المحـــور الثانـــي

نظريـــــة التقـديـــــر

Estimation Theory

تمهيد:

ذكرنا في المحور السابق بأن العينة sample هي جزء من المجتمع *Population* وطريقة اختيار هذا الجزء يسمي طريقة المعاينة *sampling method* والغاية الرئيسية من دراسة العينات هوللاستدلال منها على خواص المجتمع الذي تعود اليه هذه العينات.

فاذا كان لدينا المتغير العشوائي X فان توزيعه الاحتمالي (أوكثافة احتماله) تعتمد على ثابت θ واحد أوأكثر لا تعرف قيمتها، وهذه الثوابت تسمى معالم *Parameters* وفي هذا المحور سندرس طرق تقدير معالم المجتمع من مقاييس الاحصائيات التي تحسب من العينة حيث اننا نحتاج لحساب قيمة أواحصائية *Statistic* من العينة لكل معلمة من معالم المجتمع (أودالة كثافة الاحتمال).

هذا وكل قيمة تحسب من العينة تسمى تقديرا *Estimate* أما الطريقة التي استخدمت في التقدير فتسمي مقدرة *Estimator* فالتقدير غير ثابت من عينة إلى أخرى عند استخدام نفس الطريقة بينما المقدر يكون ثابتا الا اذا تغيرت طريقته، هذا والتقدير اما ان يكون : تقدير المعلمة بنقطة *Point Estimation* أوتقدير المعلمة بفترة *Interval Estimation*

1– التقديـــر بنقطـــة *Point Estimation*:

اذا حسبت قيمة مفردة من العينة كتقدير لمعلمة من المجتمع فالطريقة تسمى تقدير النقطة لأن نقطة واحدة فقط من فضاء العينة قد استخدم تقدير للمعلمة .

مثال (1)

إن قيمة الوسط الحسابي للعينة $\overline{X}$ هوتقدير نقطة للمعلمة μ_x الوسط الحسابي للمجتمع الذي تعود اليه هذه العينة .

كما أن التباين S^2 للعينة هوتقدير نقطة لتباين المجتمع δ_x^2 .

وكذلك النسبة $\hat{p}$ هوتقدير لنسبة المجتمع p .

خصائص المقدر الجيد هي :

عدم التحيز *Unbiasedness*

إن المقدر $\hat{\theta}$ يعتبر مقدرا غير متحيز اذا كان توقعه يساوي قيمة المعلمة θ أي $E\left(\hat{\theta}\right)=\theta$

مثال

افرض بأن $X_1, X_2, \ldots\ldots\ldots X_n$ هي عينة عشوائية من مجتمع له وسط حسابي μ_x

لذا فإن $E\left(\bar{X}\right)=\mu_X$

وكذلك $E\left(X_i\right)=\mu_X$

لذا فإن كلا من $\bar{X}$ و X_i هي مقدر غير متحيز للوسط الحسابي للمجتمع μ_X .

الاتساق *Consistency*

يكون المقدر متسقا اذا كانت قيمته لا تختلف اختلاف جوهرية عن قيمة المعلمة الحقيقية للمجتمع بزيادة حجم العينة أي أن $\lim_{n\to\infty} P\left(\left|\hat{\theta}-\theta\right| \prec \varepsilon\right)=1$

حيث أن ε هوالفرق بين المقدر والمعلمة .

الكفاءة *Efficiency*

إن كفاءة المقدر غير المتحيزة $\hat{\theta}_1$ إلى المقدر غير المتحيز $\hat{\theta}_2$ هونسبة تباين المقدر $\hat{\theta}_2$ الى تباين المقدرة $\hat{\theta}_1$ أي $e\left(\hat{\theta}_1\right)=\dfrac{V\left(\hat{\theta}_2\right)}{V\left(\hat{\theta}_1\right)}$

ومن التعريف أعلاه يتضح بأن المقدر الأقل تباينا هوالأعلى كفاءة.

