معنوية $\alpha = 0.05$ ودرجات حرية $df = \infty$ فإن القيمة الجدولية تساوي 1.96[°] وهي أصغر من قيمة إحصائية الاختبار وبالتالي نرفض فرضية العدم عند مستوى معنوية $\alpha = 0.05$ ونستنتج بأن هناك فرق ذي دلالة إحصائية بين متوسط الدخل للأمريكيين البيض ومتوسط الدخل للأمريكيين من أصل ياباني .

الحل في الحالة الثانية :

الفرضية الأولية تقول بأنه لا توجد فروق ذات دلالة إحصائية ما بين متوسط الدخل للأمريكيين البيض ومتوسط الدخل للأمريكيين من أصل صيني، مقابل الفرضية البديلة القائلة بوجود فرق معنوي بين هذين المتوسطين .أي نكتب :

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

– **حساب الخطأ المعياري المقدر** : لحساب إحصائية الاختبار t لا بد من حساب الخطأ المعياري المقدر مفترضين أن تباين المجتمعين متساويين ، بموجب العلاقة التالية :

[°] وهي نفس القيمة الحرجة التي يمكن أن نجدها في جدول التوزيع الطبيعي المعياري عند مستوى المعنوية ذاته واختبار من طرفين ولهذا السبب عندما يكون حجم العينتين كبيراً نقول بأن إحصائية الاختبار تتبع التوزيع الطبيعي المعياري .

$$S_p = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1.n_2}}$$

$$S_p = \sqrt{\frac{(2122)(14225)^2+(470)(10289)^2}{2123+471-2}\cdot\frac{2123+471}{2123(471)}} = 692.23$$

– درجات الحرية في مثالنا تساري :

$$v = n_1 + n_2 - 2 = 2123 + 471 - 2 = 2592$$

– **حساب إحصائية الاختبار:** إن إحصائية الاختبارt تتبع ، عندما تكون H_0 صحيحة ، توزيع ستودنت مع $v = n_1 + n_2 - 2$: درجة حرية ، وتحسب بالعلاقة:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p} = \frac{24891 - 21439}{692.23} = 4.987$$

وبمقارنة قيمة إحصائية الاختبار المحسوبة مع قيمة t الحرجة والمستخرجة من جدول **توزيع ستودنت عند 2592 درجة حرية ، ومستوى معنوية $\alpha = 0.01$ والتي تساوي 2.575 وهي أصغر من قيمة إحصائية الاختبار وبالتالي نرفض فرضية العدم عند مستوى معنوية $\alpha = 0.01$ ونستنتج بأن هناك فرق ذي دلالة إحصائية بين متوسط** الدخل للأمريكيين البيض ومتوسط الدخل للأمريكيين من أصل صيني .

مثال (11–5) :

استنادا إلى نتائج المسح العام[6] للمجتمع الأمريكي عام 1996 والمسمى GSS96[7] لقد بلغ متوسط عدد السنوات التي قضاها الرجال في التعليم 13.59 سنة ، في حين بلغ متوسط عدد السنوات التي قضتها النساء في التعليم 13.26 سنة وذلك كما هو موضح في الجدول التالي.

---

[6] GSS96: The General Society Survey has been conducted annually since 1972,for the National Data Program for the Social Sciences at the National Opinion Research Center at University of Chicago.

[7] هذا المسح يتم سنوياً منذ عام ١٩٧٢ وهو بمثابة برنامج المعلومات للعلوم الاجتماعية يقوم بإجرائه مركز بحث الرأي الوطني في جامعة شيكاغو .

المطلوب:

هل يعتبر الفرق بين متوسط عدد السنوات التي قضاها الرجال في التعليم و متوسط عدد السنوات التي قضتها النساء في التعليم:

$$\mu_1 - \mu_2 = 13.59 - 13.26 = 0.33$$

ذي دلالة إحصائية وذلك عند مستوى معنوية ألفا 0.05 وكذلك 0.01 .

عدد سنوات التعليم للرجال والنساء طبقاً لنتائج GSS96

	الرجال العينة الأولى	النساء العينة الثانية
$\bar{X}$ الوسط الحسابي	13.59	13.26
الانحراف المعياري S	3.03	2.91
التباين S^2	9.18	8.47
حجم العينة	635	811

المصدر : Chava Frankfort مرجع سابق ذكره ، ص ٤٧٥ .

