الفصل الرابع

افتراض نموذج احتمال من بيانات العينة

Assuming a Probability Model From the Sample Data

الآن بعد أن جمعنا البيانات، ورسمنا المدرج التكراري، وقدرنا المقاييس للنزعة المركزية والتغيّر، مثل المتوسط والوسيط والانحراف المعياري، فإننا جاهزين لافتراض نموذج احتمال للمجتمع أو العملية التي حصلنا منها على العينات. عند هذه النقطة، نقوم بافتراض تقريبي باستخدام مقادير بسيطة للمتوسط، والوسيط، والانحراف المعياري، والمدرج التكراري. ولكن من المهم ملاحظة أن هناك اختبارات أكثر صرامة، مثل اختبار χ^2 للطبيعية [7] لتحديد ما إذا كان نموذج احتمال خاص مناسبًا للافتراض من مجتمع لبيانات العينة.

بمجرد افتراضنا لنموذج الاحتمال المناسب، يمكننا اختيار الاختبارات الإحصائية المناسبة التي ستتيح لنا اختبار الفروض واستخلاص النتائج بمستوى معين من الثقة. وسيحدد نموذج الاحتمال مستوى الثقة لدينا عند قبول أو رفض فرضية ما.

هناك سؤالان أساسيان نحاول معالجتهما عند افتراض نموذج احتمال لمجموعتنا الأساسية:

١- ما هو مقدار ثقتنا بأن إحصائيات العينة تمثل المجتمع الإحصائي الكامل؟

٢- هل الاختلافات في الإحصائيات بين مجتمعين كبيرين، وناتجة عن عوامل أكثر منها عن الصدفة فقط؟

نحتاج، من أجل تأكيد أي مستوى "الثقة" عند القيام بالاستنتاج الإحصائي، إلى نموذج رياضي يصف احتمال إمكانية حدوث أي قيمة بيانات. وتُسمى هذه النماذج توزيعات الاحتمال.

هناك عدد من نماذج الاحتمال التي يتم افتراضها تكراراً لوصف العمليات البيولوجية. على سبيل المثال، عند وصف تغيّر معدل ضربات القلب، فإن احتمال ملاحظة فترة زمنية محددة بين ضربات القلب المتتالية يمكن وصفه عن طريق التوزيع الأسي [1, 8]. يوضح الشكل (٣،٦) في الفصل الثالث مدرجا تكرارياً (رسماً بيانياً) لعينات مأخوذة من توزيع أسي. لاحظ أن هذا التوزيع مائل بشدة إلى اليمين. وبالنسبة للفواصل R-R، فإن دالة الاحتمال هذه منطقية من الناحية الفسيولوجية لأن خلايا القلب الفردية لها فترة استعصاء تمنعها من التقلص في أقل من فترة زمنية دنيا. ومع ذلك، قد تحدث فترة زمنية طويلة جداً بين الضربات، مما يؤدي إلى ظهور بعض الفترات الزمنية الطويلة بشكل غير منتظم.

إن نموذج الاحتمال الأكثر افتراضاً لمعظم التطبيقات العلمية والهندسية هو التوزيع الطبيعي أو توزيع جاوس (normal or Gaussian distribution). وهذا التوزيع موضَّح بالخط الأسود المتصل في الشكل (٤،١) وغالباً ما يُشار إليه بمنحني الجرس لأنه يشبه الجرس الموسيقي.

إن المعادلة التي تعطي الاحتمال، $f(x)$، لملاحظة قيمة معينة لـ x من المجتمع الإحصائي هي:

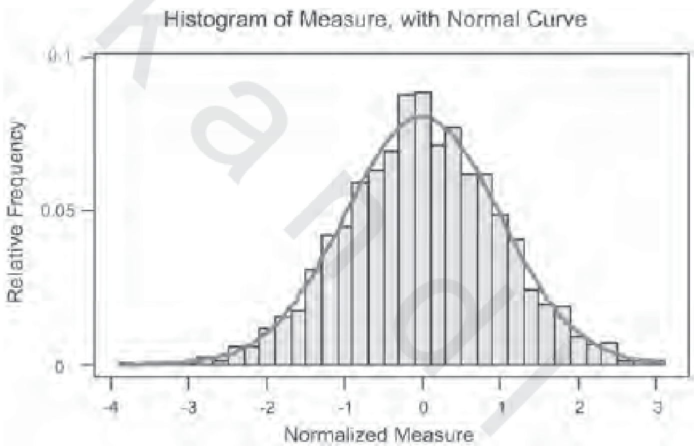
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad \infty < x < \infty$$

حيث μ هو المتوسط الحقيقي للمجتمع أو العملية و σ هو الانحراف المعياري لنفس المجتمع الإحصائي أو العملية. والرسم بياني لهذه المعادلة مُعطى مُوضَّحاً بالمنحنى المتصل الأملس في الشكل (٤,١). المساحة تحت المنحنى تساوي واحد.

لاحظ أن التوزيع الطبيعي هو:

١- منحني متناظر على شكل جرس موصوف تماماً بواسطة، متوسطه، μ، وانحرافه المعياري، σ.

٢- من خلال تغيير μ و σ، نمد ونزيح التوزيع.

