المحور الثاني: التقدير الاحصائي وفترات الثقة

تمهيد

رأينا في المحور السابق أن الهدف الأساسي من دراسة خواص العينات وتوزيعاتها كان الاستدلال على معالم المجتمع الذي تم سحب منه تلك العينات. وفي هذا الصدد، تجدر الإشارة إلى مسألة بالغة الأهمية وهي ضرورة معرفة أن القيم التي يتم الحصول عليها عن طريق أسلوب المعاينة لا تعكس بالضرورة القيم الحقيقية المناظرة لها في المجتمعات الأصلية، وهذا ما يدفعنا لمحاولة إجراء عمليات **تقدير** لتلك المعالم أو المجالات التي تقع ضمنها تلك القيم بدرجة ثقة معينة، وهذا قصد التأكد من مدى دقة تقييم الاحصاءات للمعلمات المناظرة لها.

ما المقصود بالتقدير (Estimation)؟

هو أسلوب إحصائي يستخدم لتقدير معالم المجتمع التي غالبا ما تكون **مجهولة** عن طريق استخدام مقاييس (إحصائيات) العينة. وهناك نوعان أو أسلوبان للتقدير الإحصائي، يسمى الأول **التقدير بنقطة أو بالقيمة الواحدة** ويسمى الثاني **التقدير بفترة (فترة أو مجال التقدير أو الثقة)**

أولا- التقدير بقيمة أو بنقطة

هو تقدير معلمة المجتمع المجهولة **بقيمة واحدة فقط** تحسب من العينة العشوائية المسحوبة من هذا المجتمع.

فمثلا عند سحب عينة عشوائية من مجتمع ما وقمنا بحساب:

- ☜ المتوسط الحسابي للعينة $\bar{X}$ فإنه يعتبر تقدير بقيمة واحدة لمعلمة المجتمع μ
- ☜ تباين العينة S^2 فإنه يعتبر تقدير بقيمة واحدة لمعلمة المجتمع σ^2
- ☜ الانحراف المعياري للعينة S فإنه يعتبر تقدير بقيمة واحدة لمعلمة المجتمع σ ...إلخ

ونقول عندئذ:

$\bar{X}$	هو مقدر لـ	μ
S^2	هو مقدر لـ	σ^2
S	هو مقدر لـ	σ
$\vdots$	$\vdots$	$\vdots$

ونكتب حينها:

$$\hat{\mu} = \bar{X}$$

حيث يشير الرمز $\hat{}$ للتقدير

ويسمى $\hat{\mu}$ بمتوسط المجتمع المقدر أو المتوسط المقدر للمجتمع، وهكذا بالنسبة لباقي المعالم

ثانيا- التقدير بفترة (التقدير بمجال أو المجالي)

نرمز لدرجة الثقة بالرمز $(1-\alpha)$ " **احتمال وقوع المعلمة المجهولة داخل الفترة المقدرة"**

ومكمل هذه القيمة يسمى مستوى المعنوية أو مستوى الدلالة ويرمز له بالرمز α " **احتمال عدم وقوع المعلمة المجهولة داخل الفترة المقدرة"**

$$\text{درجة الثقة} + \text{مستوى المعنوية} = 1$$

فمثلا إذا كانت درجة الثقة هي 95 % (0.95) فإن قيمة مستوى المعنوية تساوي 5 % (0.05)