الحل :

- **الفرضية الأولية (العدم)** تقول بأنه لا توجد فروق ذات دلالة إحصائية ما بين متوسط عدد السنوات التي قضاها كل من الرجال والنساء في التعليم ، مقابل الفرضية البديلة القائلة بوجود فرق معنوي بين هذين المتوسطين ، أي نكتب :

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

- **حساب الخطأ المعياري المقدر :**

لحساب إحصائية الاختبار t لا بد من حساب الخطأ المعياري المقدر ، مفترضين أن تبايني المجتمعين متساويين ، بموجب العلاقة التالية:

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \cdot \frac{n_1 + n_2}{n_1 . n_2}}$$

$$S_p = \sqrt{\frac{(635-1)3.03^2 + (811-1)2.91^2}{635+811-2} \cdot \frac{635+811}{635(811)}} = 0.16$$

- درجات الحرية في مثالنا تساوي : $v = n_1 + n_2 - 2 = 635 + 811 - 2 = 1444$

- **حساب إحصائية الاختبار:**

إن إحصائية الاختبار t تتبع ، عندما تكون H_0 صحيحة ، توزيع t لستودنت مع $v = n_1 + n_2 - 2$ درجة حرية ، وتحسب بالعلاقة :

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p} = \frac{13.59 - 13.26}{0.16} = 2.0625$$

وبمقارنة قيمة إحصائية الاختبار المحسوبة مع قيمة t الحرجة والمستخرجة من جدول توزيع ستودنت عند 1444 درجة حرية ، نلاحظ أنه عند مستوى معنوية $\alpha = 0.05$ ودرجات حرية $df = \infty$ فإن القيمة الجدولية تساوي 1.96 وهي أصغر من قيمة إحصائية الاختبار وبالتالي نرفض فرضية العدم عند مستوى معنوية $\alpha = 0.05$ ونستنتج بأن هناك فرق بين متوسط عدد السنوات التي قضاها الرجال في التعليم و متوسط عدد السنوات التي قضتها النساء في التعليم .

جدول توزيع ستودنت من أجل بعض القيم المختارة لدرجات الحرية

$\alpha/2$ ONE -TAILED TEST	..10	..05	..025	..01	..005	..0005
df	α TWO -TAILED TEST	..10	..05	..02	..01	..001
30	1.31	1.697	2.042	2.457	2.75	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2	2.39	2.66	3.46
120	1.289	1.658	1.98	2.358	2.617	3.373
$df = \infty$	1.282	1.645	1.96	2.326	2.576	3.291

عند مستوى معنوية $\alpha = 0.01$ ودرجات حرية $df = \infty$ فإن القيمة الجدولية تساوي 2.576 وهي أكبر من قيمة إحصائية الاختبار وبالتالي نقبل فرضية العدم عند مستوى معنوية $\alpha = 0.01$ ونستنتج بعدم وجود فرق ذي دلالة إحصائية بين متوسط عدد السنوات التي قضاها الرجال و متوسط عدد السنوات التي قضتها النساء في التعليم .

3-3 المعاينة من مجتمع غير طبيعي :

عندما تكون المعاينة من مجتمعات غير طبيعية ، نتبع نفس الإجراءات كما في حالة المعاينة من مجتمعات طبيعية إذا كان حجم العينتين n_1 , n_2 العشوائيتين والمسحوبتين من مجتمع غير طبيعي كبيراً ، وذلك كنتيجة لنظرية النهاية المركزية ، حيث أن إحصائية الاختبار z تتبع توزيع طبيعي إذا كانت H_0 صحيحة وكان تباينا المجتمعين معلومين ، أما في حالة كون تباين المجتمعين غير معلوم فيستخدم تباين العينات الكبيرة كتقدير لها ولا يتم الترجيح بهذه الحالة لأن تساوي تباين المجتمعين ليست فرضية ضرورية عند استخدام الإحصائية z .

مثال (5-12) :

لمعرفة فيما إذا كان متوسط دخل الأسرة الشهري للمرضى المقبولين في المشافي الخاصة A هو أكبر من متوسط دخل الأسرة الشهري للمرضى المقبولين في المشافي العامة B . تم سحب عينتين عشوائيتين الأولى تضم 75 مريضاً تم قبولهم في المشافي الخاصة A فكان متوسط دخل الأسرة في العينة 6800 ل.س والثانية تضم 80 مريضاً تم قبولهم في المشافي العامة B فكان متوسط دخل الأسرة في العينة 5450 ل.س . الجواب سيكون نعم، إذا استطعنا رفض فرضية العدم القائلة بأن $\mu_1 \leq \mu_2$ علما أن المجتمعين طبيعيين وانحرافهما المعياري على الترتيب $\sigma_1 = 600$ و $\sigma_2 = 500$.