مثال الوسط الحسابي للعينة $\bar{X}$ هومقدر غير متحيز ل μ_X وكذلك الوسيط للعينة هومقدر غير متحيز لـ μ_X ولكن تباين $\bar{X}$ هو $\frac{\delta_X^2}{n}$ بينما تباين الوسيط هو $\frac{\pi}{2}\frac{\delta_x^2}{n}$

لذا فان كفاءة الوسط الحسابي هو $e\left(\bar{X}\right)=\frac{V\left(M_e\right)}{V\left(\bar{X}\right)}=\frac{\pi/2 * \delta_x^2/n}{\delta_x^2/n}=\frac{\pi}{2}$

أي أن الوسط الحسابي $\bar{X}$ هوأكفأ تقديرا من الوسيط.

الكفاية *Sufficiency* :

يكون المقدر $\hat{\theta}$ مقدرا كافيا للمعلمة θ اذا كان قد شمل كل المعلومات ذات العلاقة به θ المتوافرة في العينة.

فعند سحب عينة عشوائية في مجتمع يتوزع توزعا طبيعيا فأن $\bar{X}$ هومقدر كاف لـ μ_X لأنه لا يمكن اضافة أي شيء على $\bar{X}$ لجعله مقدرا أحسن لـ μ_X لأن $\bar{X}$ يحوي على جميع المعلومات المتعلقة ب μ_X من العينة

. هذا وهناك عدة طرق مهمة لايجاد تقدير المعلمة بنقطة وهي :

طريقة الامكان الأكبر *Maximum Likelihood Method*

طريقة العزوم *Method of Moments*

طريقة مربع كاي المصغرة *Minimum Chi−Square Method*

طريقة المربعات الصغرى *Method of Least Square*

2- **التقدير بمجال *Interval Estimation***

يفضل أحيانا إعطاء تقدير للمعلمة المجهولة θ بحيث لا تكون بعيدة عن الحقيقة و الواقع و باحتمال معين و الوسيلة المستخدمة في البحث عن تقديرات موثوقة باحتمال معين ضمن مجال محدد ، و يطلق على هذا المجال الذي يحتوي فيه التقدير إسم مجال التقدير

فمن أجل الحصول على مجال التقدير للمعلمة المجهولة θ علينا أن نشكل مجالا من الشكل $\hat{\theta} \pm k$ حيث تتعلق $\hat{\theta}$ بالعينة الخاصة المختارة و يتحدد العدد k بواسطة توزيع المعاينة بالإحصاء $\hat{\theta}$

إن قولنا أن التقدير $\hat{\theta}$ مساو تماما للمعلمة θ يعني أن: $\hat{\theta} - k \prec \theta \prec \hat{\theta} + k$

و يمكن و بواسطة توزيع المعاينة للإحصاء θ أن نحدد قيمة k بحيث يكون الإحتمال :

$$P\left(\hat{\theta} - k \prec \theta \prec \hat{\theta} + k\right)$$

مساويا لقيمة معينة تهم الباحث و لتكن $1-\alpha$ و عليه يمكن أن نعبر عن هذا المجال بما يلي :

$$prob\left(T_1 \prec \theta \prec T_2\right) = 1-\alpha \quad \text{أي} \quad prob\left[\theta \in \left(T_1, T_2\right)\right] = 1-\alpha$$

حيث أن: θ المعلمة المراد تقديرها، $\left(T_1, T_2\right)$ يسمى مجال ثقة ، T_1 حد الثقة الأدنى، T_2 حد الثقة الأعلى، $1-\alpha$ معامل الثقة حيث $1 \prec \alpha \prec 0$ ويحدد مسبقا، $\left(1-\alpha\right) \times 100\%$ درجة الثقة للمعلمة θ .

مجال الثقة المتوسط الحسابي للتوزيع الطبيعي .

أ –حالة التباين معروفا .