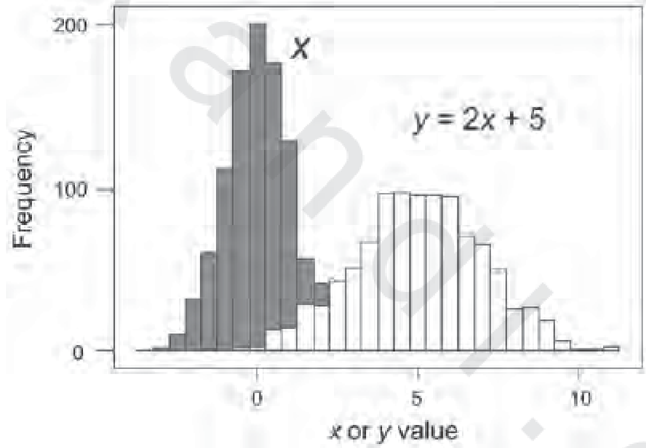


الشكل (١ , ٤). يتم هنا توضيح المدرج التكراري لـ 1000 عينة مأخوذة من توزيع طبيعي. وتم وضع المنحنى الطبيعي المثالي الممثل لدالة توزيع الاحتمال الطبيعي على المدرج التكراري.

يوضح الشكل (٤,١) أيضاً مدرجا تكرارياً يتم الحصول عليه عندما نختار عشوائياً 1000 عينة من المجتمع الإحصائي الموزعة بشكل طبيعي، ولها متوسط قيمته صفر تباين (variance) قيمته 1. ومن المهم الاعتراف أنه عند زيادة حجم العينة n، فإن المدرج التكراري يقترب من التوزيع الطبيعي المثالي المُبين بالخط المتصل الأملس. ولكن، عند أحجام عينة صغيرة، قد يبدو المدرج التكراري مختلفاً جداً عن المنحنى

الطبيعي. وهكذا، قد يكون من الصعب، من أحجام عينة صغيرة، تحديد ما إذا كان النموذج المُفترض مناسباً للمجتمع أو العملية، وأن أية اختبارات إحصائية نقوم بها قد لا تسمح لنا باختبار الفروض واستخلاص النتائج بأي مستوى من الثقة الحقيقية.

يمكننا القيام بعمليات خطية على المتغير العشوائي المُوزَّع بشكل طبيعي، x، لإنتاج متغير عشوائي آخر مُوزَّع بشكل طبيعي، y. تشمل هذه العمليات ضرب x بثابت وإضافة ثابت (تعويض) إلى x. يوضح الشكل (٤،٢) المدرجات التكرارية لعينات مأخوذة من كل من المجموعتين x و y. نلاحظ أن توزيع y مُزاح (المتوسط يساوي الآن 5)، وأن التغيّر قد ازداد بالنسبة إلى x.



الشكل (٢، ٤). مدرجات تكرارية مُبينة لعينات مأخوذة من مجتمعين x و y، حيث y هو مجرد دالة خطية لـ x. لاحظ أن المتوسط والتغيّر لـ y يختلف عن x، ولكن كليهما توزيع طبيعي.

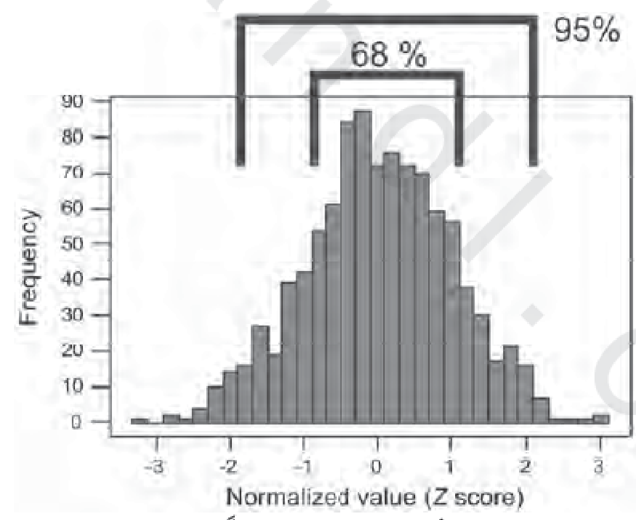
إن أحد الاختبارات التي قد نستخدمها لتحديد مدى جودة ملاءمة نموذج الاحتمال الطبيعي للبيانات المتوفرة لدينا هو عدد العينات التي تندرج ضمن الانحرافين المعياريين ±1 و ±2 للمتوسط. إذا كانت البيانات والمجتمع الإحصائي أو العملية مُنمذجة جيداً بواسطة التوزيع الطبيعي، فإن 68% من العينات ينبغي أن يقع ضمن الانحراف

المعياري ±1 عن المتوسط و 95% من العينات ينبغي أن تقع ضمن الانحرافات المعيارية ±2 عن المتوسط. هذه النسب مُوضَّحة في الشكل (٤،٣). ومن المهم تذكر هذه الأرقام القليلة، لأننا سوف نستخدم في كثير من الأحيان مجال الـ 95% عند استخلاص النتائج من التحليل الإحصائي.

هناك وسيلة أخرى لتحديد مدى جودة تمثيل بياناتنا التي تم أخذ عيناتها، x، لتوزيع طبيعي وهي تقدير معامل بيرسون للانحراف (عدم التماثل) (PCS) [5]. ويُعطى معامل الانحراف بواسطة المعادلة التالية:

$$PCS = \frac{3\left|\overline{x} - x_{median}\right|}{s}$$

إذا كان $PCS > 0.5$، فإننا نفترض أن العينات لم تكن مأخوذة من مجتمع مُوزَّع بشكل طبيعي.



الشكل (٣ ,٤). مدرج تكراري لعينات مأخوذة مجتمع إحصائي مُوزَّع بشكل طبيعي. وبالنسبة للتوزيع الطبيعي، ينبغي لـ 68% من العينات أن تقع ضمن الانحراف المعياري ±1 عن المتوسط. (صفر في هذه الحالة) و 95% من العينات ينبغي أن تقع ضمن الانحرافات المعيارية ±2 (1.96 على وجه الدقة) عن المتوسط.