الحل :

تمثل البيانات عينتين عشوائيتين مسحوبتين من مجتمعين طبيعيين تباينهما معلوم.

- الفرضيات: هي اختبار فرضية العدم القائلة بعدم وجود فرق ذي دلالة إحصائية بين متوسط دخل الأسرة الشهري في المجتمعين مقابل الفرضية البديلة التي تقول بأن متوسط دخل الأسرة الشهري للمرضى المقبولين في المشافي الخاصة هو أكبر من متوسط دخل الأسرة الشهري للمرضى المقبولين في المشافي العامة.

$$H_0 = \mu_1 - \mu_2 = 0$$
$$H_A = \mu_1 - \mu_2 > 0$$

إن الفرض البديل هو من طرف واحد أيمن .

إحصائية الاختبار z تتبع توزيع طبيعي معياري بحسابها نجد :

$$Z = \frac{(6800 - 5450) - 0}{\sqrt{(600)^2 / 75 + (500)^2 / 80}} = 1350 / 89 = 15.17$$

عند مستوى معنوية $\alpha = 0.01$ والاختبار من طرف واحد فإن القيمة الحرجة لـ z تساوي 2.33 . وعند مستوى معنوية $\alpha = 0.05$ والاختبار من طرف واحد فان القيمة الحرجة لـ z تساوي 1.645 .

- **اتخاذ القرار:** رفض H_0 لأن $z_{computed} = 15.17 > 2.33$ تقع في منطقة الرفض . وبالتالي نستنتج ، بأن متوسط دخل الأسرة لمرضى المشافي الخاصة هو أكبر من متوسط دخل الأسرة الشهري لمرضى المشافي العامة ، أي أن المجتمعين غير متساويين.

4-§ اختبار الفرضيات للنسبة P :

اختبار الفرضيات حول النسبة في مجتمع يعالج بنفس الطريقة التي اتبعناها بحالة اختبار الفرضيات لمتوسط المجتمع وخاصة عندما تتحقق الشروط الضرورية لاستخدام التوزيع الطبيعي . بفرض أن نسبة ظاهرة معينة في المجتمع هي P ، و $\hat{p}$ هي نسبة هذه الظاهرة في عينة حجمها n ($n \geq 30$). والمطلوب معرفة فيما إذا كان هناك فرق معنوي بين نسبة الظاهرة في كل من العينة والمجتمع .

إن إحصائية الاختبار المستخدمة هنا هي :

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 . q_0 / n}} \qquad (5-6)$$

وهذه الإحصائية تتبع التوزيع الطبيعي المعياري تقريباً.

مثال (5–13) :

نظراً لكثرة المخالفات لأنظمة المرور ووقوع العديد من الحوادث المرورية بسبب هذه المخالفات ، لنفترض أن وزارة الداخلية في القطر العربي السوري مهتمة بمعرفة نسبة السائقين الشباب الذين يضعون بانتظام حزام الأمان ، في مسح إحصائي شمل 300 سائق شاب ، أجاب 123 منهم بأنهم يستعملون الحزام بانتظام ، هل يمكننا أن نستنتج من هذه البيانات بأن نسبة من يستخدمون الحزام بانتظام في المجتمع المعاين لا تساوي 0.50 ؟

الحل :

البيانات تتألف من إجابات 300 سائق شاب ، 123 منهم يستخدم الحزام بانتظام أي أن :

$$\hat{p} = \frac{123}{300} = 0.41$$

إن توزيع المعاينة لهذه النسبة يتبع تقريبا التوزيع الطبيعي إذا كانت فرضية العدم $H_0 : P = 0.50$ صحيحة . كما أن الخطأ المعياري يحسب بالعلاقة :

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{0.5(1-0.5)}{300}}$$

لقد استخدمنا القيمة الافتراضية لـ p عند حساب الخطأ المعياري لأن الاختبار بكامله مبني على صحة فرضية العدم ، واستخدام نسبة العينة بحساب الخطأ المعياري $\sigma_{\hat{p}}$ لن يكون متسقاً مع هذا المفهوم .

– **الفرضيات** : هي اختبار فرضية العدم القائلة بأن نسبة الشباب الذين يستخدمون الحزام بانتظام في المجتمع المعاين تساوي ٠,٥٠. مقابل الفرضية البديلة التي تقول بأن نسبة الشباب الذين يستخدمون الحزام بانتظام في المجتمع المعاين لا تساوي 0.50 أي :

$$H_0 : P = 0.50$$
$$H_1 : p \neq 0.50$$

إن الفرض البديل في هذه الحالة هو اختبار من طرفين .

