المحاضرة الخامسة

الإحصاء الحيوي

السنة الثالثة صيدلة

د. وديع علي

العينات والتقدير وفترات الثقة

Samples, Estimation and Confidence Intervals

تعتبر طريقة العينات من الطرائق الإحصائية المستخدمة في عمليات جمع المعلومات الإحصائية وتحليلها، وذلك لأنها توفر كثيراً من المال والجهد والوقت في الحصول على المعلومات اللازمة.

وتعد طريقة العينات في كثير من الأحيان الأسلوب الإحصائي الوحيد الذي يمكن تطبيقه في الحصول على المعلومات، كما في حالة تحليل الدم أو تحليل التربة، أو تحليل المياه، أو مراقبة الإنتاج، أو تقدير صلاحية الأغذية أو الأدوية...إلخ.

أنواع العينات:

تصنّف العينات حسب تصميم عمليات المعاينة , والذي يعتمد على طبيعة المجتمع المدروس وعلى هدف الدراسة... وأنواعها هي:

1. **المعاينة العشوائية البسيطة:** وهي التي يتم فيها سحب عينة عشوائية من عناصر المجتمع المدروس وباحتمالات متساوية وتطبق على المجتمعات المتجانسة.
2. **المعاينة العشوائية الطبقية:** وهي التي يتم فيها تقسيم المجتمع المدروس غير المتجانس إلى طبقات متجانسة تم تطبيق المعاينة البسيطة على كل طبقة على حده، ثم يتم تركيب النتائج وإجراء الحسابات اللازمة.
3. **المعاينة العنقودية:** وهي التي يتم فيها سحب العينات العنقودية على مرحلتين أو أكثر: تُسحب عينة فروع من الجذع ثم عينات عناصر من الفروع (عينات من المدن، ثم عينات من الأسر في كل مدينة مسحوبة).
4. **المعاينة المنتظمة :** وهي التي يتم فيها سحب عناصر العينة بطريقة عددية منتظمة, كأن نسحب عنصراً بعد مرور كلk عنصراّ.

أما خطوات البحث المتبعة في طريقة العينات فهي نفسها الخطوات المتبعة في حالة المسح الشامل. غير أن طريقة العينات تحتاج إلى خطوة إضافية وهامة هي خطوة سحب العينة، وهو ما سنعالجه فيما يأتي،

كيفية سحب العينة العشوائية البسيطة:

يُشترك في سحب العينة أن يكون سحباً عشوائياً ولذلك نقوم بما يلي:

1. نحصر عناصر المجتمع المدروس ونحدد عددها N.
2. نعطي كل عنصر رقماً متسلسلاً من 1 حتى N.
3. نضع أرقام هذه العناصر في صندوق خاص ونخلطها جيداً.

4. نقوم بسحب الأرقام من الصندوق بطريقة عشوائية ونخلط الأرقام قبل كل عملية سحب.

أما السحب بحد ذاته فيمكن أن يكون على شكلين هما:

.السحب مع الإعادة: وفيه يُعاد الرقم المسحوب إلى الصندوق ليتعرض للسحب مرةً ثانية أو ثالثة ...الخ.

. السحب بدون إعادة، وفيه لا يُعاد الرقم المسحوب إلى الصندوق وبذلك لا يتعرض للسحب مرةً ثانية.

ويمكننا سحب الأرقام عشوائياً بواسطة استخدام عجلات الرهان، أو بواسطة جداول الأرقام العشوائية، أو بواسطة علاقات رياضية مبرمجة على الآلات الحاسبة بمفتاح Run. وسنقتصر في بحثنا على العينات الكبيرة المسحوبة مع الإعادة لسهولة التعامل معها.

5. نعتبر الأرقام المسحوبة في الخطوة السابقة هي أرقام العناصر التي ستدخل في العينة، ونبحث عنها في المجتمع لدراستها واستخلاص المعلومات اللازمة منها.

إن عملية السحب العشوائي لعناصر العينة تضمن لنا عدم التحيّز في الانتقاء، كما تؤمّن تمثيلاً جيداً لجميع عناصر المجتمع المدروس.

أما بالنسبة لحجم العينة اللازم سحبه في دراسة معينة فيتوقف على مقدار تجانس المجتمع بالنسبة للخاصة المدروسة، وعلى مقدار الدقة المطلوبة d وعلى معامل الثقة المفروض z وعلى مقادير أخرى كالتكلفة وطريقة السحب. وسنعتمد على أبسط العلاقات الرياضية التي تعطينا حجم العينة n باحتمال ثقة قدره (0.95) وهي:

$$n = \frac{4\sigma^2}{d^2}$$

حيث أن σ^2 هو تباين الخاصة المدروسة في المجتمع ويتم تقديره من خلال تباين عينة اختباريه. وأن d هي الدقة المطلوبة في التقدير وتُحدد مسبقاً من قبل المسؤولين عن البحث.

