المحاضرة السادسة / الفصل الثاني
"التقدير الإحصائي"

1- مقدمة:

تعتمد التوزيعات الإحصائية على معالم محددة ،فمثلا يعتمد التوزيع **ذي الحدين** على المعلمة P (نسبة النجاح في التجربة)، ويعتمد توزيع **بواسون** على معلمة محددة تسمى بالوسط الحسابي وهي λ، ويعتمد التوزيع **الطبيعي** على المعلمتين μ ، σ^2 (أي الوسط الحسابي و التباين). وفي الكثير من الحالات تكون هاته المعالم مجهولة، وهدف أي إحصائي معرفتها، وليس من السهل معرفة قيمة هاته المعالم بالضبط، لهذا نضطر إلى تقديرها، حيث أن هناك طريقتين للتقدير هما :

أ- **التقدير بنقطة** "Point Estimation":

يعني اختيار قيمة واحدة كتقدير لمعلمة مجهولة كأن يستخدم متوسط الدخل الشهري للأسرة المحسوب من عينة عشوائية من الأسر المسحوبة من مجتمع معين **كتقدير** لمتوسط الدخل الشهري للأسرة في ذلك المجتمع.

ب- **التقدير بفترة** "Interval Estimation":

من الطبيعي أن **التقدير بنقطة** لأية معلمة لا نتوقع فيه أن يقدر تلك المعلمة بدون خطأ؛ بعبارة أخرى نتوقع أن يكون تقدير أي معلمة مطابقا تماما لقيمة المعلمة المطلوب تقديرها. فمثلا لا نتوقع أن يقدر الوسط الحسابي للعينة ($\overline{x}$) متوسط المجتمع المسحوب منه العينة (μ) بدون أي خطأ؛ بعبارة أخرى لا نتوقع أن يكون $\overline{x}$ مطابقا تماما لـ μ، ولذلك فإنه قد يكون من المرغوب فيه تحديد فترة يتوقع أن تقع قيمة المعلمة داخلها و هذه الفترة تسمى **بفترة الثقة**، حيث تقع المعلمة داخل حدود تلك الفترة بدرجة ثقة (أو احتمال) يحدد ذلك. فمثلا بدلا من استخدام $\overline{x} = 300.00$ دج كتقدير بنقطة لمتوسط دخل الأسرة الشهري بالدينار في مجتمع معين (أي كتقدير لـ μ) تستخدم فترة ثقة لذلك، كأن يقال بأن متوسط دخل الأسرة الشهري بالدينار (μ) يقع في الفترة من 270.00 إلى 330.00 دينار بدرجة ثقة 95%.

أي أن: $270.00 \leq \mu \leq 330.00$

باحتمال قدره: 0.95

2-التقدير بنقطة Point Estimation:

إن أحد أهداف النظرية الإحصائية هو إيجاد الطرق التي تمكننا من الاستدلال على قيمة معالم المجتمع ، و ذلك عن طريق النتائج التي نحصل عليها من عينة عشوائية نختارها من بين مفردات المجتمع.

فإذا كنا نرغب في تقدر أحد معالم المجتمع و ليكن θ من خلال عينة من المشاهدات $(x_1,x_2, \ldots\ldots,x_n)$ مسحوبة من مجتمع. فإن القيمة التي يتم حسابها للمعلمة θ من واقع هذه المشاهدات تسمى **تقديرا** (Estimate)، بينما الدالة أو الصيغة الرياضية الإحصائية التي تستخدم للوصول إلى هذا التقدير تسمى **مقدرا** (Estimator)، والمقدر هو إحصاء أو دالة تعتمد على المشاهدات، بينما التقدير هو قيمة هذه الدالة عند التعويض بقيم المشاهدات فيها، ولهذا فإن التقدير يختلف من عينة لأخرى رغم استخدام نفس المقدر. و هذا أمر طبيعي، حيث هناك اختلاف بين قيم المشاهدات من عينة لأخرى رغم أن المقدر له نفس الصيغة التي يتم التعويض فيها.

على العموم يمكن تقدير معلمة المجتمع بقيمة واحدة و هي قيمة المقدر التي نحصل عليها من بيانات العينة حيث:

- الوسط الحسابي لعينة عشوائية ($\overline{x}$) هو مقدر لمتوسط المجتمع (μ).
- تباين العينة (S^2) هو مقدر لتباين المجتمع (σ^2).
- نسبة صفة معينة في العينة p هي تقدير نقطي للنسبة الحقيقية لنفس الصفة في المجتمع P.

