الإحصاء والإحتمالات- المحاضرة السادسة

Statistics and probabilities- Lecture

Dr Fadi KHALIL

Doctor lecturer in statistics and programing

في هذه المحاضرة سيتم التطرق لـ

- مفاهيم إحصائية مهمة: نظرية النهاية المركزية Central limit theorem:

- مفهوم مجال الثقة Interval of confidence.

- فكرة اختبار الفرضيات Hypothesis testing

  - الاختبار ثنائي الاتجاه Two tailed test
  - الاختبار أحادي الاتجاه One tailed test

مفاهيم إحصائية مهمة: نظرية النهاية المركزية Central limit theorem:

تنص نظرية النهاية المركزية على أنّ مجموع عدد كبير من المتغيرات العشوائية المستقلة عن بعضها والخاضعة لنفس التوزيع الإحتمالي تقترب نحو التوزيع الطبيعي، وبشكل أكثر تحديداً:

بفرض X_n ،..... X_2، X_1 تمثل n متغير عشوائي مستقلة عن بعضها وتخضع لنفس التوزيع الإحتمالي ولها متوسط μ وتباين σ^2 فإنّ توزيع المقدار، $X_1 + X_2 + \cdots X_n$ يخضع للتوزيع الطبيعي $N\sim(n\mu, n\sigma^2)$، بالتالي إن توزيع المقدار :

$$\frac{(X_1+X_2+\cdots X_n)-n\mu}{\sigma\sqrt{n}} \sim N(0,1)$$

وكذلك إنّ المقدار $\bar{X}=\frac{X_1+X_2+\cdots X_n}{n}$ يخضع للتوزيع الطبيعي $N\sim(\mu, \frac{\sigma^2}{n})$، أي:

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

مثال :

يعتقد المهندسون المدنيون أن مقدار الوزن W (بوحدات 1000 رطل) يعبر عن مقدار الوزن الذي يمكن أن تتحمله مسافة معينة من الجسر دون حدوث أضرار هيكلية ، يتم توزيعه عادةً بمتوسط 400 وانحراف معياري 40. ومع افتراض أن الوزن (في وحدة 1000 رطل) من السيارة متغير عشوائي بمتوسط 3 وانحراف معياري 0.3 فكم عدد السيارات التي يجب أن تكون على طول الجسر حتى يتجاوز احتمال الضرر الهيكلي 0.1؟

بفرض أن وزن السيارة الواحدة هو متغير عشوائي X_i ، بالتالي إنّ الاحتمال :

$$P\left(\sum_{i=1}^{n} X_i > W\right)$$

$$P\left(\sum_{i=1}^{n} X_i - W > 0\right)$$

يعبر عن احتمال حدوث أضرار هيكلية في الجسر،

وأنّ $\sum_{i=1}^{n} X_i$ وفقاً لنظرية النهاية المركزية يخضع لتوزيع طبيعي بمتوسط $3n$ وتباين 0.09n ،

وكذلك المقدار $\sum_{i=1}^{n} X_i - W$ يخضع للتوزيع الطبيعي بمتوسط: $3n - 400$ ، وتباين $0.09n + 1600$

بالتالي :

$$P\left(\frac{\sum_{i=1}^{n} X_i - W - (3n-400)}{\sqrt{0.09n+1600}} > \frac{0-(3n-400)}{\sqrt{0.09n+1600}}\right)$$

$$P\left(\text{z} > \frac{-3n+400}{\sqrt{0.09n+1600}}\right)$$

بحيث أنّ Z تخضع للتوزيع الطبيعي المعياري ، ووفقاً لمعطيات المسألة،

$$P\left(\text{z} > \frac{3n+400}{\sqrt{0.09n+1600}}\right) = 0.1$$

ومن جدول التوزيع الطبيعي المعياري : إن القيمة التي احتمالها 0.1 هي القيمة 1.28،

أي انّ $\frac{-3n+400}{\sqrt{0.09n+1600}} = 1.28$

أي انّ n تقريباً 117. بالتالي عند تجاوز عدد السيارات 117 هناك فرصة 0.1 لأن يحدث ضرر هيكلي في الجسر.