عند جمع البيانات، فإنه عادة ما يتم جمعها بالعديد من الأنواع المختلفة من الوحدات الفيزيائية (فولت، مئوية، نيوتن، سنتمتر، غرام، ...إلخ). ومن أجل استخدام الجداول التي تم وضعها لنماذج احتمالات، فهناك حاجة إلى تنظيم البيانات بحيث يكون للبيانات المُنتظمة متوسط قيمته صفر، وانحراف معياري قيمته 1. ويُسمى مثل هذا التوزيع الطبيعي بالتوزيع الطبيعي المعياري وهو مُوضَّح في الشكل (٤.١).

التوزيع الطبيعي المعياري له توزيع متماثل على شكل جرس وفيه $\mu=0$ و $\sigma=1$.

لتحويل البيانات المُوزَّعة بشكل طبيعي إلى قيمة طبيعية معيارية، فإننا نستخدم الصيغ التالية:

$$z=(x-\mu)/\sigma \text{ أو } z=(x-\bar{x})/s,$$

وذلك اعتماداً على ما إذا كنا نعرف المتوسط الحقيقي، μ، والانحراف المعياري، α، أو ليس لدينا سوى تقديرات العينة، $\bar{x}$ أو s.

لأجل أية عينة فردية أو نقطة بيانات، x_i، من عينة ذات متوسط، $\bar{x}$، وانحراف معياري، s، يمكننا تحديد نقاطها z من الصيغة التالية:

$$z_i=\frac{x_i-\bar{x}}{s}.$$

وبالنسبة لعينة فردية، فإن النقاط z هي قيمة "منتظمة" أو "قياسية". يمكننا استخدام هذه القيمة مع معادلات دالة كثافة الاحتمال أو جداول الاحتمال القياسية [3] لتحديد احتمال ملاحظة مثل قيمة عينة من المجتمع.

يمكن التفكير أيضاً بالنقاط z بأنها مقياس لمسافة العينة الفردية، x_i، عن متوسط العينة، $\bar{x}$، في وحدات الانحراف المعياري. على سبيل المثال، إذا كانت لنقطة عينة، x_i z نقطة مقدارها $z_i=2$، فهذا يعني أن نقطة البيانات، x_i، تبعد انحرافين معياريين عن متوسط العينة.

إننا نستخدم نقاط z منتظمة بدلاً من البيانات الأصلية عند إجراء التحليل الإحصائي ؛ لأن جداول البيانات المنتظمة يتم تنظيمها بالفعل ومتوفرة في معظم كتب الإحصاء أو حزم البرامج الإحصائية. بالإضافة إلى ذلك، وباستخدام القيم المنتظمة، ليس هناك داعٍ للقلق بشأن المطال المطلق للبيانات أو الوحدات المستخدمة لقياس البيانات.

(٤,١) التوزيع الطبيعي المعياري
THE STANDARD NORMAL DISTRIBUTION

يوضح الجدول (٤.١) التوزيع الطبيعي المعياري.

يوفِّر الجدول z المرتبط بهذا الشكل مداخل الجدول الذي يعطي احتمال أن $z \leq a$ ، الذي يساوي المساحة تحت المنحنى الطبيعي إلى يسار $z = a$. إذا كانت البيانات تأتي من توزيع طبيعي، فإن الجدول يخبرنا عن احتمال أو "فرصة" قيمة العينة أو النتائج التجريبية التي لها قيمة أقل من أو تساوي a.

وهكذا، يمكننا أخذ أي عينة وحساب نقاطها z على النحو الموصوف سابقاً ومن ثم استخدم الجدول z لإيجاد احتمال ملاحظة قيمة z أقل من قيمة منتظمة محددة أو تساويها، a. على سبيل المثال، إن احتمال ملاحظة قيمة z أقل من أو تساوي 1.96 هو 97.5%. وهكذا، فإن احتمال ملاحظة قيمة z أكبر من 1.96 هو 2.5%. بالإضافة إلى ذلك، وبسبب التماثل في التوزيع، نعلم أن احتمال ملاحظة قيمة z أقل من -1.96 هو أيضاً 97.5%، واحتمال ملاحظة قيمة z أقل من أو تساوي -1.96 هو 2.5%. وأخيراً، فإن احتمال ملاحظة قيمة z بين -1.96 و 1.96 هو 95%. ينبغي للقارئ دراسة الجدول z والرسم البياني المرتبط به للتوزيع z للتحقق من أن الاحتمالات (أو المساحات الواقعة تحت دالة كثافة الاحتمال) المذكورة لاحقا صحيحة.

نحن بحاجة في كثير من الأحيان لتحديد احتمال بحيث تقع نتيجة تجريبية بين قيمتين أو تكون النتيجة أكبر من قيمة محددة a أو أقل أو أكبر من قيمة محددة b. ولإيجاد هذه المساحات، يمكننا استخدام الصيغ الهامة التالية، حيث Pr هو الاحتمال:

$$\Pr(a \le z \le b) = \Pr(z \le b) - \Pr(z \le a)$$

المساحة بين $z = a$ و $z = b$

$$\Pr(z \le a) = 1 - \Pr(z < a)$$

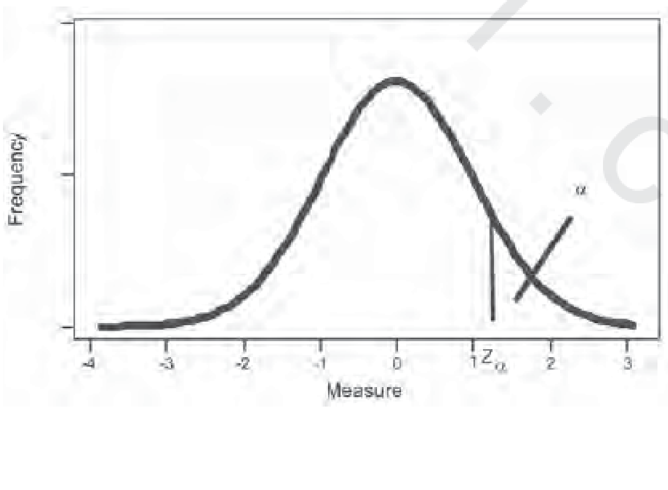
المساحة إلى يمين $z = a$

المساحة في الذيل الأيمن

وهكذا، من أجل أية ملاحظة أو قياس، x، من أي توزيع طبيعي:

$$\Pr(a \le x \le b) = \Pr\left(\frac{a-\mu}{\sigma} \le z \le \frac{b-\mu}{\sigma}\right),$$