بصفة عامة، لإيجاد فترة الثقة لمعلمة مجهولة، ولتكن على سبيل المثال θ، يجب إيجاد إحصائيتين ولتكونا L_1 و L_2 بحيث يكون:

$$P(L_1 \leq \theta \leq L_2) = 1-\alpha$$

L_1: هو الحد الأدنى لفترة أو مجال الثقة

L_2: هو الحد الأعلى لفترة أو مجال الثقة

أ- فترة أو مجال الثقة حول متوسط المجتمع μ

هنا يمكن التمييز بين عدة حالات، هي كالآتي

☜ **فترة الثقة حول المتوسط μ عندما يكون التباين σ^2 معلوما**

إذا كان المجتمع يتبع التوزيع الطبيعي، فإن توزيع المعاينة للمتوسط الحسابي للعينة سيتبع هو الآخر التوزيع الطبيعي، وذلك سواء كان حجم العينة صغيرا أو كبيرا. وفي هذه الحالة، يمكننا تحويل - كما رأينا ذلك في المحور الأول- المتغير العشوائي $\bar{X}$ إلى المتغير المعياري Z بالشكل:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \backsim N(0,\ 1)$$

يمكن ملاحظة هنا أن المتغير Z هو دالة في المقدر $\bar{X}$ والمعلمة المجهولة هي μ والتوزيع الاحتمالي للمتغير Z هو التوزيع الطبيعي المعياري الذي لا يعتمد على μ ولذلك يمكن استخدام Z لإيجاد فترة الثقة لمتوسط المجتمع μ. ويعطى مجال الثقة للمتوسط الحسابي للمجتمع μ عندما يكون تباين المجتمع معلوما عند مستوى الثقة $(1-\alpha)$ كالتالي:

مجال الثقة $=$ التقدير بنقطة $\pm$ معامل الثقة $\times$ الخطأ المعياري

وبالشكل الرياضي:

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

عموما، يمكن تلخيص خطوات تقدير متوسط المجتمع فيما يلي:

- ✓ حساب المتوسط الحسابي للعينة $\bar{X}$
- ✓ حساب الخطأ المعياري للوسط و الذي يساوي $\frac{\sigma}{\sqrt{n}}$
- ✓ ضرب الخطأ المعياري للوسط في معامل الثقة المناسب (أو الدرجة المعيارية) حسب درجة الثقة المطلوبة، أي حساب $Z\frac{\sigma}{\sqrt{n}}$
- ✓ طرح حاصل الضرب السابق من متوسط العينة للحصول على الحد الأدنى لفترة التقدير، ثم جمع حاصل الضرب مرة أخرى بمتوسط العينة للحصول على الحد الأعلى لفترة التقدير.

ملاحظات مهمة

1- إذا كانت المعاينة من مجتمع (محدود) أو السحب بدون إرجاع، يتم استبدال العلاقة السابقة بإضافة معامل التصحيح كالتالي:

$$P\left(\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \leq \mu \leq \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right) = 1 - \alpha$$

يمكن كتابة أشهر وأهم درجات ومعاملات الثقة (للتوزيع الطبيعي) في الجدول التالي، مع الإشارة إلى أن 95 % و99 % هي أشهرها وأكثرها استخداما في مجال البحث العلمي

مستوى الثقة $(1-\alpha)$		مستوى المعنوية المناظر α		معامل الثقة Z بدليل $\frac{\alpha}{2}$
النسبة	القيمة	النسبة	القيمة	
% 90	0.90	% 10	0.1	1.65
% 95	0.95	% 5	0.05	1.96
% 99	0.99	% 1	0.01	2.58

☜ فترة الثقة حول المتوسط μ عندما يكون التباين σ^2 مجهولا

❖ حالة العينات الكبيرة

في هذه الحالة يتم استبدال الانحراف المعياري للمجتمع المجهول بالانحراف المعياري للعينة كتقدير له، فتصبح الصيغة الرياضية لفترة الثقة عندئذ كما يلي:

$$P\left(\bar{X} - Z\alpha_{/2}\frac{S}{\sqrt{n-1}} \le \mu \le \bar{X} + Z\alpha_{/2}\frac{S}{\sqrt{n-1}}\right) = 1 - \alpha$$