لنفرض أن المطلوب هو إنشاء مجال ثقة للمتوسط الحسابي μ_X للتوزيع الطبيعي إذا كانت δ_X^2 معلومة . بما أن المتوسط الحسابي للعينة $\bar{X}$ كمقدر للمعلمة μ_X للتوزيع الطبيعي (التباين معلوم) له خصائص جيدة مثل الكفاءة و الكفاية وعدم التحيز فإننا سوف نستخدمه لإنشاء مجال ثقة للمتوسط الحسابي للمجتمع μ_X ضمن مجال محدد بمعامل ثقة $1-\alpha$ من خلال العلاقة :

$$prob\left[\mu_X \in \left(T_1, T_2\right)\right] = 1-\alpha$$

رأينا سابقا ، أن المتوسط الحسابي لعينة عشوائية حجمها n من توزيع طبيعي أو أن حجم العينة كبير و ذلك بغض النظر عن نوع المجتمع المدروس وسطه الحسابي μ_X و تباينه δ_X^2 يتبع أيضا التوزيع الطبيعي بمتوسط حسابي ولكن بتباين $\delta_{\bar{X}}^2 = \dfrac{\delta_X^2}{n}$

و بالتالي فإن الكمية المحورية المناسبة هي : $Z = \dfrac{\left(\bar{X} - \mu_X\right)}{\delta/\sqrt{n}}$

وهي دالة في المقدر $\bar{X}$ و المعلمة μ_X و توزيعهما هو التوزيع الطبيعي المعياري الذي لا يعتمد على المعلمة μ_X ، و من ثم فإن احتمال أن تقع هذه القيمة بين القيمتين :

$Z_{\left(1-\frac{\alpha}{2}\right)}$ و $-Z_{\left(1-\frac{\alpha}{2}\right)}$ يساوي $1-\alpha$

$$prob\left(-Z_{\left(1-\frac{\alpha}{2}\right)} \prec Z \prec Z_{\left(1-\frac{\alpha}{2}\right)}\right) = 1-\alpha$$

وبالتعويض عن Z نحصل على :

$$prob\left(-Z_{\left(1-\frac{\alpha}{2}\right)} \prec \frac{\left(\bar{X} - \mu_X\right)}{\delta_X/\sqrt{n}} \prec Z_{\left(1-\frac{\alpha}{2}\right)}\right) = 1-\alpha$$

وبعد ضرب كل طرف من أطراف هذه المتباينة بالعدد $\delta_X/\sqrt{n}$ ثم طرح من جميع الأطراف المقدار $\bar{X}$ ، نجد بعد ضربها بإشارة ناقص و عندها يمكننا أن نكتب :

$$prob\left(\bar{X} - Z_{\left(1-\frac{\alpha}{2}\right)} \delta_X/\sqrt{n} \prec \mu_X \prec \bar{X} + Z_{\left(1-\frac{\alpha}{2}\right)} \delta_X/\sqrt{n}\right) = 1-\alpha$$

وهذا ما أشرنا إليه سابقا في الصورة العامة كما يلي $prob\left(T_1 \prec Z \prec T_2\right) = 1-\alpha$

و هكذا نجد أن $(1-\alpha)\times 100\%$ مجال الثقة للوسط μ_X في مجتمع فيه δ_X معلوم هو مجال

حيث أن:

حد الثقة الأدنى هو $T_1 = \bar{X} - Z_{\left(1-\frac{\alpha}{2}\right)} \delta_X \big/ \sqrt{n}$ ،

حد الثقة الأعلى $T_2 = \bar{X} + +Z_{\left(1-\frac{\alpha}{2}\right)} \delta_X \big/ \sqrt{n}$

و المعلمة θ محل التقدير هنا تمثل المتوسط الحسابي μ_X

ب –حالة التباين غير معروف .

كما يدل عنوان هذه الفقرة فإننا نواجه معلمتين إثنتين مجهولتين ، إذغالبا ما نفكر في

إيجاد تقدير للمتوسط في مجتمع إحصائي تباينه δ_X^2 مجهول .

و يختلف الآن الموضوع عن الحالة السابقة إذ يجب تقدير تباين هذا المجتمع لكي نتمكن من

تقدير المتوسط الحسابي للمجتمع μ_X

لنفرض أن المطلوب هو إنشاء مجال ثقة للمتوسط الحسابي μ_X للتوزيع الطبيعي إذا كانت δ_X^2

غير معلومة .