مثال :

قرر عالم فلك قياس المسافة بين مرصده و نجم بعيد. ولكن بسبب الأحوال الجوية ، فإن أي قياس لن يعطي المسافة الدقيقة d. نتيجة لذلك ، قرر الفلكي إجراء سلسلة من القياسات ثم استخدام قيمة المتوسط كتقدير للمسافة الفعلية بين المرصد والنجم البعيد. إذا كان الفلكي يعتقد أن قيم القياسات المتتالية هي متغيرات عشوائية مستقلة بمتوسط d سنة ضوئية وانحراف معياري لسنتين ضوئيتين ، فكم عدد القياسات التي يجب أن يجعله متأكداً بنسبة 95٪ على الأقل من أن دقة تقديره في حدود ± 0.5 سنة ضوئية؟

الحل :

على فرض أن القياسات المأخوذ للمسافة هي متغيرات عشوائية مستقلة X_i ، بالتالي إنّ $\bar{X}$ يخضع للتوزيع الطبيعي بمتوسط d وتباين $\frac{4}{n}$

$$P(-0.5 \leq \bar{X} - d \leq +0.5)$$

$$P\left(\frac{-0.5}{2/\sqrt{n}} \leq \frac{\bar{X}-d}{2/\sqrt{n}} \leq \frac{+0.5}{2/\sqrt{n}}\right)$$

$$= 2P(\text{z} \leq +\sqrt{n}/4) - 1$$

ومن معطيات المسألة :

$$2P(\text{z} \leq +\sqrt{n}/4) - 1 \geq 0.95$$

$$P(\text{z} \leq +\sqrt{n}/4) \geq 0.975$$

في التوزيع الطبيعي المعياري $P(Z \leq 1.96) = 0.975$

بالتالي $\frac{\sqrt{n}}{4} = 1.96$

أي على الأقل يجب أن يكون عدد القياسات 62 قياس للحصول على الدقة المطلوبة.

10- مجال الثقة *Interval of confidence*:

مجال الثقة هو شكل من أشكال تقدير معالم المجتمع. وعموماً هناك نوعين من أنواع التقدير:

a. التقدير النقطي.

b. التقدير المجالي.

يعتمد التقدير النقطي، كما رأينا في فقرة المؤشرات الإحصائية، على حساب قيمة محددة لتقدير أحد معالم المجتمع اعتماداً على مؤشرات النزعة المركزية أو التشتت في حين يعتمد التقدير المجالي على حساب مجال من الثقة حول التقدير النقطي بحيث يعطي معلومة أدق عن معلمة المجتمع.

مجال الثقة يعبر عن مجال تقديري من القيم التي يقع ضمنها قيمة معلمة المجتمع حسب احتمال محدد مسبقاً وكلما كان مجال الثقة كبيراً كلما كان احتمال وقوع قيمة معلمة المجتمع الحقيقية ضمن هذا المجال كبيراً. إضافة إلى ذلك، غالباً ما يرتبط بمجال الثقة قيمة احتمالية تعبر عن احتمال وقوع قيمة معلمة المجتمع الحقيقية ضمنه.

يسمى هذا المجال بمجال الثقة لأنّه يترافق بدرجة من الثقة المفترضة حول المقدّر المحسوب. ويعطى نصف طول المجال كما يلي:

$$\bar{x} \pm Z_{\alpha/2} * {}^{\sigma}/_{\sqrt{n}}$$

بعبارة أخرى هو التقدير النقطي المحسوب $(\bar{x})$ اعتماداً على بيانات العينة مضافاً إليه (أو مطروحاً منه) عدد معين $(Z_{1-\alpha/2})$ من الأخطاء المعيارية $({}^{\sigma}/_{\sqrt{}})$. من المفترض أنّ كل من التقدير النقطي و الخطأ المعياري يمكن حسابهما انطلاقاً من بيانات العينة. أمّا عدد الأخطاء المعيارية فيتم حسابه تبعاً لنقطتين:

a. التوزيع الاحتمالي المفترض للتقدير النقطي.

b. درجة الثقة $(1-\alpha)$ أو الاحتمال المرغوب لوقوع التقدير النقطي ضمن المجال المحسوب.