حيث μ هو متوسط التوزيع الطبيعي و σ هو الانحراف المعياري للتوزيع الطبيعي.



الجدول (١ , ٤). دالة توزيع z معياري: المساحات تحت دالة الكثافة الطبيعية القياسية.

0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00	Z
0.5359	0.5319	0.5279	0.5239	0.5199	0.5160	0.5120	0.5080	0.5040	0.5000	0.0
0.5753	0.5714	0.5675	0.5636	0.5596	0.5557	0.5517	0.5478	0.5438	0.5398	0.1
0.6141	0.6103	0.6064	0.6026	0.5987	0.5948	0.5910	0.5871	0.5832	0.5793	0.2
										...
0.9633	0.9625	0.9616	0.9608	0.9599	0.9591	0.9582	0.9573	0.9564	0.9554	1.7
0.9706	0.9699	0.9693	0.9686	0.9678	0.9671	0.9664	0.9656	0.9649	0.9641	1.8
0.9767	0.9761	0.9756	0.9750	0.9744	0.9738	0.9732	0.9726	0.9719	0.9713	1.9
0.9817	0.9812	0.9808	0.9803	0.9798	0.9793	0.9788	0.9783	0.9778	0.9772	2.0
										...
0.9936	0.9934	0.9932	0.9931	0.9929	0.9927	0.9925	0.9922	0.9920	0.9918	2.4
0.9952	0.9951	0.9949	0.9948	0.9946	0.9945	0.9943	0.9941	0.9940	0.9938	2.5
0.9964	0.9963	0.9962	0.9961	0.9960	0.9959	0.9957	0.9956	0.9955	0.9953	2.6
										...
0.9990	0.9990	0.9989	0.9989	0.9989	0.9988	0.9988	0.9987	0.9987	0.9987	3.0

قيم z المُستخدمة بشكل شائع

Z_α	المساحة في الذيل الأيمن، α
1.282	0.1
1.645	0.05
1.96	0.025
2.326	0.010
2.576	0.005

وبعبارة أخرى، نحن بحاجة إلى معايرة أو إيجاد القيم z لكل من البارامترات، a و b، لإيجاد المساحة الواقعة تحت المنحنى الطبيعي المعياري (التوزيع z) التي تمثل التعبير على الجانب الأيسر من المعادلة المذكورة سابقاً.

مثال (٤,١)

إن متوسط امتصاص الدهون بالنسبة للذكور الذين تتراوح أعمارهم من 6 – 9 سنوات هو 28 غراما، بانحراف معياري مقداره 13.2 غرام. افترض أن الامتصاص مُوزع بشكل طبيعي. إن امتصاص ستيف (Steve) هو 42 غراما وامتصاص بن (Ben) هو 25 غراما.

١– ما هي نسبة المساحة بين امتصاص ستيف اليومي وامتصاص بن اليومي؟

٢– إذا أردنا عشوائياً اختيار ذكر يتراوح عمره بين 6 و 9 سنوات، ما هو احتمال أن يكون امتصاصه للدهون 50 غرام أو أكثر؟

الحل

x = امتصاص الدهون

١– يمكن القول إن المشكلة على النحو التالي: ما هو $\Pr(25 \le x \le 42)$ ؟

بافتراض أن التوزيع طبيعي، فإننا نتحول إلى نقاط z:

ما هو $\Pr(((25-28)/13.2) < z < ((42-28/13.2)))$ ؟

$$= \Pr(-0.227 \le z \le 1.06) = \Pr(z \le 1.06) - \Pr(z \le -0.227)$$

(وذلك باستخدام صيغة $\Pr(a \le z \le b)$)

$$= \Pr(z \le 1.06) - [1 - \Pr(z \le 0.227)] = 0.8554 - [1 - 0.5910].$$

$0.4464 \; or \; 44.6\%$ (من المساحة تحت المنحنى z)

٢– يمكن القول إن المشكلة على النحو التالي: "ما هو $\Pr(x > 50)$" ؟

بالمعايرة إلى النقطة z، ما هو $\Pr(z > (50-28/13.2))$ ؟

$$= \Pr(z > 1.67)$$

$$= 1 - \Pr(z \le 1.67) = 1 - 0.9525 = 0.0475,$$

$or \; 4.75\%$ من المساحة تحت المنحنى z.

