

سنعتبر البيانات التوضيحية والتي استخدمناها من قبل (1, 2, 3, 7, 12) والتي متوسطها هو 5. سنفترض مؤقتاً أن هذه البيانات هي لمجتمع وليس لعينة، (أهمية هذه التفرقة سوف تتضح فيما بعد). أخذاً في الاعتبار أنها مجتمع، فإن متوسط مربعات الانحرافات بعداً عن المتوسط يسمى تباين Variance المجتمع.

التباين لتلك البيانات ينفذ كما يلي:

$$\text{Variance} = \frac{(1-5)^2 + (2-5)^2 + (3-5)^2 + (7-5)^2 + (12-5)^2}{5} = 16.4$$

للتباين، كمقياس للاختلاف، ميزة على متوسط الانحرافات المطلقة، فهو يساعد في تسهيل العمليات الحسابية، وكنتيجه لذلك، فإنه يعد أهم مقاييس الاختلاف في عملية تطوير النظرية الإحصائية.

يرمز لتباين المجتمع بـ σ^2 (سجماً تربيعاً) وتحدد بصفة عامة بالصيغة التالية:

$$\sigma^2 = \frac{\sum_{i=1}^N (\chi_i - \mu)^2}{N} \quad (2.4)$$

والأهمية الرئيسية لتباين المجتمع أنه المقياس الأساسي للاختلاف بين مفردات المجتمع، لذلك، فإنه غالباً المؤشر ذو الأهمية الكبرى. ومع ذلك، فإنه في معظم التطبيقات الإحصائية نعتمد على بيانات العينة. وكنتيجه لذلك، فإنه نادراً ما نجد أنفسنا نحسب تباين المجتمع، وبدلاً لذلك فإننا نهتم عند استخدام بيانات العينة بتقدير estimate لتباين المجتمع.

نفرض أننا نتعامل مع بيانات عينة، هل يمكنك أن تتوقع ما هو شكل صيغة التباين للعينة؟ لتأخذ الآن لحظة لتحاول في ذلك. (ملحوظة: تفحص الصيغة (2.4) وضع التعديلات في الرموز لتعكس تباين العينة بدلاً من المجتمع). عندما تشتق نظرياً صيغة تباين العينة، ينتج نظير دقيق لتباين المجتمع وذلك باحلال متوسط العينة $\bar{x}$ محل μ ، وحجم العينة n محل حجم المجتمع N في الصيغة (2.4). ومع ذلك فقد تبين أن القسمة على n تسبب ضعفاً في التقدير الناتج لـ σ^2 ، بمعنى أنه للعديد من العينات العشوائية، يكون هذا التقدير في الغالب الأعم أقل من σ^2 عن كونه أكبر من σ^2 . هذه الحالة تسمى بالتحيز bias (الفصل الخامس يعطي مناقشة أكثر استفاضة عن التحيز). في العادة التحيز يعد أمراً غير مرغوب فيه، لأن الغرض الأساسي من قياس تباين العينة هو تقدير تباين المجتمع، ومع ذلك فقد تبين أن القسمة على $(n-1)$ بدلاً من n يزيل هذا التحيز. لذلك فإن تباين العينة مع القسمة على $(n-1)$ يقال عنه أنه غير متحيز Unbiased وهذه الصيغة هي التي نستخدمها خلال هذا الكتاب. وتباين العينة يرمز له بـ S^2 ويحدد بالصيغة التالية:

$$S^2 = \frac{\sum_{i=1}^n (\chi_i - \bar{X})^2}{n-1} \quad (2.5)$$

درجات الحرية Degrees of Freedom

هناك حد إحصائي في مقام تباين العينة وهو $(n-1)$. أنه يسمى بـ درجات الحرية degrees of freedom، لماذا هذا الاسم؟ سنوضح ذلك بمثال. نفرض أنك طلبت من "محمد" أن يختار ثلاثة أعداد وأنه اختار 2, 4, 16. لنفرض أنك أيضاً طلبت من "حازم" أن يختار ثلاثة أعداد بشرط أن يكون متوسطها يساوي 5، فكان اختياره على الصورة 5, 10. لاحظ أن محمد كان حراً تماماً في اختيار أي ثلاثة أعداد، لكن حازم كان حراً فقط في اختيار عددين، لأن العدد الثالث يجب أن يكون العدد الذي يعطي المجموع 15 حتى يكون المتوسط 5. بسبب اشتراط ثبات المتوسط، فإننا نقول أن هناك درجتان فقط من الحرية. والآن، دعنا نعمم ذلك، إذا طلب من شخص ما اختيار n من الأعداد بحيث يكون لها متوسط محدد، فإن درجات الحرية يصبح $(n-1)$ ، وأن آخر عدد يختار يجب أن يكون القيمة التي تجعل المتوسط المحدد سلفاً يمكن تحقيقه. حيث أن تباين العينة يعتمد على انحرافات قيم البيانات بعداً عن المتوسط، فإننا نقول أن له $(n-1)$ درجة حرية.