– إحصائية الاختبار z تقترب من التوزيع الطبيعي المعياري ، وتحسب كما يلي :

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 . q_0 / n}} = \frac{0.41 - 0.5}{\sqrt{0.5(0.5)/300}} = -3.11$$

عند مستوى معنوية $\alpha = 0.05$ والاختبار من طرفين فإن القيمة الحرجة لـ z تساوي ± 1.96 . القرار هو رفض فرضية العدم H_0 لوقوع قيمة إحصائية الاختبار 3.11- في منطقة الرفض ، وبالتالي نستنتج بأن نسبة من يستخدمون بانتظام حزام الأمان في المجتمع لا تساوي ٠,٥٠ .

مثال (5–14) :

يدّعي مصنع للأدوية المسجلة أن دواء من إنتاجه له فاعلية بنسبة 90 % في التخفيف من الحساسية لفترة 8ساعات في عينة من 200 شخصاً مصاباً بالحساسية ، أدى الدواء إلى تخفيف آلام 160 منهم ، قرر فيما إذا كان إدعاء المصنع صحيح أم غير صحيح عند مستوى معنوية $\alpha = 0.01$.

الحل:

نكوّن الفرضية الأولية والبديلة كالتالي :

$H_0 : P = 0.90$ (الإدعاء صحيح)

$H_1 : P < 0.90$ (الفرض البديل من الطرف الأيسر ، الإدعاء باطل)

- إحصائية الاختبار z تقترب من التوزيع الطبيعي المعيــاري ، إذا كانــت H_0 صحيحة ، وتحسب كما يلي :

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 . q_0 / n}} = \frac{0.80 - 0.90}{\sqrt{0.9(0.1)/200}} = -4.73$$

عند مستوى معنوية $\alpha = 0.01$ والاختبار من طرف أيسر فإن القيمة الحرجة لـــ z تساوي -2.33 .

القرار هو رفض فرضية العدم H_0 لوقوع 4.73- في منطقة الـــرفض . وبالتـــالي يكون إدعاء مصنع الأدوية غير صحيح .

مثال (5–15) :

لنفترض بأن إحدى الصحف المحلية قد نشرت بأن نسبة البطالة (العاطلين عن العمل) بين الشباب في سورية يعادل 14 % ، وللتحقق من صحة هذه النسبة سحبت عينة عشوائية تضم 100 شاب فتبين أن 20 شاباً منهم لا يعمل .

المطلوب:

قرر فيما إذا كان إدعاء الصحيفة صحيحاً أم غير صحيح بخصوص نسبة البطالة عنـــد مستوى معنوية $\alpha = 0.05$.

الحل :

الفرضيات : هي اختبار فرضية العدم القائلة بأن نسبة الشباب العاطلين عن العمـــل في المجتمع المعاين تساوي 0.14 مقابل الفرضية البديلة، التي تقول بأن نسبة الشباب الذين لا يعملون في المجتمع المعاين هي مختلفة عن ذلك .

$$H_0 : P = 0.14$$
$$H_1 : p \neq 0.14$$

إن الفرض البديل في هذه الحالة هو اختبار من طرفين ويعني بأن الادعاء باطل.

- إحصائية الاختبار z تقترب من التوزيع الطبيعي المعياري ، إذا كانت H_0 صحيحة ، وتحسب كما يلي :

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 . q_0 / n}} = \frac{0.20 - 0.14}{\sqrt{0.14(0.86)/100}} =$$
$$= \frac{0.06}{0.0347} = 1.73$$

عند مستوى معنوية $\alpha = 0.05$ والاختبار من طرفين فإن القيمة الحرجة لـ z تساوي 1.96±. القرار هو قبول فرضية العدم H_0 لوقوع القيمة المحسوبة لإحصائية الاختبار 1.73 في منطقة القبول . وبالتالي يكون إدعاء الصحيفة صحيحاً ، أي لا توجد فروق حقيقية بين النسبة الحقيقية للعاطلين عن العمل في المجتمع والنسبة المنشورة في الصحيفة المحلية ، أو أن النتائج التي زودتنا بها العينة غير كافية لرفض فرضية العدم .

§-5 اختبار الفرضيات للفرق بين نسبتي مجتمعين :

يستخدم اختبار الفرضيات للفرق بين نسبتي مجتمعين $P_1 - P_2$ لمعرفة ما إذا كان الفرق يرجع إلى عامل الصدفة أم أن هناك فروقاً حقيقية وذلك من خلال بيانات عينتين عشوائيتين مسحوبتين من المجتمعين قيد الدراسة ؟ مثال نسبة التدخين لعينتين من الطلاب في جامعتين مختلفين، ونسبة الشفاء من مرض معين باستخدام عقارين مختلفين لعينتين من المرضى ... الخ .