رأينا سابقا أن المقدار Z له توزيعا طبيعيا معياريا إذا كان المجتمع المدروس طبيعيا أي :

$$Z = \frac{\left(\bar{X} - \mu_X\right)}{\delta_X \big/ \sqrt{n}} \sim N\ (0.1)$$

أو توزيعا قريبا من التوزيع الطبيعي المعياري إذا كان المجتمع المدروس غير طبيعي و ذلك طبقا لنظرية النهاية المركزية بشرط أن يكون حجم العينة n كبيرا .

لكن إذا كانت n صغيرة و حيث أنه لا يجوز لنا اللجوء إلى نظرية النهاية المركزية ، فمن الطبيعي وضع S الإنحراف المعياري المحسوبة من العينة التي تم سحبها بصورة عشوائية بدلا من δ_X

$$\bar{X} = \frac{1}{n}\sqrt{\sum_{i=1}^{n} X_i} \quad \text{و} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

و نظرا لكون العينة صغيرة فإن هذا يجعلنا نشك في أن تكون S ممثلة لقيمة تقريبية وهكذا نفقد كل أساس نظري أو منطقي يسمح لنا بالقول بأن توزيع المتغير الجديد: $\frac{\left(\bar{X} - \mu_X\right)}{S/\sqrt{n}}$

هو التوزيع الطبيعي المعياري . و مع أن توقع S^2 هو $E\left(S^2\right) = \delta^2$ تماما إلا أن احتمال كون S^2 أقل من δ^2 يزيد على النصف .كما نرى من توزيع S^2

رأينا أن المقدار $\frac{(n-1)S^2}{\delta^2}$ له توزيع $\chi^2_{(n-1)}$ أي نكون بصدد متغير كاي مربع بحيث أن :

$$\chi^2 = (n-1)\frac{S^2}{\delta^2} = \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\delta}\right)^2$$

و الذي يتبع توزيع χ^2 بدرجات حرية $(n-1)$ وهذا ما يحملنا على الاعتقاد بأن منحنى الكثافة للمتغير $\frac{\left(\bar{X} - \mu_X\right)}{S/\sqrt{n}}$ سيكون أكثر إنتشارا .

فعندما تزداد n فإن S^2 تقترب من δ^2 و يقترب التوزيع الجديد من التوزيع الطبيعي المعياري .

وحيث أن توزيع t هو توزيع لمتغير يمثل حاصل قسمة متغير يتبع التوزيع الطبيعي المعياري على الجذر التربيعي لمتغير مستقل منه يتبع توزيع χ^2 مقسوما على درجات حريته فإن :

$$t=\frac{Z}{\sqrt{\chi^2/_{n-1}}}$$ يتبع توزيع t بدرجات حرية $(n-1)$ و بما أن :

فإننا نصل إلى $$t=\left(\frac{\bar{X}-\mu_X}{\delta_X/\sqrt{n}}\right)\Bigg/\left(\sqrt{\frac{(n-1).S^2}{(n-1).\delta^2}}\right)$$

$$t=\left(\frac{\bar{X}-\mu_X}{S/\sqrt{n}}\right)\sim t_{(n-1)}$$

أي أن لها توزيع t بدرجات حرية $(n-1)$ و بالتالي فإن t هي الكمية المحورية المناسبة

لهذه الحالة أي حالة التباين مجهول، و عليه فإن :

$$prob\left(-t_{\left(1-\frac{\alpha}{2},n-1\right)}\prec T\prec t_{\left(1-\frac{\alpha}{2},n-1\right)}\right)=1-\alpha$$

حيث أن $t_{\left(1-\frac{\alpha}{2},n-1\right)}$ هي قيمة t المجدولة

و بالتعويض عن t نحصل على ما يلي

$$prob\left(-t_{\left(1-\frac{\alpha}{2},n-1\right)}\prec\frac{\left(\bar{X}-\mu_X\right)}{S/\sqrt{n}}\prec t_{\left(1-\frac{\alpha}{2},n-1\right)}\right)=1-\alpha$$

وبعد ضرب كل طرف من أطراف هذه المتباينة بالعدد $S/\sqrt{n}$ ثم طرح من جميع الأطراف