(٥-٥) توزيع المعاينة للأحصاء $\bar{X}$: ملخص: The Sampling Distribution of $\bar{X}$: A Summary

من مناقشة هذا الفصل، يمكن أن نلخص الوضعين اللذين يحددان أي توزيعات المعاينة لـ $\bar{X}$ يمكن أن تستخدم لعمل إستنتاجات إحصائية حول μ :

١. عندما يكون الإنحراف المعياري للمجتمع σ معلوماً: هذا يؤدي إلى الأحصاء Z ، ومع ذلك فإنه من النادر أن نعرف σ .

٢. عندما يكون الإنحراف المعياري للمجتمع σ غير معلوماً: هذا يؤدي إلى الأحصاء T ، وهذا هو الوضع العادي في التطبيقات الأحصائية الحقيقية. وحيث أن أغلب التطبيقات تنتمي إلى ثاني هذه الأوضاع، فإن التوزيع المناسب في أغلب التحليلات المتعلقة بمتوسط المجتمع هو توزيع T أكثر من التوزيع الطبيعي المعياري. وفيما يلي ملخصاً لهذين الوضعين:

ملخص

توزيع المعاينة الذي يستخدم في عمل إستنتاجات حول μ اعتماداً على $\bar{X}$

1- إذا كانت قيمة الإنحراف المعياري في المجتمع معلومة، وكان:

(أ) توزيع المجتمع هو التوزيع الطبيعي.

(ب) توزيع المجتمع ليس الطبيعي، ولكن حجم العينة n كبيراً بدرجة كافية ($n \geq 30$)، فإن

توزيع المعاينة للأحصاء Z ، حيث: $Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ هو تقريباً التوزيع الطبيعي المعياري.

2- إذا كانت قيمة الإنحراف المعياري في المجتمع غير معلومة، وكان:

(أ) توزيع المجتمع هو التوزيع الطبيعي.

(ب) حجم العينة كبيراً بدرجة كافية ($n \geq 30$)، فإن توزيع المعاينة للأحصاء T ، حيث:

$T = \frac{\bar{X} - \mu}{S / \sqrt{n}}$ هو تقريباً توزيع T بدرجات حرية $(n-1)$.

توزيع المعاينة للنسبة .

ملخص: توزيع المعاينة الذي يستخدم في عمل استنتاج حول اعتماداً على P

توزيع المعاينة للنسبة يقترب بدرجة كافية من التوزيع الطبيعي، وله:

المتوسط: $\mu = P$

الانحراف المعياري: $\sigma = \sqrt{\frac{P(1-P)}{n}}$

إذا كان: $n \geq 30$; $nP \geq 5$ and $n(1-P) \geq 5$

لذلك، إذا ما تم معايرة P ، فإن توزيع المعاينة للأحصاء Z ، حيث $Z = \frac{\bar{X} - P}{\sqrt{P(1-P)/n}}$

يكون له تقريباً التوزيع الطبيعي المعياري.

(٣-٥-٥) توزيع المعاينة لـ $\bar{X}$ عندما يكون المجتمع له توزيع غير طبيعي:

The Sampling Distribution of $\bar{X}$ When the Population Has a Nonnormal Distribution

في كثير من الحالات، لا نستطيع تعين هوية توزيع المجتمع وبالتالي لا يمكن تحديد توزيع المعاينة لـ $\bar{X}$ ، ومع ذلك فقد تمكن علماء الإحصاء من إثبات أن توزيع المعاينة لـ $\bar{X}$ هو تقريبا التوزيع الطبيعي في حالة العينات ذات الأحجام الكبيرة أيا كان توزيع المجتمع. هذه النتيجة الحاسمة في الاستنتاج الإحصائي تعرف باسم "نظرية النهاية المركزية" Central Limit Theorem ويمكن تلخيصها على النحو التالي:

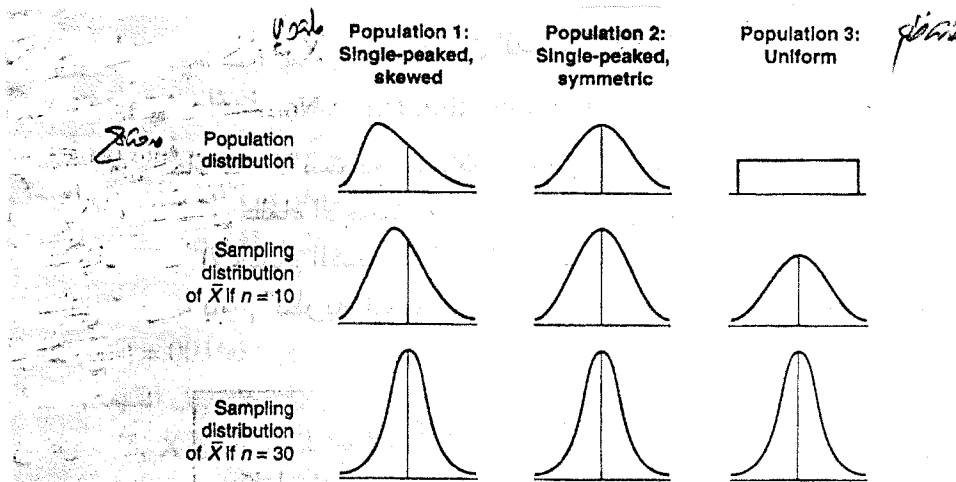
نظرية النهاية المركزية
كلما زاد حجم العينة، كلما اقترب توزيع المعاينة لـ $\bar{X}$ من التوزيع الطبيعي بغض النظر عن توزيع المجتمع.

الحقيقة الجديدة بالملاحظة حول توزيع المعاينة لـ $\bar{X}$ أنه يميل تجاه التوزيع الطبيعي للعينات كبيرة الحجم بغض النظر عن طبيعة توزيع المجتمع، وهكذا فتوزيع المعاينة لـ $\bar{X}$ في جميع العينات الممكنة كبيرة الحجم هو تقريبا توزيع طبيعي بمتوسط μ وخطأ معياري $\sigma / \sqrt{n}$ حيث μ, σ هما متوسط المجتمع وانحرافه المعياري على التوالي، لذلك نجد أن توزيع Z ، حيث:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

هو تقريبا توزيع طبيعي معياري طالما أن n كبيرة بدرجة كافية.

الآن إلى أي درجة يجب أن تكون n كبيرة حتى يكون لـ $\bar{X}$ توزيع طبيعي؟ الأجابة تعتمد على درجة اقتراب توزيع المجتمع من التوزيع الطبيعي. شكل (٧-٥) يوضح بعض الأمثلة لتوزيعات المعاينة لـ $\bar{X}$ عند أحجام عينات مختلفة عندما تتم المعاينة من مجتمعات ذات توزيعات مختلفة. يلاحظ أن الأقتراب تجاه الاعتدالية (الطبيعي) هو أسرع في المجتمع (2) عن المجتمع (1). يحدث هذا لأن توزيع المجتمع (1) ملتوي إلى حد بعيد بينما توزيع المجتمع (2) تقريبا طبيعي (معتدل). يلاحظ أيضا أن توزيع المعاينة لـ $\bar{X}$ بالنسبة للمجتمع (3) يميل إلى الاعتدالية بسرعة حتى ولو كان توزيع المجتمع ليس له الشكل الاعتدالي (الطبيعي). وعلى ذلك، أيا كان توزيع المجتمع، فإن توزيع المعاينة لـ $\bar{X}$ هو تقريبا توزيع طبيعي عند أحجام العينات $n \geq 30$.



شكل (٧-٥): تأثير حجم العينة على شكل توزيع متوسط العينة عندما تتم المعاينة من مجتمعات ذات توزيعات مختلفة

(5-5) توزيع المعاينة لـ $\bar{X}$ عندما يكون الانحراف المعياري للمجتمع σ غير معلوم: مقدمة لتوزيع T:
The Sampling Distribution of $\bar{X}$ When The Population Standard Deviation σ is Unknown: An Introduction to the T Distribution

عند مناقشة توزيع المعاينة لـ $\bar{X}$ في البندين الأخيرين، إفتراضنا أن الانحراف المعياري للمجتمع σ معلوماً. والآن نتذكر من الصيغة (5.3) أنه إذا كان متوسط العينة $\bar{X}$ له توزيع طبيعي، فإن:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