إن اختبار الفرضيات للفرق بين نسبة مجتمعين الأكثر استخداما هو أن يكون الفرق مساويا إلى الصفر، مع العلم أن هناك إمكانية لكون الفرق مساويا لقيمة أخرى، سواء كان الاختبار من طرف واحد أو من طرفين ، عندما نختبر فرضية العدم :

$$H_0 = P_1 - P_2$$

فإننا نفترض تساوي نسبة المجتمعين، ونستخدم هذه الحقيقة لدمج نتائج العينتين في تقدير مجمع للنسبة الافتراضية المشتركة. وإذا اتبعنا هذه الطريقة نحسب:

$$\tilde{p} = \frac{x_1 + x_2}{n_1 + n_2} \qquad (5-7)$$

حيث :

x_1 , x_2 عدد حالات تحقق الخاصية المدروسة في العينتين ، وهذا التقدير المجمع لـــ $P = p_1 = p_2$ يستخدم لحساب الخطأ المعياري المقدر:

$$\hat{\sigma}_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n_1} + \frac{\tilde{p}(1-\tilde{p})}{n_2}} \qquad (5-8)$$

إن إحصائية الاختبار:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)_0}{\hat{\sigma}_{\hat{p}_1 - \hat{p}_2}} \qquad (5-9)$$

تتبع توزيع طبيعي معياري إذا كانت H_0 صحيحة .

مثال (5-16) :

في دراسة أكاديمية مخصصة لمقارنة نتائج دورة تدريبية حديثة في تقانات المعلوماتية مع نتائج دورة تدريبية نمطية معروفة ، كانت نتائج الاختبار لـ 78 شخصاً من أصل 100 شخص اتبعوا الدورة النمطية إيجابياً ، في حين كانت نتائج الاختبار لـ 90 شخصاً من أصل 100 شخص اتبعوا الدورة الحديثة إيجابياً ، هل هذه المعطيات كافية لتشير بأن الدورة الحديثة هي أكثر فعالية من الدورة النمطية وذلك عند مستوى معنوية 0.05 ؟

الحل :

تتألف البيانات من نتائج عينتين عشوائيتين تضم كل منها 100 شخص اتبعوا نوعين

من الدورات فكانت النتائج التالية :

$$\hat{p}_1 = 0.78 \quad , \quad \hat{p}_2 = 0.90 \quad , \tilde{p} = \frac{90+78}{100+100} = 0.84$$

إن توزيع المعاينة للفرق $\hat{p}_1 - \hat{p}_2$ يقترب من التوزيع الطبيعي بمتوسط: $p_2 - p_1 = 0$
وخطأ معياري مقدر يساوي :

$$\hat{\sigma}_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n_1} + \frac{\tilde{p}(1-\tilde{p})}{n_2}}$$

إذا كانت فرضية العدم صحيحة وتقديرات العينة مرجحة .

- الفرضيات : نكوّن الفرضية الأولية والبديلة كالتالي :

$$H_0 : p_2 - p_1 = 0$$
$$H_A : p_2 - p > 0$$

والفرض البديل بهذه الحالة هو من طرف واحد أيمن .

- إحصائية الاختبار z تقترب من التوزيع الطبيعي المعياري ، وتحسب كما يلي :

$$z = \frac{(0.90 - 0.78)}{\sqrt{(0.84)(0.16)/100 + (0.84)(0.16)/100}}$$
$$= 0.12/0.518 = 2.32$$

عند مستوى معنوية $\alpha = 0.05$ والاختبار من طرف واحد فإن القيمة الحرجة لـــــ z تساوي ١.645 . القرار هو رفض فرضية العـــدم H_0 لوقـــوع ٢.32 في منطقـــة الرفض . وبالتالي نستنتج بأن العلاج الجديد أكثر فعالية من العلاج النمطي .