في هذا السياق، من المتعارف عليه إحصائياً وجود العديد من مجالات الثقة التي تختلف عن بعضها حسب عدد الاخطاء المعيارية المأخوذة بعين الاعتبار، بحيث يكون كل خطأ معياري مرتبط بقيمة احتمالية معينة (0,01 – 0,05 – 0,10).

فمثلاً يمكن القول أن متوسط هطول الأمطار في المنطقة الساحلية هذا العام تتراوح بين 1 ملم و 1,5 ملم باحتمال ثقة $1-\alpha$ قدره 95% أي باحتمال خطأ α قدره 5%.

لتفسير مجال سيتم تناول الحالة الأبسط لمجال لثقة وهو مجال الثقة حول متوسط المجتمع μ:

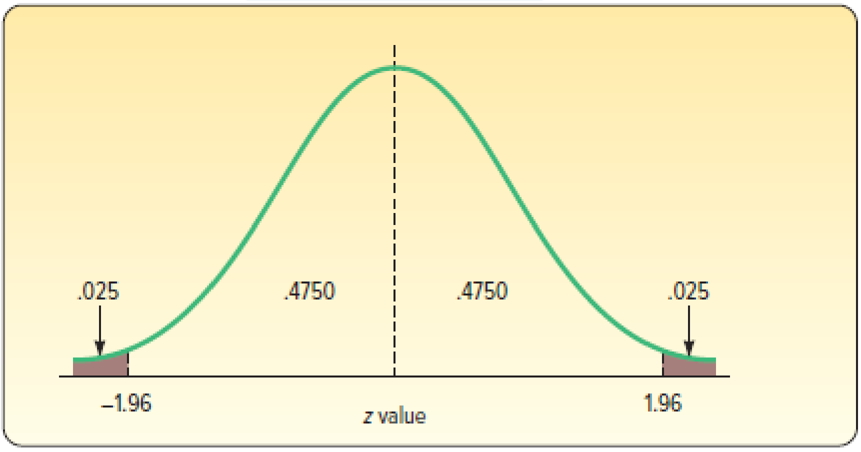
$$p\left[\bar{x} - Z_{\frac{\alpha}{2}} * {}^{\sigma}/_{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2} * {}^{\sigma}/_{\sqrt{n}}\right] = 1-\alpha$$

أي أن التقدير النقطي لمتوسط العينة $\bar{x}$ يقع في مجال، حده الأدنى $\bar{x} - Z_{\alpha/2} * {}^{\sigma}/_{\sqrt{n}}$ وحده الأعلى $\bar{x} + Z_{\alpha/2} * {}^{\sigma}/_{\sqrt{n}}$، باحتمال ثقة قدره 95% ، و احتمال أن يقع خارجه باحتمال قدره 5%.

وتجدى الإشارة إلى أن التقدير النقطي إما قد يكون لأحد مؤشرات النزعة المركزية والتشتت، أو عبارة عن الفرق بين مقدّرين من عينتين مختلفتين (مستقلتين أو مرتبطتين)، أو يأخذ الشكل النسب. و لكن مجالات الثقة حول هذه المقاييس تتشابه بالمبدأ وتختلف عن بعضها بطريقة الحساب و طريقة تقدير الخطأ المعياري.

بعد التمهيد السابق، لا بدّ من بيان كيفية استخراج المقدار $Z_{\alpha/2}$ ، فمثلاً عند مقدار ثقة 95% (أي عند احتمال وقوع قيمة المجتمع الحقيقية ضمن المجال هو 95%) أي $\alpha = 0.05$ وبالتالي قيمة $Z_{\alpha/2} = Z_{0.05/2} = Z_{0.025} = 1.96$

وهي القيمة الموضحة في الشكل الآتي :



أي أنّ الجزء المتبقي من المساحة تحت المنحنى هو $1-\alpha = 0.95$ وبما إنّ الجدول المعتمد في هذا المقرر يحسب المساحة المحصورة بين القيمة المعيارية وقيمة $Z = 0$.