❖ حالة العينات الصغيرة

في حالة العينات الصغيرة التي حجمها أقل من 30، فإن توزيع المعاينة للمتوسط يتبع توزيع المجتمع، وعليه إذا كان المجتمع يتبع التوزيع الطبيعي فإن توزيع المعاينة للمتوسط يتبع هو كذلك التوزيع الطبيعي. ولتقدير متوسط مجتمع مجهول انطلاقا من بيانات عينة في حالة عدم معرفة قيمة تباين (الانحراف المعياري) المجتمع، فإنه يتم استخدام **توزيع ستودنت Student**

ويستخدم في هذه الحالة الانحراف المعياري للعينة S كتقدير للانحراف المعياري **غير المعلوم** للمجتمع. وتكتب الصيغة الرياضية لمجال الثقة في هذه الحالة بالشكل:

$$P\left(\bar{X} - T\alpha_{/2}\frac{S}{\sqrt{n-1}} \le \mu \le \bar{X} + T\alpha_{/2}\frac{S}{\sqrt{n-1}}\right) = 1 - \alpha$$

❖ فترة الثقة باستخدام نظرية تشيبيشيف *Chebyshev*

تستخدم نظرية تشيبيشيف إذا كان **توزيع المجتمع مجهولا** أو **لا يتبع التوزيع الطبيعي**، ويكون حجم العينة المسحوبة **صغير**. ويعطى مجال الثقة في هذه الحالة بالشكل:

$$P\left(\bar{X} - K\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + K\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

حيث:

$$1 - \alpha = 1 - \frac{1}{K^2}$$

ب- مجال الثقة للفرق بين متوسطي مجتمعين

في كثير من الأحيان قد نرغب في مقارنة متوسطي مجتمعين، وهنا قد تواجهنا حالتين. فقد يكون المجتمعين مستقلين أو غير مستقلين

☜ مجال الثقة للفرق بين متوسطي مجتمعين مستقلين

وهنا كذلك يمكن التمييز بين حالتين

❖ حالة التوزيعات الطبيعية المستقلة بتباين معلوم

في هذه الحالة يكون:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

وعندئذ، يكتب مجال الثقة للفرق بين متوسطي المجتمعين بالشكل:

$$P\left[(\bar{X}_1 - \bar{X}_2) - Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le (\mu_1 - \mu_2) \le (\bar{X}_1 - \bar{X}_2) + Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right] = 1 - \alpha$$

❖ حالة التوزيعات الطبيعية المستقلة بتباين مجهول

إذا كان σ_1^2 و σ_2^2 مجهولين نستبدلهما بتباين العينة S_1^2 و S_2^2 وهنا نميز بين حالتين:

○ حالة العينات الكبيرة

حسب نظرية النهاية المركزية، يكون لدينا:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1 - 1} + \frac{S_2^2}{n_2 - 1}}} \sim N(0,1)$$

ومنه، فمجال الثقة للفرق بين متوسطي المجتمعين يكون بعد إجراء التحويل بالشكل:

$$P\left[(\bar{X}_1 - \bar{X}_2) - Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1 - 1} + \frac{S_2^2}{n_2 - 1}} \le (\mu_1 - \mu_2) \le (\bar{X}_1 - \bar{X}_2) + Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1 - 1} + \frac{S_2^2}{n_2 - 1}}\right] = 1 - \alpha$$

○ حالة العينات الصغيرة

وهنا بدوره يمكن التمييز بين حالتين

✓ عندما يكون التباينين مجهولين ومتساويين

في هذه الحالة يكون مجال الثقة للفرق $(\mu_1 - \mu_2)$ عندما يكون فيها $\sigma_1^2 = \sigma_2^2$

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim T_{\left(\frac{\alpha}{2}, n_1 + n_2 - 2\right)}$$

ومنه، فمجال الثقة للفرق بين متوسطي المجتمعين يكون بعد إجراء التحويل بالشكل:

$$P\left[(\bar{X}_1 - \bar{X}_2) - T_{\frac{\alpha}{2}} S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le (\mu_1 - \mu_2) \le (\bar{X}_1 - \bar{X}_2) + T_{\frac{\alpha}{2}} S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right] = 1 - \alpha$$