مثال (5–17):

السؤال هل يوافق أو يحبذ الأمريكيون الطريقة التي يتصرف بموجبها (يعالج بها بيل

كلينتون عمله كرئيس) رئيسهم في عمله ؟ في استقصاء أجرته جريدة CBS/ نيويورك تايمز New York Times/CBS News في شهر آب من عام 1998 م ، شمل 92 مواطن أمريكي أسود و 788 مواطن أمريكي أبيض ، فكانت النتيجة أن 0.94 من الأمريكيين السود يوافقون بيل كلينتون على الطريقة التي يعالج بها عمله كرئيس ، بينما كانت النتيجة هي أن 0.60 فقط من الأمريكيين البيض يوافقون بيل كلينتون على الطريقة التي يعالج بها عمله كرئيس ، أي أن :

$$\hat{p}_1 = 0.94 \quad , \quad \hat{p}_2 = 0.60$$

والمطلوب:

اختبار معنوية الفرق بين نسبتي العينتين، مفترضين أن العينتين عشوائيتين ومستقلتين ومسحوبتين من مجتمعين طبيعيين.

الحل :

الافتراضات هي : استخدام عينتين عشوائيتين ، مستقلتين وبحجم $n_1 + n_2 > 100$ ، مستوى المعنوية ألفا 0.05 .

الفرضيات:

- الفرضية الأولية (العدم) تقول بعدم وجود فروق ذات دلالة إحصائية ما بين وجهة نظر الأمريكيين السود ووجهة نظر الأمريكيين البيض فيما يتعلق بأداء بيل كلينتون لعمله كرئيس .
- الفرضية البديلة القائلة بوجود فرق معنوي بين النسبتين:

$$H_0 : P_1 = P_2$$
$$H_1 : P_1 \neq P_2$$

وكما نعلم[8] بأن توزيع المعاينة للفرق بين نسبة عينتين هو تقريباً طبيعي عندما يكون حجم العينة كبيراً ($n_1 + n_2 > 100$) .بمتوسط يساوي :

[8] **Chava Frankfort – Nachmais & Anna Leon- Guerrero** " Social statistics for a diverse Society " Pine Forge Press, A Sage Publications Company, Thousand Oaks, California, Second Edition 2000, pp 497.

$$\mu_{p_1-p_2} = P_1 - P_2$$

وبخطأ معياري مقدر Estimated Standard Error يساوي :

$$S_{p_1-p_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

إحصائية الاختبار Z تقترب من التوزيع الطبيعي المعياري ، وتحسب بالعلاقة التالية :

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{S_{p_1-p_2}} = \frac{0.94 - 0.60}{0.03} = 11.33$$

حيث :

$$S_{p_1-p_2} = \sqrt{\frac{0.94(1-0.94)}{92} + \frac{0.60(1-0.60)}{788}} = 0.03$$

عند مستوى معنوية $\alpha = 0.05$ والاختبار من طرفين فإن القيمة الحرجة لـــ z تساوي ±1.96 . إن القرار هو رفض H_0 لوقوع قيمة إحصائية الاختبار 11.33 في منطقة الرفض ، لأنها أكبر من القيمة الحرجة لـــ z والتي تساوي ±1.96 .

وبالتالي نستنتج بأن هناك فرق ذي دلالة إحصائية بين نسبة الأمريكيين السود ونسبة الأمريكيين البيض فيما يتعلق برأيهم حول الطريقة التي يمارس بها بيل كلينتون عمله كرئيس للولايات المتحدة الأمريكية.

مثال (5–18):

لمقارنة نسبة الرجال والنساء المؤيدين لعملية الإجهاض لأي سبب كان ، تم انتقاء عينة عشوائية جزئية من المسح العام (استقصاء) الذي جرى عام 1996 والخاص بولاية كاليفورنيا فكانت النتائج كما هي موضحة بالجدول.

والمطلوب:

اختبار معنوية الفرق بين نسبة العينتين ، مفترضين أن العينتين عشوائيتين ومستقلتين .

النساء	الرجال	البيان
$\hat{p}_2 = 0.44$	$\hat{p}_1 = 0.46$	نسبة الموافقين
$n_2 = 505$	$n_1 = 399$	حجم العينة

الحل :

الافتراضات : استخدام عينتين عشوائيتين ، مستقلتين وبحجم $n_1 + n_2 > 100$.

– **الفرضيات :** سنفترض أن الاختبار من طرفين Two-tailed test أي أن الفرق بين نسبة المجتمعين (الرجال والنساء الأمريكيين) فيما يتعلق بوجهة نظرهم حول عملية الإجهاض غير متساوية .

– الفرضية الأولية تقول بعدم وجود فروق ذات دلالة إحصائية ما بين وجهتي نظر الرجال الأمريكيين والنساء الأمريكيات فيما يتعلق بعملية الإجهاض.