وقبل أن نبحث في قضايا التقدير نورد بعض التعاريف المتعلقة بالمجتمع والعينة والمؤشرات المستخدمة في تقدير معالم المجتمع.

إن عملية استنتاج معلمات المجتمع عن طريق دراسة خواص العينات تسمى بالاستدلال الإحصائي والذي يمكن تقسيمه إلى نوعين:

- التقدير الإحصائي
- اختبار الفرضيات الإحصائي

ومن التعريفات الهامة التي يقوم عليها الاستدلال الإحصائي تعريفي المعلمة والإحصاء.

المعلمة (Parameter) عبارة عن خاصية أو مقياس يتم حسابها من المجتمع محل الدراسة.

مثال:

- نسبة البطالة في سورية
- متوسط العمر الافتراضي لجهاز معين.

أي أن المعالم هي مقاييس تحدد خصائص المجتمع (التوزيع)

الإحصاء (Statistic) عبارة عن خاصية أو مقياس ينم حسابه من العينة المسحوبة من المجتمع محل الدراسة.

التقدير (Estimation) أسلوب إحصائي يستخدم لتقدير معلمة مجهولة عن طريق استخدام مقاييس العينة.

يوجد أسلوبين لتقدير معلمة المجتمع المجهولة وهما:

التقدير بنقطة:

تستخدم بيانات العينة لتقدير معلمة المجتمع المجهولة بنقطة واحدة فقط، أي بقيمة واحدة فقط.

التقدير بفترة:

تستخدم بيانات العينة لتقدير معلمة المجتمع المجهولة بفترة لها حد أدنى وحد أعلى.

من أهم معالم المجتمع متوسط المجتمع μ :

نستخدم المتوسط الحسابي $\bar{x}$ للعينة كتقدير نقطة (Point Estimation) لمتوسط المجتمع μ أي أن:

$$\tilde{\mu} = \bar{x}$$

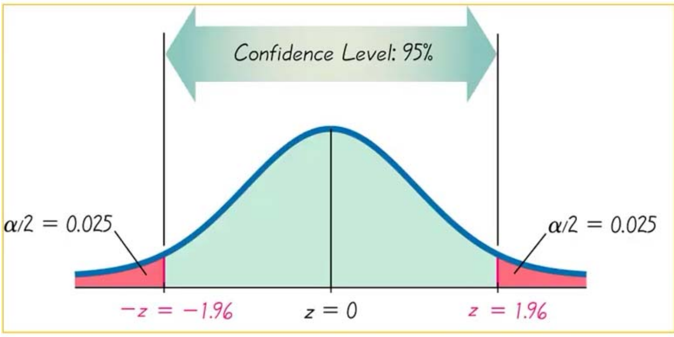
ومن المتوقع أن لا تكون قيمة متوسط العينة مساوية لقيمة متوسط المجتمع، ولكن قريبة منها، ويسمى الفرق المطلق بين قيمة الإحصاء وقيمة المعلمة الفعلية المراد تقديرها بالخطأ المعياري.

الخطأ المعياري (Standard Error) للمتوسط: هو الانحراف المعياري لتوزيع متوسطات العينات. بمعنى انحراف متوسطات العينات عن متوسط مجتمعها ويرمز له بالرمز $\sigma_{\bar{x}}$ حيث:

$$\tilde{\sigma}_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

التقدير بفترة لمتوسط المجتمع:

ذكرنا سابقاً أن التقدير بنقطة نادراً ما يساوي المعلمة التي نرغب في تقديرها، لذلك نلجأ إلى التقدير بفترة تحتوي على قيمة معلمة المجتمع، وتسمى هذه الفترة بتقدير فترة الثقة، واحتمال وقوع المعلمة في هذه الفترة يسمى درجة الثقة ويرمز لدرجة الثقة بـ $1-\alpha$ ومكمل هذه القيمة يسمى مستوى الدلالة ويرمز له بالرمز α. فمثلاً إذا كانت درجة الثقة 95% فإن مستوى الدلالة $\alpha = 0.05$.



الحالة الأولى:

لتقدير فترة ثقة لمتوسط المجتمع μ نستخدم متوسط العينة $\bar{x}$ وتستخدم العلاقة الآتية:

$$\bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

حيث تتحدد قيمة $Z_{\frac{\alpha}{2}}$ حسب درجة الثقة (أو مستوى الدلالة) كما يلي:

- إذا كانت درجة الثقة 90 % فإن $Z_{\frac{\alpha}{2}} = 1.65$
- إذا كانت درجة الثقة 95 % فإن $Z_{\frac{\alpha}{2}} = 1.96$
- إذا كانت درجة الثقة 99 % فإن $Z_{\frac{\alpha}{2}} = 2.575 \approx 2.58$.

مثال:

عينة عشوائية حجمها $n = 25$. أخذت من مجتمع طبيعي انحرافه المعياري $\sigma = 4$ فأعطت المعدل $\bar{x} = 60$

. أوجد فترة ثقة 99% لمتوسط المجتمع μ.