يكون له توزيع طبيعي معياري. هذه النتيجة تفترض مسبقاً أن σ هي ثابت معلوم. ولكن إذا كانت σ غير معلومة، فإن Z تكون دالة في مؤشر غير معلوم ومن ثم لا يمكن تحديد قيمة Z لعينة محددة. ويبدو أن هذا يخلق مشكلة، حيث أنه من الناحية العملية، نادراً ما تكون قيمة الانحراف المعياري في المجتمع معلومة. من المتوقع أن يكون ردك الطبيعي على هذه النقطة أن تقول: لماذا لا يتم إستبدال σ بمقدرها أي بالانحراف المعياري في العينة S ؟ حسناً، هذا هو بالضبط ما فعلناه. إستبدال σ بالتقدير S في الصيغة (5.3) يؤدي إلى الكمية T (وتسمى بالأحصاء T)، حيث:

$$T = \frac{\bar{X} - \mu}{S / \sqrt{n}} \quad (5.4)$$

ومما يؤسف له أن توزيع المعاينة للأحصاء T ليس توزيع طبيعي معياري، حتى ولو كان توزيع المجتمع هو التوزيع الطبيعي، ولكي نفهم السبب في ذلك قارن الكميتين:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \quad \text{and} \quad T = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

الآن كم عدد الأحصاءات (متغيرات عشوائية) التي تراها في هذه الكميات؟ الأجابة: بالنسبة إلى Z ، هناك متغير عشوائي واحد يسمى $\bar{X}$ (تذكر أن μ ، σ كلها ثوابت) أما الأحصاء T فيعتمد على متغيرين عشوائيين: $\bar{X}$ ، S . إدخال إحصاء إضافي S يزيد من إختلاف قيمة T من عينة إلى أخرى مقارنة مع Z ، لذا يجب ألا نتوقع أن تكون توزيعات Z ، T هما نفس الشيء. في الحقيقة فإن توزيع الأحصاء T لجميع العينات العشوائية كل ذات الحجم n والمسحوبة من مجتمع توزيعه طبيعي يسمى توزيعات الطالب T : Student's T distribution والذي يختصر غالباً إلى توزيع T (*).

(*) قدم توزيع T في عام ١٩٠٨ بواسطة W.S.Gosset والذي نشر أبحاثه تحت اسم مستعار «طالب». كثير من المؤلفين يستخدم الحرف الصغير t للإشارة إلى هذا التوزيع. هنا نستخدم الحرف الكبير T لنحافظ على الاتساق العملي في استخدام الحروف الكبيرة لتدل على المتغيرات العشوائية والحروف الصغيرة للإشارة إلى قيم المتغيرات العشوائية.

(٧-٧-٥) توزيع المعاينة لـ S^2 : مقدمة لتوزيع كاي تربيع:

The Sampling Distribution of S^2 : An Introduction to the Chi-Square Distribution

نفرض أنه سحبت عينة عشوائية حجمها n من مجتمع يتبع التوزيع الطبيعي. ولعمل استنتاجات حول تباين المجتمع σ^2 اعتماداً على S^2 فإننا نحتاج إلى معرفة توزيع المعاينة لـ S^2 ، ومن ثم اشتقاقه رياضياً، لكن ذلك قد يسبب إرباكاً وعدم متابعة في فهم الموضوع. عموماً كثيراً ما نستخدم الأحصاء: $(n-1) S^2 / \sigma^2$ وهو -كما نرى- دالة في S^2 . يطلق على هذا الأحصاء -إحصاء كاي تربيع Chi-Square Statistic-. وقد أثبت علماء الإحصاء أنه إذا كان توزيع المجتمع هو التوزيع الطبيعي، فإن توزيع المعاينة لأحصاء كاي تربيع:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \quad (5.16)$$

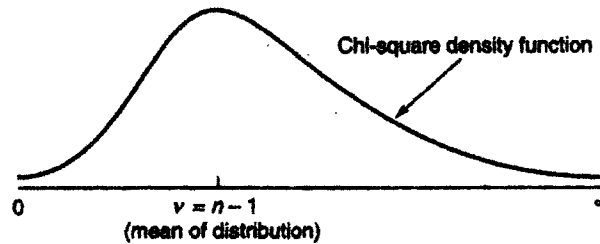
هو توزيع كاي تربيع بدرجات حرية $(n-1)$. الرمز χ هو حرف يوناني كاي. درجات الحرية $(n-1)$ المقترنة بالأحصاء كاي تربيع تعكس حقيقة أن هناك $(n-1)$ من درجات الحرية مقترنة بتباين العينة S^2 .