– الفرضية البديلة القائلة بوجود فرق معنوي بين هاتين النسبتين .

$$H_0 : P_1 = P_2$$
$$H_1 : P_1 \neq P_2$$

إن توزيع المعاينة للفرق بين نسبتي عينتين هو تقريباً طبيعي عندما يكون حجم العينة كبيراً ($n_1 + n_2 > 100$). بمتوسط يساوي : $P_1 - P_2$ ، وبخطأ معياري مقدر يساوي :

$$S_{p_1 - p_2} = \sqrt{\frac{0.46(1-0.46)}{399} + \frac{0.44(1-0.44)}{505}} = 0.03$$

إحصائية الاختبار Z تقترب من التوزيع الطبيعي المعياري ، وتحسب بالعلاقة التالية :

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{S_{p_1 - p_2}} = \frac{0.46 - 0.44}{0.03} = 0.67$$

عند $\alpha = 0.01$ والاختبار من طرفين ، فإن القيمة الحرجة لـ z تساوي ± 2.58 ، القرار هو قبول الفرضية H_0 لوقوع قيمة إحصائية الاختبار 0.67 في منطقة القبول

، لأنها أصغر من القيمة الحرجة لـــ z والتي تساوي ± 2.58 . وبالتالي نستنتج بعدم وجود فرق ذي دلالة إحصائية بين نسبة الرجال الأمريكيين ونسبة النساء الأمريكيات فيما يتعلق برأيهم ووجهة نظرهم حول عملية الإجهاض لأي سبب كـــان. كمـــا أن الفرق المشاهد بين وجهتي نظر الرجال والنساء والمحسوب من واقع بيانـــات المســـح الحقيقي من المحتمل أنه لا يعكس الفرق الحقيقي بين وجهتي نظر الرجال والنســـاء في المجتمع ككل .

§-6 اختبار الفرضيات حول تباين المجتمع :

سنقتصر في مناقشتنا هنا على اختبار الفرضيات حول تباين المجتمـــع عنـــدما تكون البيانات المتاحة متأتية من عينة عشوائية بسيطة مسحوبة من مجتمع طبيعي .

إن إحصائية الاختبار المستخدمة لاختبار تباين المجتمع ستكون :

$$\chi^2 = (n-1)s^2 / \sigma^2 \qquad (5-10)$$

وتتبع ، عندما تكون فرضية العدم صحيحة، توزيع كاي تربيع مع $n-1$ درجة حرية.

مثال (5-19) :

تضم عينة عشوائية بسيطة 15 مشرفاً اجتماعياً يشتركون في دورة تدريبيـــة متقدمـــة لاكتساب مهارات وتقانات معلوماتية، خضعوا لاختبار بهـــدف قيـــاس المهـــارات والتقانات التي اكتسبوها خلال هذه الدورة . لقد كان تباين مشاهدات العينة مساو إلى 1225 درجة ، هل نستطيع أن نستنتج من هذه البيانات بأن تباين المجتمـــع هـــو مختلف عن 2500 ؟

الحل:

تتألف البيانات من معدلات الاختبار في المهارات والتقانات المكتسبة لـــعينة عشوائية تضم 15 مشرفاً اجتماعياً مسحوبة من مجتمع طبيعي، لقد كان تباين العينة العشوائية :

$s^2 = 1225$

- **الفرضيات :**

$$H_0 : \sigma^2 = 2500$$
$$H_1 : \sigma^2 \neq 2500$$

والفرض البديل في هذه الحالة من طرفين .

- **إحصائية الاختبار:**

إن إحصائية الاختبار $\chi^2 = (n-1)s^2 / \sigma^2$

تتبع ، عندما تكون H_0 صحيحة ، توزيع كاي تربيع مع $n-1$ درجة حرية .

وبحسابها نجد :

$$\chi^2 = (n-1)s^2 / \sigma^2 = 14(1225) / 2500 = 6.86$$

- بفرض أن $\alpha = 0.05$ والاختبار من طرفين فإن القيم الحرجة لـــ χ^2 تساوي 5.629 و 26.119 .

- القرار هو رفض H_0 باستثناء حالة وقوع قيمة إحصائية الاختبار بين القيمتين الحرجتين الصغرى والكبرى . وبما أن :

$$\chi^2_{(0.025,14)} = 5.629$$

$$\chi^2_{(0.975,14)} = 26.119$$

- ونظراً لوقوع قيمة إحصائية الاختبار بين القيمتين الحرجتين الصغرى والكبرى :

$$5.629 < 6.86 < 26.119$$

فإننا لا نستطيع رفض H_0 . وبالتالي واستنادا لهذه البيانات لا يمكننا أن نستنتج بأن تباين المجتمع الإحصائي هو مختلف عن 2500 .