مثال (٤,٢)

بفرض أن المواصفات على مجال إزاحة مُفرِّض (أداة تفريض) (indentor) طرف هو 0.001±0.5 ميللي متر. إذا كانت هذه الإزاحات مُوزَّعة بشكل طبيعي، ولها متوسط = 0.47 وانحراف معياري = 0.002، ما هي النسبة المئوية للمُفرِّضات الواقعة ضمن هذه المواصفات ؟

الحل

x = الإزاحة

يمكن القول إن المشكلة على النحو التالي: "ما هو $\Pr(0.499 \le x \le 0.501)$"؟

باستخدام نقاط z،

$$\Pr(0.499 \le x \le 0.501) = \Pr((0.499 - 0.47)/0.002 \le z \le (0.501 - 0.47/0.002))$$

$$= \Pr(14.5 \le x \le 15.5) = \Pr(z \le 15.5) - \Pr(z \le 14.5) = 1 - 1 = 0$$

ومن المفيد ملاحظة أنه إذا كان توزيع المجتمع الإحصائي وبيانات العينة المرتبطة بها ليس طبيعياً (غير متماثل)، فإنه يمكن في كثير من الأحيان استخدام التحويلات لجعل البيانات طبيعية، ويمكن بعد ذلك استخدام الإحصائيات التي يغطيها هذا الكتاب لإجراء التحليل الإحصائي على البيانات التي تم تحويلها. وتشمل هذه التحويلات على البيانات الخام اللوغاريتم، والجذر التربيعي، والمقلوب (المعكوس).

(٤,٢) التوزيع الطبيعي ومتوسط العينة

THE NORMAL DISTRIBUTION AND SAMPLE MEAN

إن جميع التحليل الإحصائي يتبع نفس الإجراء العام:

١ – افتراض توزيع أساسي للبيانات والبارامترات المرتبطة به (على سبيل المثال، متوسط العينة).

٢- تدريج البيانات أو البارامتر بالنسبة إلى توزيع "معياري".

٣- تقدير فترات ثقة باستخدام جدول معياري للتوزيع المُفترض. (والسؤال الذي نطرحه هو: "ما هو احتمال ملاحظة النتيجة التجريبية عن طريق الصدفة وحدها؟")

٤- إجراء اختبار الفرضية (على سبيل المثال، الاختبار t للطالب).

بدأنا هذا الكتاب وركزنا في معظمه على التوزيع الطبيعي أو نموذج الاحتمال بسبب انتشاره في المجال الطبي الحيوي وإلى حد ما يُسمى نظرية النزعة المركزية. إن أحد أكثر الاختبارات الإحصائية الأساسية التي نقوم بها هو المقارنة بين المتوسطات لمجتمعين أو أكثر. إن متوسط العينة هو في حد ذاته تقدير تم إيجاده من عدد محدود من العينات. وهكذا، فإن متوسط العينة، $\overline{x}$ ، هو في حد ذاته متغير عشوائي تتم نمذجته بتوزيع طبيعي [4].

هل هذا النموذج لـ $\overline{x}$ منطقي؟ الجواب نعم، لعينات كبيرة، بسبب نظرية النزعة المركزية، والتي تنص على [4, 10]:

> إذا كانت نقاط أو عينات البيانات الفردية (كل عينة هي متغير عشوائي)، x_i ، تأتي من أي توزيع احتمال عشوائي، فإن مجموع (وبالتالي، متوسط) نقاط البيانات تلك يكون مُوزَّعًا بشكل طبيعي عندما يصبح حجم العينة، n، كبيراً.

وهكذا، حتى إذا كانت كل عينة، مثل رمية النرد (الزهر)، تأتي من توزيع غير طبيعي (nonnormal) (على سبيل المثال، التوزيع المنتظم، مثل رمية النرد)، فإن مجموع تلك العينات الفردية (مثل المجموع الذي نستخدمه لتقدير متوسط العينة، $\overline{x}$) سيكون له توزيع طبيعي. ويمكن للمرء بسهولة افتراض أن العديد من العمليات البيولوجية أو الفسيولوجية التي نقيسها هو مجموع عدد من العمليات العشوائية مع توزيعات احتمال مختلفة، وبذلك، فإن الافتراض بأن العينات تأتي من توزيع طبيعي هو افتراض معقول.

(٤,٣) فترة الثقة لمتوسط العينة
CONFIDENCE INTERVAL FOR THE SAMPLE MEAN

كل إحصاء عينة في حد ذاته متغير عشوائي بنوع محدد لتوزيع الاحتمال. وهكذا، عندما نستخدم العينات لتقدير الإحصائيات الحقيقية للمجتمع الإحصائي (عادة ما تكون في الواقع غير معروفة ولا يمكن الحصول عليها)، فإننا نريد أن يكون لدينا مستوى معين من الثقة بأن تقديراتنا للعينة قريبة من إحصائيات المجتمع الإحصائي الحقيقية أو تكون ممثلة للمجتمع أو العملية.

عند تقدير فترة الثقة لمتوسط العينة، نطرح السؤال التالي: "ما هو مقدار تطابق متوسط العينة (مُقدَّراً من عدد محدود من العينات) مع المتوسط الحقيقي للمجتمع؟"

لتحديد مستوى الثقة للتقديرات الإحصائية أو الاستنتاجات الإحصائية، هناك حاجة أولاً إلى افتراض توزيع أو نموذج احتمال للعينات والمجتمع الإحصائي ومن ثم حاجة إلى تقدير فترة الثقة باستخدام التوزيع المُفترض.

إن فترة الثقة الأبسط الذي يمكننا أن نبدأ به فيما يتعلق بإحصائياتنا الوصفية هي فترة الثقة لمتوسط العينة، $\overline{x}$. سؤالنا هو: ما هو مدى تطابق متوسط العينة مع المتوسط الحقيقي للمجتمع أو العملية، μ؟

قبل أن نتمكن من الإجابة على هذا، نحن بحاجة إلى افتراض نموذج احتمال أساسي لمتوسط العينة، $\overline{x}$، والمتوسط الحقيقي، μ. وكما ذُكر سابقاً، يمكن إظهار أن متوسط العينة، $\overline{x}$، لحجم العينات الكبير مُنمذجاً جيداً من خلال التوزيع الطبيعي أو نموذج الاحتمال. وهكذا، سوف نستخدم هذا النموذج عند تقدير فترة الثقة لمتوسط العينة، $\overline{x}$.