المقدار $\bar{X}$ ، و بعد ضربها (-1) نجد :

$$prob\left(\bar{X}-t_{\left(1-\frac{\alpha}{2},n-1\right)}S/\sqrt{n}\prec\mu_X\prec\bar{X}+t_{\left(1-\frac{\alpha}{2},n-1\right)}S/\sqrt{n}\right)=1-\alpha$$

حيث أن:

حد الثقة الأدنى هو $T_1 = \bar{X} - t_{\left(1-\frac{\alpha}{2}, n-1\right)} S / \sqrt{n}$ ،

حد الثقة الأعلى $T_2 = \bar{X} + t_{\left(1-\frac{\alpha}{2}, n-1\right)} S / \sqrt{n}$

فترة الثقة للتباين δ_X^2 :

كتقدير نقطي لتباين مجتمع إحصائي δ_X^2 نأخذ عادة تباين العينة S^2 و الذي كما رأينا سابقا، تقديرا غير متحيز لتباين المجتمع المجهول δ_X^2 $\qquad E\left(S^2\right) = \delta_X^2$

و نعلم أن المقدار : $\dfrac{(n-1).S^2}{\delta_X^2}$ يتبع توزيع χ^2 بدرجات حرية $(n-1)$ و ذلك بفرض أن العينة التي سحبناها و حسبنا منها الإحصاء السابق قد أخذت من مجتمع طبيعي تباينه δ_X^2 لذلك يمكننا أن نكتب:

$$prob\left(\chi^2_{\left(\frac{\alpha}{2}, n-1\right)} \prec \frac{(n-1)S^2}{\delta_X^2} \prec \chi^2_{\left(1-\frac{\alpha}{2}, n-1\right)}\right) = 1-\alpha$$

و بتقسيم كافة أطراف المتباينة السابقة على $(n-1).S^2$ ثم نأخذ مقلوب هذه المتباينة المزدوجة ، نصل إلى ما يلي :

$$prob\left(\frac{(n-1)S^2}{\chi^2_{\left(1-\frac{\alpha}{2}, n-1\right)}} \prec \delta_X^2 \prec \frac{(n-1)S^2}{\chi^2_{\left(\frac{\alpha}{2}, n-1\right)}}\right) = 1-\alpha$$

وهذا يعني أنه بالنسبة إلى مجال الثقة للتباين المجهول δ_X^2 في مجتمع إحصائي فإن :

الحد الأدني هو $\dfrac{(n-1)S^2}{\chi^2_{\left(1-\frac{\alpha}{2}, n-1\right)}}$ و الحد الأعلى هو $\dfrac{(n-1)S^2}{\chi^2_{\left(\frac{\alpha}{2}, n-1\right)}}$

ومنه فإن مجال الثقة للإنحراف المعياري هو :

$$prob\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)}}} \prec \delta_X \prec \sqrt{\frac{(n-1)S^2}{\chi^2_{\left(\frac{\alpha}{2},n-1\right)}}}\right)=1-\alpha$$

مجال الثقة للنسبة

إذا كان المطلوب هو تقدير النسبة p فإننا نفترض أن المجتمع له التوزيع الثنائي و المشكلة هي تقدير معلمة هذا التوزيع p

باعتماد f_n و الذي يمثل النسبة في العينة المسحوبة من المجتمع محل الدراسة و الخاضع للتوزيع الثنائي و ذلك ضمن مجال محدد بمعامل الثقة $1-\alpha$ أي احتمال وقوع تلك القيمة p ضمن المجال (T_1,T_2) أي: $prob\left[p\in(T_1,T_2)\right]=1-\alpha$

نعلم أنه و في حالة العينات الكبيرة يمكن تقريب التوزيع الثنائي بالتوزيع الطبيعي .