تمارين غير محلولة

١) هل يمكننا أن نستنتج ، أن متوسط السعرات الحرارية التي يحصل عليها الشخص في منطقة حضرية بدولة متقدمة هو أقل من 2000 عند مستوى معنوية 0.05 ؟ علماً أنه تم سحب عينة عشوائية تضم 500 شخص من هذا المجتمع فكان متوسط السعرات الحرارية لأفراد العينة 1985 وانحرافها المعياري 210 .

٢) اختيرت عينة عشوائية تضم 25 طفلاً مولودين حديثاً في مشافي حلب، فكان الانحراف المعياري لأوزان الأطفال في العينة مساويا إلى 150 غرام . هل هذه البيانات كافية للتدليل بأن تباين المجتمع هو أكبر من 10000 عند مستوى معنوية 0.05 ؟

٣) في استطلاع للرأي حول الإنفاق الشهري على المواد الغذائية في محافظة حلب ، تم سحب عينتين عشوائيتين الأولى عمالية والثانية من ذوي الدخل المحدود فتبين ما يلي :

$$n_1 = 327 \quad , \quad \bar{x}_1 = 6000 \quad , \quad s_1 = 1040$$
$$n_2 = 286 \quad , \quad \bar{x}_2 = 6400 \quad , \quad s_2 = 1180$$

والمطلوب : هل يمكننا أن نستنتج بأن أسر أصحاب الدخل المحدود تنفق على المواد الغذائية أكثر من الأسر العمالية وذلك ضمن مستوى معنوية 0.05 ؟

٤) ليكن معلوماً أن الدخل الشهري للعاملين في مجال المعلوماتية في ثانويات محافظة حلب يخضع للتوزيع الطبيعي بمتوسط قدره 8500 ل.س ، وتباين غير معلوم . تم انتقاء عينة عشوائية تضم 9 أساتذة من خريجي الهندسة المعلوماتية فكانت دخولهم الشهرية بآلاف الليرات السورية كما يلي :

5 ، 6 ، 7 ، 7 ، 8 ، 8 ، 10 ، 10 ، 11

والمطلوب :

حساب الوسط الحسابي ، التباين ، الانحراف المعياري للدخول الشهرية.

- اختبار الفرض بأن متوسط المجتمع يختلف عن 8500 ل.س وذلك عند مستوى معنوية 0.05 ؟ بناء مجال ثقة لمتوسط المجتمع بدرجة ثقة 0.95 ؟

٥) بلغ متوسط دخل الأسرة في مدينة حمص 5000 ل.س ، وللتأكد من صحة هذه المعلومة أو الادعاء قمنا بسحب عينة عشوائية مؤلفة من 35 عائلة فوجدنا أن متوسط دخولهم هو 4660 ل.س وأن تباين هذه الدخول يساوي 1465 ل.س (تباين مقدر للمجتمع ، حيث يستعاض عن تباين المجتمع بتباين العينة) .

والمطلوب : اختبار فيما إذا كان متوسط الدخل للعائلات في مدينة حمص هو مختلف عن 5000 أم لا ، وذلك عند مستوى معنوية 0.05 وأيضاً 0.01 ؟

٦) يقال بأن متوسط سعر الكغ الواحد من البرتقال قد بلغ في شهر أيار بمدينة حلب خلال العام الماضي 15 ل.س . تم أخذ مشاهدات متتالية بأيام مختلفة ولفترة 15 يوم للتعرف على متوسط أسعار البيع في المدينة فكانت النتائج التالية:

15 – 13-14 – 17 -15–12 – 18–13 – 12 – 20 - 13

13 – 12 – 20 - 13

والمطلوب: اختبار فيما إذا كان متوسط سعر البرتقال هو مختلف عن 15 ل.س وذلك عند مستوى معنوية 0.05 ؟

٧) أجريت تجربة على 15 شخصاً من رواد الفضاء ، في مجال يحاكي مجال انعدام الوزن وجد أن متوسط زيادة ضربات القلب لديهم 25ضربة / دقيقة ، بانحراف معياري قدره 5ضربة / الدقيقة ، احسب أقصى خطأ في تقدير المتوسط عند مستوى معنوية 0.01 ؟

٨) لتحديد نسبة ذوي الضغط العالي من بين الأشخاص البالغين ، كان المطلوب هو تحديد حجم العينة n حتى يمكننا التأكد باحتمال . 99. من أن الخطأ لا يتجاوز 0.05 إذا علمنا أن قيمة النسبة P في المجتمع تساوي 0.20