وهكذا يتم تقدير $\overline{x}$ من العينة. ثم نسأل، ما هو مدى تطابق $\overline{x}$ (متوسط العينة) مع المتوسط الحقيقي للمجتمع، μ؟

يمكن إظهار أنه إذا أخذنا عدة مجموعات لـ n عينة وقدّرنا $\bar{x}$ لكل مجتمع ،

١- المتوسط أو متوسط العينة لـ $\bar{x} = \mu$ ، و

٢- الانحراف المعياري لـ $\bar{x} = s/\sqrt{n}$.

وهكذا، عندما يصبح حجم العينة، n، كبيراً، فإن توزيع $\bar{x}$ يقترب من التوزيع الطبيعي.

عندما تكون n كبيرة، فإن $\bar{x}$ يتبع التوزيع الطبيعي، وقد يتم استخدام النقطة z لـ $\bar{x}$ لتقدير ما يلي :

$$\Pr(a \le \bar{x} \le b) = \Pr\left(\frac{a-\mu}{\sigma/\sqrt{n}} \le z \le \frac{b-\mu}{\sigma/\sqrt{n}}\right).$$

يفترض هذا التعبير أن n كبيرة وأننا نعرف قيمة σ .

ننظر الآن إلى الحالة التي قد تكون فيها n كبيرة، ولكننا لا نعرف قيمة. في مثل هذه الحالات، نستبدل σ بـ s للحصول على التعبير التالي :

$$\Pr(a \le \bar{x} \le b) = \Pr\left(\frac{a-\mu}{s/\sqrt{n}} \le z \le \frac{b-\mu}{s/\sqrt{n}}\right),$$

حيث يُسمى $s/\sqrt{n}$ الخطأ المعياري للعينة ويمثل الانحراف المعياري لـ $\bar{x}$.

لنفترض الآن n كبيرة، ونريد تقدير فترة الثقة 95% لـ $\bar{x}$. نقوم أولاً بتدريج متوسط العينة، $\bar{x}$ ، بالنسبة لقيمة z (لأن نظرية النزعة المركزية تنص على أن $\bar{x}$ مُوزَّعة بشكل طبيعي)

$$z = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

نحن نتذكر أن 95% من قيم z تقع بين ± 1.96 (حوالي 2σ) من المتوسط، وبالنسبة للتوزيع z،

$$\Pr(-1.96 \le z \le 1.96) = 0.95$$

وبتعويض z المساوية

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

نحصل على

$$0.95 = \Pr\left(-1.96 \le \frac{\bar{x} - \mu}{s/\sqrt{n}} \le 1.96\right)$$

إذا استخدمنا الرموز التالية من حيث الخطأ المعياري للعينة :

$$SE(\bar{x}) = \frac{s}{\sqrt{n}}$$

بإعادة ترتيب حدود الصيغة سابقاً، نلاحظ أن احتمال وقوع μ بين ± 1.96 (أو 2) انحراف معياري لـ $\bar{x}$ هو 95%:

$$0.95 = \Pr\left(\bar{x} - 1.96SE(\bar{x}) \le \mu \le \bar{x} + 1.96SE(\bar{x})\right)$$

لاحظ أنه يُشار إلى 1.96 بـ $z_{\alpha/2}$. هذه القيمة لـ z هي قيمة z التي من أجلها تكون المساحة في الذيل الأيمن للتوزيع الطبيعي مساوية $\alpha/2$. إذا أردنا تقدير فترة ثقة 99%، فإننا نستبدل $z_{0.01/2}$ ، التي هي 2.576، في الموضع 1.96 أعلاه.

وهكذا، بالنسبة لـ n كبيرة وأي مستوى وثوقية، $1-\alpha$، فإن فترة الثقة $1-\alpha$ للمتوسط الحقيقي للمجتمع الإحصائي، μ ، يُعطى بواسطة المعادلة التالية :

$$\mu = \bar{x} \pm z_{\alpha/2}SE(\bar{x})$$

وهذا يعني أن هناك $(1-\alpha)$ في المئة احتمال أن المتوسط الحقيقي يقع ضمن المجال أعلاه المتمركز حول $\bar{x}$.

مثال (٤,٣)

تقدير فترات الثقة

المسألة : مُعطى مجتمع بيانات لها $\bar{x}=505$ و $s=100$. إذا كان عدد العينات 1000، ما هي فترة الثقة 95% بالنسبة لمتوسط المجتمع، μ ؟

الحل

إذا افترضنا حجم عينة كبير، يمكننا استخدام توزيع z لتقدير فترة الثقة لمتوسط العينة باستخدام المعادلة التالية :

$$\mu = \bar{x} \pm z_{\alpha/2} SE(\bar{x})$$

وبالتعويض بالقيم التالية :

$$\bar{x} = 505;$$
$$SE(\bar{x}) = s/\sqrt{n} = 100/\sqrt{1000}$$

وبالنسبة فترة ثقة مقدارها 95%، فإن $\alpha = 0.05$.