فإذا كان المتغير العشوائي (عدد مرات النجاح في العينة) هو X و كانت العينة كبيرة و حجمها n فإن $E(X)=\mu_X=n.p$

ومنه $E\left(\frac{X}{n}\right)=E(\hat{p})=\frac{1}{n}E(X)=\frac{1}{n}.n.p=p$

كذلك $V(X)=n.p.q$

ومنه $V\left(\frac{X}{n}\right)=V(\hat{p})=\frac{1}{n^2}V(X)=\frac{1}{n^2}.n.p.q=\frac{p.q}{n}$

و إذا كانت n كبيرة و و فقا لنظرية النهاية المركزية فإن التوزيع التقريبي ل X هو :

$Z=\frac{(X-n.p)}{\sqrt{n.p.q}}$ له تقريبا توزيع طبيعي معياري و بالتالي فإن :

$$prob\left(-Z_{\left(1-\frac{\alpha}{2}\right)} \prec Z \prec Z_{\left(1-\frac{\alpha}{2}\right)}\right) = 1-\alpha$$

و بالتعويض بقيمة Z نحصل على $prob\left(-Z_{\left(1-\frac{\alpha}{2}\right)} \prec \frac{(X-n.p)}{\sqrt{n.p.q}} \prec Z_{\left(1-\frac{\alpha}{2}\right)}\right) = 1-\alpha$

و بالتالي فإن $prob\left(f_n - Z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{p.q}{n}} \prec p \prec f_n - Z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{p.q}{n}}\right) = 1-\alpha$

رغم توفر المعلومات عن f_n و n فإنه لا يمكن إيجاد قيمتي المجال و ذلك لعدم معرفتنا

القيمة p و q (وهما مجهولان) الموجودة في الإنحراف المعياري $\delta_p^2 = \frac{p.q}{n}$ و لهذا يجب البحث عن قيمة p هذه .

في حقيقة الأمر هناك عدة طرق تستخدم من أجل إزالة هذه الصعوبة نذكر أبسط طريقة و هي طريقة تقريب p بواسطة f_n

نحن نعلم بأن f_n يمكن أن تستعمل كمقدر فعال و بدون تحيز ل p و هذه الخواص هي التي تقودنا إلى تقدير p (نسبة المجتمع) ضمن المجال المذكور أعلاه و بناءا على ذلك فإنه من المنطقي أن نستعمل f_n أيضا من أجل تقدير p و من ثم q الموجودة تحت الجذر أي في عبارة

الإنحراف المعياري $\sqrt{\frac{p.q}{n}}$ فإن الخطأالمرتكب سيكون صغيرا جدا من أجل قيمة كبيرة ل n

و يجدر الإنتباه إلى أن تغير $\sqrt{p.q}$ طفيف جدا من أجل قيم p قريبة من0.5 لذلك فإنه

وبإحلال التقدير النقطي $f_n = \frac{X}{n}$ مكان p و إدخال f_n داخل الجذر التربيعي يمكننا من الوصول إلى مجال الثقة للنسبة p كما يلي:

$$prob\left(f_n - Z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{f_n\left(1-f_n\right)}{n}} \prec p \prec f_n - Z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{f_n\left(1-f_n\right)}{n}}\right)=1-\alpha$$

و بالتالي فإن

الحد الأدني هو $f_n - Z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{f_n\left(1-f_n\right)}{n}}$ و الحد الأعلى هو $f_n + Z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{f_n\left(1-f_n\right)}{n}}$

و هما حدا مجال الثقة للوسيط p

إن المجال السابق و هو $(1-\alpha)\times100\%$ مجال الثقة ل p محسوبا من عينة عشوائية حجمها n حيث $(n \geq 30)$ تم سحبها من مجتمع ثنائي .

ملاحظة :

بفرض أن عملية السحب كانت تتم بدون إعادة و من مجتمع صغير (منته) ففي هذه الحالة يجب أن نأخذ مؤشر عدم الإعادة بعين الإعتبار $\left(\frac{N-n}{N-1}\right)$ و عندها فإن العلاقة السابقة تصبح كما يلي

$$prob\left(f_n - Z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{f_n\left(1-f_n\right)}{n}}\sqrt{\left(\frac{N-n}{N-1}\right)} \prec p \prec f_n - Z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{f_n\left(1-f_n\right)}{n}}\sqrt{\left(\frac{N-n}{N-1}\right)}\right)=1-\alpha$$