أي للتبسيط يتم قسمة :

$$\frac{0.95}{2} = 0.4750$$

بعد ذلك يتم البحث في جدول التوزيع الطبيعي المعياري على القيمة المقابلة لاحتمال 0.4750:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884

وهي القيمة المظللة باللون الأصفر والتي تقابل قيمة 1.96 ، بعبارة أخرى إنّ قيمة $Z_{\alpha/2} = 1.96$ وبعدها يتم التعويض في مجال الثقة.

مثال :

بفرض أنّ الإشارة المرسلة من مكان A إلى مكان B تخضع للتوزيع الطبيعي بمتوسط μ وتباين $\sigma^2 = 4$، ولتقليل الخطأ في استقبال الإشارة تم إرسال الإشارة لتسع مرات متتالية، فإذا كانت الإشارات المستقبلة هي :

5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5

فما هو مجال التقدير الممكن بناؤه باحتمال ثقة 0.95حول المتوسط μ؟

مجا ل الثقة حول المتوسط باحتمال ثقة 0.95 يُعطى بالصيغة الآتية :

$$p\left[\bar{x} - Z_{\frac{\alpha}{2}} * {}^{\sigma}/_{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2} * {}^{\sigma}/_{\sqrt{n}}\right] = 0.95$$

إنّ قيمة $\bar{x}$ للعينة تحسب كما يأتي :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{9} x_i}{9} = \frac{5+8.5+12+15+7+9+7.5+6.5+10.5}{9} = 9$$

إنّ قيمة الخطأ المعياري $ {}^{\sigma}/_{\sqrt{n}} = \frac{2}{3}$

وكما تم بيانه سابقاً إنّ قيمة $1 - \alpha = 0.95$ أي $\alpha = 0.05$ وبالتالي $Z_{0.05/2} = Z_{0.025} = 1.96$

وبالتعويض :

$$p\left[9 - 1.96 * \frac{2}{3} < \mu < 9 + 1.96 * 2/3\right] = 0.95$$

$$p[7.69 < \mu < 10.31] = 0.95$$

أي أنّ المتوسط μ الحقيقي يقع ضمن مجال حده الأعلى 10.31 وحده الأدنى 7.69 باحتمال قدره 0.95، ويقع خارجه باحتمال قدره 0.05

ولإنشاء مجال حول المتوسط μ باحتمال ثقة 0.90 يتم اتباع نفس الخطوات ولكن الاختلاف فقط في إيجاد $Z_{\alpha/2}$ ،
وبما إنّ الاحتمال هو $1-\alpha = 0.90$ أي $\alpha = 0.1$ وبالتالي $Z_{0.1/2} = Z_{0.5} = 1.65$

وكذلك للتبسيط يتم قسمة :

$$\frac{0.90}{2} = 0.45$$

وبالبحث على القيمة المقابلة لهذا الاحتمال في الجدول :

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884

وهي تقريباً القيمة المقابلة لاحتمال 0.4505 وهي 1.65، وبالتالي يتم التعويض في مجال الثقة:

$$p\left[9 - 1.65 * \frac{2}{3} < \mu < 9 + 1.65 * 2/3\right] = 0.9$$

$$p[7.9 < \mu < 10.1] = 0.9$$

أي أنّ المتوسط μ الحقيقي يقع ضمن مجال حده الأعلى 10. 1 وحده الأدنى 7.9 باحتمال قدره 0.9، ويقع خارجه باحتمال قدره 0.1.