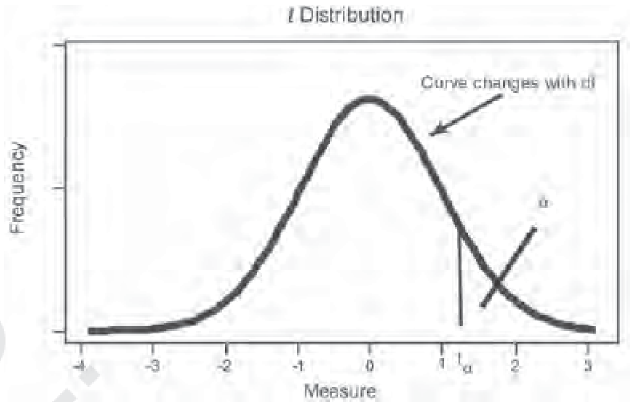
باستخدام الجدول z لتحديد موقع $z(\alpha/2)$ ، نجد أن قيمة z التي تعطي مساحة قيمتها 0.025 في الذيل الأيمن هي 1.96

وبتعويض $\bar{x}$ ، و$SE(\bar{x})$ ، و$z(\alpha/2)$ في تقدير فترة الثقة أعلاه، نجد أن مدى الثقة 95% عندما $\mu = [498.80, 511.20]$.

لاحظ أنه إذا أردنا تقدير فترة ثقة مقدارها 99%، فإننا ببساطة نستخدم قيمة مختلفة لـ z، $z(0.01/2)$ في المعادلة نفسها. إن قيمة z المرتبطة بمساحة مقدارها 0.005 في الذيل الأيمن تساوي 2.576. إذا استخدمنا هذه القيمة لـ z، فإننا نقدر فترة الثقة لـ μ بـ [496.86, 515.14]. ونلاحظ أن فترة الثقة قد اتسعت عند زيادة مستوى الثقة.

(٤,٤) التوزيع t
THE t DISTRIBUTION

بالنسبة للعينات الصغيرة، لم يعد $\bar{x}$ مُوزَّعاً بشكل طبيعي. ولذلك، فإننا نستخدم التوزيع t للطالب لتقدير الإحصاءات الحقيقية للمجتمع. يشبه التوزيع t، كما هو موضح في الجدول (٤,٢) التوزيع z ولكن مع تناقص أبطأ عند الذيول ومنطقة وسطى أكثر تسطحاً.



مدخل الجدول = $t(\alpha;df)$، حيث α هي المساحة في الذيل إلى يمين $t(\alpha,df)$ و df هي درجات الحرية.

الجدول (٢ ، ٤). دالة النقط المئوية للتوزيع t للطالب.

α = المساحة إلى يمين $t(\alpha,df)$					df
0.005	**0.01**	**0.025**	**0.05**	**0.10**	
63.657	31.821	12.706	6.314	3.078	1
9.925	6.965	4.303	2.920	1.886	2
5.841	4.541	3.182	2.353	1.638	3
4.604	3.747	2.776	2.132	1.533	4
4.032	3.365	2.571	2.015	1.476	5
					...
3.169	2.764	2.228	1.812	1.372	10
3.106	2.718	2.201	1.796	1.363	11
3.055	2.681	2.179	1.782	1.356	12
3.012	2.650	2.160	1.771	1.350	13
2.977	2.624	2.145	1.761	1.345	14
2.947	2.602	2.131	1.753	1.341	15
					...
2.750	2.457	2.042	1.697	1.310	30
2.704	2.423	2.021	1.684	1.303	40
2.660	2.390	2.000	1.671	1.296	60
2.617	2.358	1.960	1.658	1.289	120
2.576	2.326	1.960	1.645	1.282	∞
2.576	2.326	1.960	1.645	1.282	z_α (عينة كبيرة)

نستخدم التوزيع t، مع ذيول متناقصة ببطء، لأنه مع العينات القليلة، لدينا يقين أقل حول التوزيع الأساسي (نموذج الاحتمال).

تُعطى الآن القيمة التي تم معايرتها من أجل $\bar{x}$ بواسطة العلاقة التالية:

$$\frac{\bar{x}-\mu}{s/\sqrt{n}}$$

المعروفة أن لها توزيعاً t بدلاً من التوزيع z الذي ناقشناه حتى الآن. وكان أول من اكتشف التوزيع t W.S. Gosset [4, 11]، وهو كيميائي عمل في مصنع للجعة. قرر Gosset نشر التوزيع t الخاص به تحت الاسم المستعار "الطالب". لذلك، فإننا غالباً ما نشير إلى هذا التوزيع بالتوزيع t للطالب.

يعد التوزيع t متماثلاً مثل التوزيع z وله بشكل عام شكل الجرس. ولكن كمية انتشار التوزيع إلى الذيول، أو عرض الجرس، تعتمد على حجم العينة، n. وخلافاً للتوزيع z، الذي يفترض حجم عينة لا نهائية، فإن التوزيع t يغير الشكل مع حجم العينة. والنتيجة هي أن فترات الثقة التي يتم تقديرها بالقيم t هي أكثر "انتشاراً" منها للتوزيع z، وخاصة بالنسبة لأحجام عينة صغيرة، لأنه مع مثل هذه الأحجام للعينات، يتم لومنا لعدم وجود عينات كافية لتمثيل مدى التغير في المجتمع الإحصائي أو العملية. وهكذا، عند تقدير فترات الثقة لمتوسط العينة، $\bar{x}$، فإنه لا يوجد لدينا ثقة كبيرة في تقديراتنا. وبذلك، فإن الفترة تتسع لتعكس هذا اليقين المنخفض بأحجام العينة الأصغر. نقدر في المقطع التالي فترات الثقة لنفس المثال المُعطى مسبقاً، ولكن باستخدام التوزيع t بدلاً من التوزيع z.

(٤,٥) فترات الثقة باستخدام التوزيع t

CONFIDANCE INTERVAL USING t DISTRIBUTION

هناك جداول t، مثل الجداول z، تكون فيها القيم t المرتبطة بمساحات مختلفة تحت منحنى الاحتمال محسوبة مسبقاً، ويمكن استخدامها للتحليل الإحصائي دون الحاجة إلى

إعادة حساب القيم t. إن الفرق بين الجدول z والجدول t هو أن القيم t الآن دالة في الحجم العينات أو درجات الحرية. يعطي الجدول (٤,٢) بضعة أسطر من الجدول t من [3].

