تعاريف ومصطلحات أساسية

- وحدة المعاينه : Sampling Unit
- المجتمع الاحصائي : .Statistical pop
- العينه والمعاينه : Sample & sampling
- حجم المجتمع وحجم العينه
- كسر المعاينه
- المتغير العشوائي : Random variable
- الوسط الحسابي : Arithmetic mean
- التباين والانحراف المعياري : Variance & Standard Deviation
- التغاير والارتباط Covariance & Correlation
- معلمة المجتمع : pop. parameter
- احصائية العينه : Asample Statistic
- الاحتمال : Probability
- التوقع : Expectation
- أهم التوزيعات الاحتمالية
- توزيع ذي الحدين : Binomial distribution
- التوزيع الطبيعي : Normal Distribution
- توزيع ستيودنت : Student Dist
- تقدير معالم المجتمع : Estimation of pop. Parameters
- التقدير والمقدر : Estimate & Estimator
- المقدر الجيد : Best Estimator
- عدم التحيز
- الاتساق
- الكفاءة
- الكفاية
- التقدير بنقطة والتقدير بفترة : Point & Interval estimate

وحدة المعاينه : Sampling Unit

وحدة المعاينه هي " الجزء أو الكيان الصغير الذي نجمع منه البيانات " . ان كل وحدة من الوحدات المكونه للمجتمع هي وحدة معاينه أي أن عدد وحدات المعاينه هي عدد وحدات المجتمع . ان وحدات المعاينه قد تكون وحدات طبيعية تتعلق بالجنس البشري وكالموظف والطالب ، والفرد والأسرة) أو وحدات مصطنعة (كالمؤسسة ، أو الوزارة أو المسكن أو المصنع) . كما أن وحدات المعاينه قد تكون متشابه من حيث الحجم أو مختلفة . وعند تنفيذ البحوث الميدانية ، يجب تحديد وتعريف وحدة المعاينه تعريفاً واضحاً لجمع البيانات من الوحدات التي يشملها البحث وعدم تداخل هذه الوحدات مع تلك التي لايشملها البحث . كذلك يجب التمييز بين وحدات المعاينه ووحدات المشاهدة (وحدة المشاهدة هي الوحدة التي يجري عليها القياس أو التصنيف) اللتين قد تتطابقا أو لا تتطابقان (مثلاً قد تكون وحدة المعاينه المصنع ووحدة القياس المدير أو العامل) .

المجتمع الاحصائي : Statistical pop.

المجتمع الاحصائي هو عبارة عن " جميع وحدات المعاينه التي نقوم بدراستها " أي هو جميع وحدات المعاينه التي نريد الاستدلال على خواصه عن طريق العينه . ويمكننا تقسيم المجتمعات إلى مجتمعات ثابته لا تخضع لتغيرات خلال فترة (قصيرة) من الزمن كالمدن والشوارع ومجتمعات غير ثابتة (حركية) تتغير بشكل سريع من فترة لأخرى مثل عدد السكان وعدد السيارات التي تمر في شارع ما ويجب تحديد المجتمع الذي سيشمله البحث تحديداً واضحاً ودقيقاً لتقييم نتائج العينه بشكل دقيق ، خاصة فيما يتعلق بعدد وحدات المجتمع حيث يمكننا التمييز بين المجتمع المحدد Finite pop. عندما يكون عدد القيم محدوداً والمجتمع غير المحدد Infinite pop. عندما يتضمن المجتمع عدداً لا نهائياً من القيم .

العينه والمعاينه : Sample & sampling

نستخدم كلمة العينه كثيراً في حياتنا اليومية ، إذ عندما يمرض الشخص يطلب الطبيب فحص عينه من دمه أي بجزء منه . كذلك عندما نريد شراء سلعه معينه كالحبوب (القمح، الأرز ...) نختار جزءاً من هذه السلعة للتأكد من جودتها ، ولاتخاذ قرار بشرائها أو عدم شرائها . إن عملية الاختيار قد تكون جيدة ومناسبة بحيث تمكننا من الوصول إلى القرار السليم ، وقد تكون خاطئة تعطي نتائج مضلله.

وتعرف العينه بأنها " جزء من المجتمع يتم اختياره لتمثيل المجتمع " أما المعاينه فتعرف بأنها " عملية اختيار جزء من المجتمع الاحصائي للاستدلال على خواص المجتمع بأكمله عن طريق تعميم نتائج العينه " .