الحل:

$$1-\alpha = 0.99 \,,\ \alpha = 1\% \,,\ \frac{\alpha}{2} = 0.005$$

$$\left[\bar{x} - Z_{0.005}\frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{0.005}\frac{\sigma}{\sqrt{n}}\right] = \left[60 - 2.575\frac{4}{\sqrt{25}}, 60 + 2.575\frac{4}{\sqrt{25}}\right]$$

$$57.94 \le \mu \le 62.06$$

الحالة الثانية: إذا كان تباين المجتمع σ^2 مجهولاً وحجم العينة كبيراً $(n > 30)$ نستخدم الفترة في الحالة الأولى مع استخدام قيمة s^2 تباين العينة بدلاً من تباين المجتمع σ^2 المجهول ويكون شكل الفترة:

$$\bar{x} - Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

مثال:

تسبب المقاعد الخالية لشركات الطيران خسارة لجزء من الدخل، بفرض أن إحدى شركات الطيران الكبرى أرادت تقدير عدد المقاعد الخالية لكل رحلة خلال العام الماضين فقامت باختيار عينة عشوائي من 225 رحلة طيران وتسجيل عدد المقاعد الخالية في كل رحلة من هذه العينة وكان المتوسط الحسابي والانحراف المعياري لعدد المقاعد الخالية في هذه العينة هما: $\bar{x} = 11.6$ و $s = 4.1$ مقعد . قدر عدد المقاعد الخالية خلال العام الماضي باستخدام فترة ثقة 95% .

الحل:

$$11.6 - 1.96 \frac{4.1}{\sqrt{225}} \leq \mu \leq 11.6 + 1.96 \frac{4.1}{\sqrt{225}}$$

$$11.15 \leq \mu \leq 12.05$$

أي أن متوسط المقاعد الخالية خلال العام الماضي يقع داخل الفترة $[11.15, 12.05]$ وذلك بدرجة ثقة 95% .

الحالة الثالثة: فترة $100(1-\alpha)\%$ ثقة للمتوسط μ إذا كان حجم العينة صغيراً $(n \leq 30)$ مسحوبة من مجتمع طبيعي تباينه غير معلوم:

$$\bar{x} - t_{\left(\frac{\alpha}{2}, n-1\right)} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\left(\frac{\alpha}{2}, n-1\right)} \frac{s}{\sqrt{n}}$$

حيث s الانحراف المعياري للعينة.

ملاحظة: إذا كانت قيمة المعلمة σ مجهولة نستخدم قيمة الانحراف المعياري للعينة s بدلاً منها كتقدير لها.

مثال:

مصنع لإنتاج المصابيح الكهربائية، اختير من إنتاجه عينة حجمها 100 مصباح لتقييم جودة الإنتاج، حيث حدد مدير المصنع معيار الجودة أن يتراوح عمر المصباح بين 1000 و 1100 يوم. إذا كان المتوسط الحسابي لعمر المصباح في العينة المختارة 1200 يوم وانحرافه المعياري 250 يوم. قدر فترة ثقة 95% لمتوسط عمر المصباح من إنتاج المصنع كله.

الحل:

حيث إن الانحراف المعياري للمجتمع مجهول نستخدم الانحراف المعياري للعينة S ويتم استخدام توزيع Z نظراً لأن حجم العينة كبير .

$$n = 100 \ , \ \bar{x} = 1200 \ , \ s = 250 \ , 1-\alpha = 0.95 \ , \ \alpha = 0.05$$

$$\frac{\alpha}{2} = 0.025 \ , \ z_{0.025} = 1.96.$$

$$1200 - 1.96\frac{250}{\sqrt{100}} \le \mu \le 1200 + 1.96\frac{250}{\sqrt{100}}$$

$$1151 \le \mu \le 1249.$$

ويدل هذا المصابيح عالية في هذا المصنع حيث إن متوسط عمر المصباح من إنتاج المصنع كله يتراوح بين 1151 و 1249 يوم .

مثال:

أخذت عينة عشوائية 15 من مجتمع طبيعي فأعطت $\bar{x} = 17.4$ و $s = 2.1$. أوجد فترة ثقة 95% للمتوسط الحسابي.

الحل:

$$1-\alpha = 95\% \ , \ \alpha = 5\% \ , \ \frac{\alpha}{2} = 0.025$$

$$n-1 = 14 \ , \ t_{(0.025,14)} = 2.145.$$

$$\left[17.4 - 2.145\frac{2.1}{\sqrt{15}}, 17.4 + 2.145\frac{2.1}{\sqrt{15}}\right] = [16.24, 18.56]$$

ملخص:

- إذا كانت σ معلومة نستخدم توزيع Z
- إذا كانت σ غير معلومة:

أ) نستخدم توزيع Z عندما يكون $n > 30$

ب) نستخدم توزيع t عندما يكون $n \ge 30$.