إذن كل مقدر من هذه المقدرات المذكورة يسمى تقديرا بنقطة لأن كل منهم عدد حقيقي واحد أو نقطة واحدة على خط الأعداد الحقيقة.

خصائص المقدر الجيد:

من أهم الخصائص التي يجب أن تتوفر في المقدر الجيد هي:

أ- عدم التحيز (Unbiasedness)
ب- الكفاءة (Efficiency)
ج- الاتساق (Consistency)

فيما يلي سنتطرق للخاصية الأولى والثانية بشيء من الإسهاب، بينما الخاصية الثالثة سنكتفي بعرضها بصورة مبسطة، لأن معالجتها تحتاج إلى مستوى عال من الرياضيات.

أ- خاصية عدم التحيز:

يقال بأن الإحصاءة $\hat{\theta}$ مقدرا غير متحيز للمعلمة المجهولة θ إذا كانت القيمة المتوقعة لـ $\hat{\theta}$ تساوي المعلمة المجهولة θ ، أي:

$$\mathbf{E}(\hat{\theta}) = \theta \ldots\ldots..(01)$$

ونعلم مما سبق أن الوسط الحسابي لتوزيع المعاينة للوسط الحسابي للعينة هو نفسه الوسط الحسابي للمجتمع الأصلي، أي أن:

$$\mathbf{E(\overline{x})=\mu}$$: أي $\mu_{\overline{X}} = \mu$

وبالتالي، فإن الإحصاءة $\overline{x}$ هي مقدر غير متحيز للمعلمة المجهولة μ، ويمكن إثبات ذلك في النظرية التالية.

نظرية (01): إذا كان $\overline{x}$ الوسط الحسابي لعينة عشوائية $(x_1,x_2......,x_n)$ مسحوبة من مجتمع وسطه الحسابي μ فإن : $\mathbf{E(\overline{x})=\mu}$

البرهان :

$$\mathbf{E(\overline{x})} = E\left(\frac{x_1+x_2+\cdots+x_n}{n}\right)$$

$$=\frac{1}{n}[\ E(x_1+x_2+\cdots\cdots+x_n)\ ]$$

$$=\frac{1}{n}[\ E(x_1)+E(x_2)+\cdots\cdots+E(x_n)\ ]$$

$$=\frac{1}{n}[\ \mu+\mu+\cdots\cdots+\mu]$$

$$=\frac{1}{n}[\ n\ \mu]$$

$$=\mathbf{\mu}$$

كذلك نستطيع أن نثبت أن تباين العينة S^2 هو مقدر غير متحيز لتباين المجتمع σ^2 في حالة سحب العينة العشوائية مع الإرجاع.

أما في حالة السحب دون إرجاع، فان المقدر غير المتحيز لتباين المجتمع هو $(\frac{N-1}{N}S^2)$.

وفي المثال (01) الموالي سوف نثبت عدديا أن تباين العينة هو مقدر غير متحيز للمعلمة المجهولة σ^2 في حالة السحب مع الإرجاع، أما الإثبات النظري فسنوضحه في النظرية التالية.

نظرية (02): إذا كان S^2 تباين عينة عشوائية $(\ x_1,x_2......,x_n\ )$ مسحوبة من مجتمع تباينه σ^2 فإن:

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

البرهان:

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

$$= \frac{1}{n-1} E[\Sigma(x_i^2 - 2x_i\overline{X} + \overline{X}^2)]$$

$$= \frac{1}{n-1} E(\Sigma x_i^2 - 2\overline{X}\Sigma x_i + n\overline{X}^2)$$

$$= \frac{1}{n-1} E(\Sigma x_i^2 - n\overline{X}^2)$$

$$= \frac{1}{n-1} [\Sigma E(x_i^2) - nE(\overline{X}^2)]$$

وذلك لأن $\Sigma x_i = n\overline{X}$ ، وبما أن تباين المجتمع يمكن حسابه كما يلي:

$$\sigma^2 = E(x)^2 - \mu^2$$

وكذلك تباين توزيع المعاينة للوسط الحسابي للعينة $\sigma^2_{\overline{X}}$ يمكن حسابه كما يلي:

$$\sigma^2_{\overline{X}} = E(\overline{X})^2 - \mu^2$$

وبالتعويض عن $E(x)^2$ و $E(\overline{X})^2$ نجد أن :

$$E(S^2) = \frac{1}{n-1} [\Sigma(\sigma^2 + \mu^2) - n(\sigma^2_{\overline{X}} + \mu^2)]$$

$$= \frac{1}{n-1} (n\sigma^2 + n\mu^2 - n\sigma^2_{\overline{X}} - n\mu^2)$$

$$= \frac{1}{n-1} (n\sigma^2 - n\sigma^2_{\overline{X}})$$

وبما أن السحب تم مع الإرجاع فإن:

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$$

إذن :

$$E(S2) = \frac{1}{n-1} \left(n\sigma^2 - n\frac{\sigma^2}{n}\right)$$

$$=\frac{1}{n-1}\left(n\sigma^2-\sigma^2\right)$$

$$=\frac{1}{n-1}(n-1)\sigma^2 \quad = \quad \boldsymbol{\sigma^2}$$

وبالتالي فإن S^2 هو مقدر غير متحيز لتباين المجتمع σ^2

ب- خاصية الكفاءة :

قد يوجد للمعلمة الواحدة أكثر من مقدر، وبالتالي يمكننا المقارنة بين هذه المقدرات من خلال المقارنة بين تبايناتهم، حيث نعتبر أن المقدر **الأقل تباينا** هو المقدر **الأكثر كفاءة** (More Efficient).

فإذا كان لدينا المقدران: $\hat{\theta}_1$ و $\hat{\theta}_2$ للمعلمة θ. وكان تباين المقدر الأول هو $\sigma^2_{\hat{\theta}_1}$ و تباين المقدر الثاني هو $\sigma^2_{\hat{\theta}_2}$، فإن المقدر $\hat{\theta}_1$ يكون أكثر كفاءة من المقدر $\hat{\theta}_2$ إذا كان: $\sigma^2_{\hat{\theta}_1} < \sigma^2_{\hat{\theta}_2}$.

على افتراض أن: $\hat{\theta}_1$ و $\hat{\theta}_2$ هي مقدرات غير متحيزة.

مثال: بالنسبة للمجتمعات الطبيعية، فإن الوسط الحسابي والوسيط يمثلان تقديرين غير متحيزين لمتوسط المجتمع μ ، و لكن تباين الوسط أقل من تباين الوسيط، وبالتالي فإن الوسط الحسابي يكون مقدرا أكثر كفاءة لمتوسط المجتمع μ. و للتأكد من ذلك لدينا: تباين الوسط هو

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$$

أما تباين الوسط فهو يساوي تقريبا:

$$\sigma^2_{Me} = \frac{\pi\sigma^2}{2n}$$

وتقاس الكفاءة النسبية بين التباينين بقسمة أحدهما على الآخر، و ذلك كما يلي:

$$0.64 = \frac{2}{\pi} = \frac{\frac{\sigma^2}{n}}{\frac{\pi\sigma^2}{2n}} = \frac{\text{تباين الوسط}}{\text{تباين الوسيط}}$$

وعليه يتضح لنا أنه عند نفس حجم العينة فإن تباين الوسط أقل من تباين الوسيط، وبالتالي فإن الوسط الحسابي يعتبر مقدرا أكثر كفاءة من الوسيط.

ج-خاصية الاتساق:

يقال بأن الإحصاءة $\hat{\theta}$ هي مقدر متسق للمعلمة θ إذا كانت $\hat{\theta}$ تقترب من θ كلما زاد حجم العينة واقترب من المالا نهاية.

وعليه يمكن إثبات أن $\hat{\theta}$ تكون مقدر متسق لـ θ ، إذا كانت $\hat{\theta}$ مقدر غير متحيز لـ θ، وكان تباين $\hat{\theta}$ يقترب من الصفر كلما إقتربت n من المالا نهاية.

فمثلا إذا كان $(\overline{x})$ هو الوسط الحسابي لعينة حجمها n مسحوبة من مجتمع وسطه μ و تباينه σ^2 ، يكون $(\overline{x})$ مقدرا متسقا لـ μ لأن:

$$E(\overline{x}) = \mu$$

$$\sigma^2_{\overline{x}} = \frac{\sigma^2}{n} \rightarrow 0 \quad As \quad n \rightarrow \omega$$

فالشرط الأول سبق التأكد منه ، أما الشرط الثاني فالتأكد منه بسيط جدا حيث ،كلما زاد حجم العينة (و قرب من المالا نهاية) فإن تباين المتوسط $\sigma^2_{\overline{x}}$ يقترب من الصفر .

وعليه يمكن القول بأن S^2 تقدير متسق لـ σ^2.