لاستخدام الجدول ينظر المرء ببساطة إلى نقطة تقاطع درجات الحرية، df، (تتعلق بحجم العينة)، وقيمة α التي يرغب فيها المرء في الذيل الأيمن. توفر نقطة التقاطع القيمة t التي من أجلها تكون المساحة تحت المنحنى t في الذيل الأيمن مساوية α. وبعبارة أخرى، فإن احتمال أن يكون t أقل من إدخال محدد أو يساويه في الجدول هو $1-\alpha$. وبالنسبة لحجم عينة محدد، n، فإن درجات الحرية هي $df = n-1$.

نستبدل الآن ببساطة t بـ z لإيجاد فترات الثقة. لذا، فإن فترات الثقة لمتوسط العينة، $\bar{x}$، باستخدام التوزيع t يصبح الآن:

$$\mu = \bar{x} \pm t\left(\frac{\alpha}{2}; n-1\right) SE(\bar{x})$$

حيث μ هو المتوسط الحقيقي للمجتمع أو العملية التي نأخذ منها العينات، و $SE(\bar{x})$ هو الخطأ المعياري للمجتمع أو العملية، و t هي قيمة t التي يوجد من أجلها مساحة قيمتها $\alpha/2$ في الذيل الأيمن، و n هو حجم العينة.

المثال (٤,٤)

فترة الثقة باستخدام التوزيع t.

المسألة

نأخذ في الاعتبار المثال نفسه المُستخدم سابقاً لتقدير فترات الثقة باستخدام القيم z. في هذه الحالة، يكون حجم العينة صغيرًا $n=20$، لذلك نستخدم الآن التوزيع t.

الحل

تقديرنا الآن فترة الثقة بالنسبة لـ $\mu = \bar{x} \pm t\left(\frac{\alpha}{2}; n-1\right) SE(\bar{x})$

مرة أخرى، نستبدل بالقيم التالية:

$$\bar{x} = 505;$$
$$SE(\bar{x}) = s/\sqrt{n} = 100/\sqrt{20}$$

وبالنسبة لفترة ثقة مقدارها 95%، فإن $\alpha = 0.05$.

باستخدام الجدول t لتحديد موقع $t(0.05/2, 20-1)$، نجد أنه من أجل 19 درجة حرية ($19\ df$)، أن قيمة t التي تعطي مساحة قيمتها 0.025 في الذيل الأيمن هي 2.093.

وبتعويض $\bar{x}$، و $SE(\bar{x})$، و $t(\alpha, n-1)$ في تقدير فترة الثقة سابقاً، نجد أن فترة الثقة 95% عندما $\mu = [458.20, 551.80]$. ونلاحظ أن فترة الثقة هذه مُتسعة مقارنة مع الفترة الذي تم تقديرها سابقاً باستخدام القيم z. وهذا مُتوقع لأن التوزيع t أوسع من التوزيع z في أحجام العينة الصغيرة، مما يعكس حقيقة أن لدينا ثقة أقل في تقديرنا لـ $\bar{x}$ وبالتالي، μ عندما يكون حجم العينة صغيرًا.

ويمكن تقدير فترات الثقة لمعظم الإحصائيات الوصفية، مثل متوسط العينة، تباين العينة، وحتى الخط لأفضل ملاءمة مُحددة من خلال خط الانحدار [3]. وكما أشير سابقاً، يعكس فترة الثقة التغيّر في تقدير البارامتر وهو متأثر بحجم العينة، وتغيّر المجتمع، ومستوى الثقة الذي ننشده. كلما كانت الثقة المطلوبة أكبر، كانت الفترة أوسع. وبالمثل كلما كان التغير في العينات أكبر، كانت فترة الثقة أوسع.

المثال (٤,٥)

مراجعة مفاهيم الاحتمال.

المسألة

تم قياس أزمنة التجميع لعينة مؤلفة من 15 مضخة حقن غلوكوز. كان متوسط الزمن لتجميع مضخة حقن غلوكوز 15.8 دقيقة، مع انحراف معياري مقداره 2.4 دقيقة. بافتراض توزيع متماثل نسبي لأزمنة التجميع.

١ – ما النسبة المئوية لمضخات الحقن التي تتطلب أكثر من 17 ثانية للتجميع؟

٢– ما فترة الثقة 99% للمتوسط الحقيقي لزمن التجميع (μ) ؟

٣– ما فترة الثقة 99% لمتوسط زمن التجميع إذا كان حجم العينة 2500؟

الحل

١ – x = زمن التجميع.

ما هو $\Pr(x>17)$ ؟

$\Pr(x>17)=\Pr(z>(17-15.8)/2.4)=1-\Pr(z\le 0.5)=1-0.6915=0.3085$ ،

أو 38.50% من مضخات الحقن

٢– $=\bar{x}\pm t(\alpha/2;n-1)SE(x)$

$$\begin{aligned}&=15.8\pm t((0.01)/2;15-1)(2.4\sqrt{15})\\&=15.8\pm 2.977(0.6196)\\&=[13.96,17.64]\end{aligned}$$

لأن حجم العينة 2500 كبير الآن، فإننا نستخدم قيمة z لتقدير فترة الثقة، $\mu=\bar{x}\pm z(\alpha/2)SE(x)$.

٣– $=\bar{x}\pm t(\alpha/2)SE(x)$

$$=15.8\pm t((0.01)/2)(2.4\sqrt{2500})$$

$$=15.8\pm 2.576(0.048)$$

$$=[15.68,15.92]$$