ملاحظة: فيما سبق تمّ بناء مجال ثقة ثنائي الجانب (أصغر من قيمة ما، وأكبر من قيمة ما) ولكن في بعض الأحيان قد يكون المطلوب هو إيجاد مجال من جانب واحد ، مثلاً أن يكون المتوسط أكبر من قيمة ما باحتمال قدره 0.95. لبناء هذا النوع من المجالات:

$$p\left[\bar{x} - Z_{\alpha} * {}^{\sigma}/_{\sqrt{n}} < \mu\right] = 0.95$$

بحيث $1.65 = Z_{0.05} = Z_{\alpha}$ ، بمعنى آخر هنا لا داعي لقسمة 0.95 على 2 بل يتم البحث عن القيمة الاحتمالية

$$0.95 - 0.5 = 0.45$$

وذلك لأنّ التوزيع الطبيعي متناظر والجدول السابق يعطي المساحة بين قيمة معيارية والقيمة $Z = 0$ ، ومن الجدول يلاحظ أنّ القيمة المقابلة لاحتمال 0.45 هي 1.65

$$p\left[9 - 1.65 * \frac{2}{3} < \mu\right] = 0.95$$

$$p[7.9 < \mu] = 0.95$$

أي أنّ احتمال أن ينتمي μ للمجال $[7.9, \quad \infty]$ هو 0.95.

مثال :

أراد قسم الجودة في شركة *Siemens* دراسة مستوى التقلبات *variability* في أداء توربين *turbine* جديد يعمل على الهيدروجين *hydrogen*. من أجل ذلك رغب في تقدير الحد الأعلى والأدنى لمتوسط تدفق الغاز من التوربين، واتضح من خلال عينة مؤلفة من 9 قياسات أنّ متوسط تدفق الغاز هو 16 m^3/h بانحراف معياري 0.25، **والمطلوب**:

1- ما هو التقدير النقطي لمتوسط تدفق الغاز في التوربين .

2- مساعدة قسم الجودة في تقدير مجال الثقة الذي من المتوقع أن يقع ضمنه متوسط تدفق الغاز عند احتمال ثقة 99%.

3- تقدير الحد الأنى لأقل قيمة لمتوسط تدفق الغاز الحقيقي عن احتمال ثقة 95%.

ملاحظة: يمكنك الاستعادنة بإحدى القيم المعيارية *critical values* في الجدول الآتي :

مستوى الدلالة significance level	الاختبار ثنائي الجانب Two sided test	الاختبار أحادي الجانب One sided test
0.01	2.56	2.32
0.05	1.96	1.64
0.1	1.64	1.28

1- ما هو التقدير النقطي لمتوسط تدفق الغاز في التوربين .

التقدير النقطي لمتوسط تدفق الغاز هو نفسه متوسط العينة أي 16 m^3/h

2- مساعدة قسم الجودة في تقدير مجال الثقة الذي من المتوقع أن يقع ضمنه متوسط تدفق الغاز عند احتمال ثقة 99%.

المطلوب بناء مجا ل الثقة ثنائي الجانب حول المتوسط باحتمال ثقة 0.99 يُعطى بالصيغة الآتية :

$$p\left[\bar{x} - Z_{\frac{\alpha}{2}} * {}^{\sigma}/_{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2} * {}^{\sigma}/_{\sqrt{n}}\right] = 0.95$$

يتم اختيار $Z_{\frac{\alpha}{2}} = 2.56$

بالتعويض:

$$p\left[16 - 2.56 * {}^{0.25}/_{\sqrt{9}} \leq \mu < 16 + 2.56 * {}^{0.25}/_{\sqrt{9}}\right] = 0.99$$

$$p[15.787 < \mu < 16.213] = 0.95$$

أي ان متوسط تدفق الغاز يتروا‌ح ضمن حد أدنى مقداره 15.787 وحد اعلى مقداره 16.213 باحتمال قدره 99% و ويقع خارجه باحتمال 1% .

3- تقدير الحد الأنى لأقل قيمة لمتوسط تدفق الغاز الحقيقي عن احتمال ثقة 95%.

المطلوب هنا بناء مجال الثقة من جانب واحد يتم اختيار القيمة المعيارية Z_α وهي $Z_{0.05}$ أي 1.64 وبالتعويض :

$$p\left[16 - 1.64 * {}^{0.25}/_{\sqrt{9}} \leq \mu < +\infty\right[= 0.95$$

$$p[15.863 < \mu < \infty[= 0.95$$

أي انّ الحد الأدنى لمتوسط تدفق الغاز هو 15.863 باحتمال قدره 0.95 واحتال ان يكون أقل من هذا المستوى هو 0.05.