حتى هذه النقطة، ربما تسأل نفسك: لماذا هذا الأحصاء يسمى كاي تربيع؟ حسناً، أنه توزيع آخر من التوزيعات الأحصائية وله استخدامات كثيرة في الأحصاء التطبيقي (شأنه في ذلك شأن توزيع T) ولكي نتعرف بصورة أوضح على توزيع كاي تربيع، دعنا ننظر بأمعان للصيغة (5.16):

1- في البداية، تأمل فيما إذا كانت هذه الكمية يمكن أن تأخذ قيماً أقل من الصفر. الأجابة بالطبع هي: لا، لأن كل S^2 ، σ^2 هي كميات مربعة وحجم العينة n هو عدد صحيح موجب ومن ثم فإن مدى المتغير العشوائي الذي يتبع كاي تربيع يبدأ من الصفر إلى (نظرياً) ما لا نهاية.

2- متوسط توزيع كاي تربيع هو ببساطة $(n-1)$ أي عدد درجات الحرية وهذا لا يدهشنا كثيراً، لأن القيمة المتوقعة لـ S^2 هو σ^2 وبالتالي فالقيمة المتوقعة (المتوسط) للأحصاء كاي تربيع $(n-1) S^2 / \sigma^2$ يساوي $(n-1)$.

3- ماذا بشأن شكل توزيع كاي تربيع؟ الشكل بكل تأكيد غير متماثل، لأن مداه محدود من اليسار بالصفر وغير محدود من اليمين، وبالتالي فشكله البياني ملتوي إلى اليمين، كما هو موضح في شكل (٥-١٥) ويكون الألتواء كبيراً عندما تكون درجات الحرية صغيرة ولكن يتضاءل الألتواء كلما ازداد عدد درجات الحرية.



شكل (٥-١٥): توزيع كاي تربيع

توزيع المعاينة لـ S^2 :

ملخص: توزيع المعاينة الذي يستخدم لعمل استنتاجات عن σ^2 اعتماداً على S^2

إذا كان توزيع المجتمع هو التوزيع الطبيعي، فإن توزيع المعاينة للإحصاء كاي تربيع، حيث:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

هو توزيع كاي تربيع بدرجات حرية $(n-1)$

جدول كاي تربيع واستخدامه:

يتواجد توزيع كاي تربيع في معظم البرامج الإحصائية الجاهزة، بالإضافة إلى جدولته بصورة موسعة، وجدول D في الملحق يعطي قيما جزئية مقترنه بأحتمالات محددة تمثل مساحات تقع على يسار تلك القيم الجزئية، وتنظيم جدول D يماثل تنظيم جدول توزيع T، بمعنى أننا في البداية نحدد عدد درجات الحرية، وهذا العدد مبين في أول عمود من الجدول تحت العنوان γ . بعد ذلك نختار الاحتمال المطلوب من بين الاحتمالات المبينه في عناوين الأعمدة. وعند الصف الذي يشير إلى درجات الحرية المناسبة والعمود ذو الاحتمال المطلوب، يعطي الجدول القيمة المناظرة. فمثلا، بفرض أن $n=20$ ونرغب في إيجاد القيم الجزئية التي تقابل الاحتمالات 0.01 و 0.95 على التوالي. بالبحث في جدول D وعند درجات الحرية $\gamma = 19$ وتحت الأعمدة 0.01 و 0.95، نجد أن القيم الجزئية هي: 7.63، 30.15 على التوالي، هذه البيانات تعنى أن احتمال أن المتغير العشوائي كاي تربيع بدرجات حرية 19، احتمال أن يأخذ قيما لا تزيد عن 7.63، 30.15 هي 0.01 و 0.95. على التوالي. بالرموز يمكن أن نكتب:

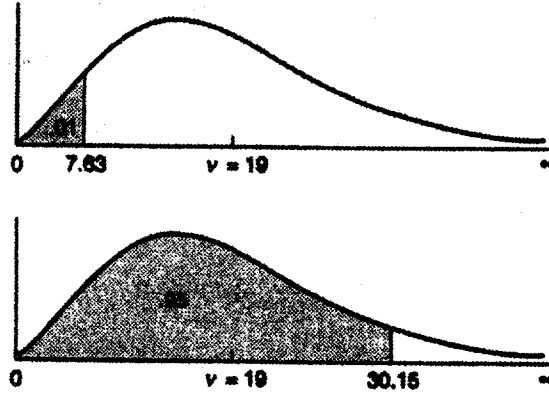
$$P(\chi_{19}^2 \leq 7.63) = 0.01$$

$$P(\chi_{19}^2 \leq 30.15) = 0.95$$

هذه القيم الجزئية واحتمالاتها المقترنه بها موضحة في شكل (١٦-٥) وينتج من قاعدة الاحتمال للحوادث المكمل أن :

$$P(\chi_{19}^2 > 7.63) = 0.99$$

$$P(\chi_{19}^2 > 30.15) = 0.05$$



شكل (١٦-٥): توضيح قيم كاي عند درجة الحرية 19