المحاضرة الثانية: توزيعات المعاينة (يتبع)

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

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

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

للإجابة على هذا السؤال يجب التطرق إلى ما يلي:

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

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

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

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

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

$$E(\overline{X}) = \mu_{\overline{X}} = \mu \quad \ldots\ldots\ (10)$$

$$Var(\overline{X}) = \sigma^2_{\overline{X}} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots (11)$$

ومنه فالإنحراف المعياري للمتغير $\overline{X}$ هو:

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

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

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

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

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

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

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

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

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

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

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

يخضع للتوزيع الطبيعي المعياري وذلك مهما كان حجم العينة.

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

المطلوب:

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

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

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

الحل:

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

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

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

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

2- إيجاد الاحتمال التالي: $P(\overline{X} > 3100)$

نقوم أولا بحساب المتغير العشوائي Z:

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

ومنه نجد:

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

$$= 0.5 - 0.3413 = 0.1587$$

3- إيجاد الاحتمال التالي:

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

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

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

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

$$= 0.4332 + 0.3413 = 0.7745$$

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

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

للإجابة عن هذا السؤال يجب التطرق إلى النظرية التالية:

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

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

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

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

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

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

المطلوب:

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

الحل:

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

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

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

نقوم أولا بحساب المتغير العشوائي Z:

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

ومنه يكون:

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

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

لدينا: $\% 4 \text{ x } 75 = 3$

إذن المراد هو حساب الاحتمال التالي:

$$P(72 < \overline{X} < 78) = P\left(\frac{72-75}{1.82} < Z < \frac{78-75}{1.82}\right)$$
$$= P(-1.65 < Z < 1.65)$$
$$= 2P(0 < Z \le 1.65)$$
$$= 0.901$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

المطلوب:

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

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

الحل:

لدينا:

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

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

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

1- حساب الاحتمال التالي: $P(\overline{X} \leq 43.5)$

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

حيث:

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

ومنه:

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

وبالتالي نجد:

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

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