✓ عندما يكون التباينين مجهولين وغير متساويين

في هذه الحالة يتم استخدام توزيع ستودنت بدرجة حرية معينة (V) عندما يكون فيها $\sigma_1^2 \neq \sigma_2^2$، حيث تكون قيمة درجة الحرية بالشكل (كما رأينا ذلك في المحور الأول):

$$V = \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^2}{\frac{\left[\frac{S_1^2}{n_1}\right]^2}{n_1 - 1} + \frac{\left[\frac{S_2^2}{n_2}\right]^2}{n_2 - 1}}$$

ويكون مجال الثقة للفرق $(\mu_1 - \mu_2)$ عندئذ بالشكل:

$$P\left[(\bar{X}_1 - \bar{X}_2) - T_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \le (\mu_1 - \mu_2) \le (\bar{X}_1 - \bar{X}_2) + T_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right] = 1 - \alpha$$

☜ مجال الثقة للفرق بين متوسطي مجتمعين غير مستقلين (مرتبطين)

يمكن هنا كذلك التمييز بين حالتين على حسب حجم العينة

❖ **حالة العينات الكبيرة**

يكون:

$$Z = \frac{\bar{d} - \mu_d}{\frac{\sigma_d}{\sqrt{n}}} \sim N(0,1)$$

ويمكن تعيين مجال ثقة تقريبي لمتوسط مجتمع الفروق بمستوى ثقة $(1-\alpha)$ كما يلي:

$$P\left[\bar{d} - Z_{\frac{\alpha}{2}} \frac{\sigma_d}{\sqrt{n}} \leq \mu_d \leq \bar{d} + Z_{\frac{\alpha}{2}} \frac{\sigma_d}{\sqrt{n}}\right] = 1 - \alpha$$

❖ **حالة العينات الصغيرة**

مادام $n < 30$ سيكون لدينا:

$$T = \frac{\bar{d} - \mu_d}{\frac{S_d}{\sqrt{n}}} \sim T_{\left(\frac{\alpha}{2}, n-1\right)}$$

ويكون مجال الثقة التقريبي لمتوسط مجتمع الفروق بمستوى ثقة $(1-\alpha)$ كما يلي:

$$P\left[\bar{d} - T_{\frac{\alpha}{2}} \frac{S_d}{\sqrt{n}} \leq \mu_d \leq \bar{d} + T_{\frac{\alpha}{2}} \frac{S_d}{\sqrt{n}}\right] = 1 - \alpha$$

ج- مجال الثقة للنسبة

يكتب مجال الثقة بالشكل:

$$P\left[\bar{p} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}\bar{q}}{n}} \leq p \leq \bar{p} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}\bar{q}}{n}}\right] = 1 - \alpha$$

أما إذا كان المجتمع محدود أو السحب مع عدم الإرجاع، فإننا نستخدم معامل الإرجاع في العلاقة السابقة ويصبح مجال التقدير كالتالي:

$$P\left[\bar{p} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}\bar{q}}{n}} \sqrt{\frac{N-n}{N-1}} \leq p \leq \bar{p} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}\bar{q}}{n}} \sqrt{\frac{N-n}{N-1}}\right] = 1 - \alpha$$

د- مجال الثقة للفرق بين نسبتي مجتمعين

يكتب مجال الثقة بالشكل:

$$P\left[(\bar{p}_1 - \bar{p}_2) - Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}_1\bar{q}_1}{n_1} + \frac{\bar{p}_2\bar{q}_2}{n_2}} \leq (p_1 - p_2) \leq (\bar{p}_1 - \bar{p}_2) - Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}_1\bar{q}_1}{n_1} + \frac{\bar{p}_2\bar{q}_2}{n_2}}\right] = 1 - \alpha$$

إذا كان المجتمعان محدودان أو السحب مع عدم الإرجاع، نضيف لهذه العلاقة معامل التصحيح