ولتوضيح هذين المفهومين ، نورد المثال التالي :

نفرض أننا نريد دراسة مستوى الرضا الوظيفي لموظفي احدى الجهات ، ونظراً لضخامة عدد موظفي هذه الجهة فقد تقرر اختيار عدد من الموظفين يمثلون المجتمع . ان الموظفين الذين تم اختيارهم هم العينة ، اذ يشكلون جزءاً من المجتمع يتضمن خصائصه . أما عملية اختيار هذه العينه وتعميم النتائج للاستدلال على خصائص المجتمع فتسمى " معاينه " .

حجم المجتمع وحجم العينه

يقصد بحجم المجتمع عدد جميع وحدات المعاينه التي يتكون منها المجتمع ويرمز له عادة بالرمز N أما حجم العينه فهو عدد وحدات المعاينه التي تـم اختيارها ويرمز له عادة بالرمز n . ويعتبر حجم العينه صغيراً اذا كان أقل من 30 أي اذا كانت $n < 30$.

كسر المعاينه

يمثل كسر المعاينه الواحدات المختارة في العينه إلى عدد وحدات المعاينه في المجتمع ، أي يساوي نسبة حجم العينه إلى حجم المجتمع ويرمز له عادة بالرمز f حيث $f=\frac{n}{N}$ وعندما يكون لدينا عينات جزئية (يشكل مجموعها العينه) أي $n=n_1+n_2+\ldots+n_L$ حيث L عدد الأقسام (كما هو الحال في المعاينه الطبقية التي سندرسها فيما بعد) نجد أن كسر المعاينه للطبقة i يساوي $f_i=\frac{n_i}{N_i}$ حيث N_i حجم المجتمع في الطبقة i و n_i حجم العينه في الطبقة i ويكون لدينا عدة كسور للمعاينه (L كسراً) :

$$f_1=\frac{n_1}{N_1},f_2=\frac{n_2}{N_2},\ldots f_L=\frac{n_L}{N_L}$$

المتغير العشوائي : Random variable

عندما نقيس وزن أو طول أو عمر شخص ما ، فاننا نشير إلى النتائج التي نحصل عليها بقيم معينه يعبر عنها بمتغير . لذا يمكننا تعريف المتغير بأنه " رمز يمكن أن يأخذ أية قيمة سبق تحديدها تسمى مجال هذا المتغير " ويرمز عادة للمتغيرات برموز (x,y,z,...) مثلاً عندما يكون لدينا N قيمة تمثل أعمار الأطفال يمكننا التعبير عن هذه القيم بالمتغير X حيث $X:X_1,X_2,X_3,\ldots,X_N$ حيث يشير الدليل إلى رقم القيمة أي رقم الوحدة الاحصائية وعندما نحصل على النتائج (القيم) نتيجة العوامل العشوائية (عوامل الحظ أو الصدفة) يسمى المتغير متغير عشوائي كما تسمى النتائج التي نحصل عليها بالمشاهدات أو المفردات observations ويمكننا تعريف المتغير العشوائي بأنه دالة ذات قيم عددية حقيقية معرفة على فضاء العينة .

ونستطيع التمييز بينو نوعين من المتغيرات العشوائية :

(أ) متغير عشوائي متقطع Discrete Random variable

وهو المتغير العشوائي الذي نحصل عليه عندما يكون هناك تقطعات أو قفزات بين القيم وعند عدم وجود قيم بين كل قيمتين من القيم ويأخذ عدداً محدداً من القيم مثلاً عدد أفراد الأسرة للموظفين في احدى الجهات هو متغير عشوائي متقطع يأخذ القيم

X: 0,1,2,3,4,...,10

(ب) متغير عشوائي متصل Continuous Random Variable

المتغير العشوائي المتصل هو المتغير العشوائي الذي لايتضمن فجوات أو تقطعات كما هو الحال في المتغير المتقطع. أي أن المتغير العشوائي المتصل هو المتغير الذي يمكن أن يأخذ أية قيمة ضمن مجال القيم للمتغير الذي ندرسه . مثلاً درجات حرارة المرضى y يمكن أن تأخذ عدة قيم تتراوح بين 36 و 41 أي Y:37,37.5,39.1,38,38.2 .

ان البيانات التي يمكن التعبير عنها بمتغيرات متقطعة تسمى بيانات متقطعة والبيانات التي يمكن التعبير عنها بمتغيرات متصلة تسمى بيانات متصلة وعندما يأخذ المتغير قيمة وحيدة فقط يسمى: "ثابت".

وكذلك يمكننا التمييز بين المتغيرات الكمية التي يمكن قياسها كالأطوال والأوزان وغيرها والمتغيرات النوعية أو الاسمية التي تعبر عن الظواهر التي لايمكن قياسها كالجنس أو اللون مثلاً نعبر عن متغير الجنس للمرضى

X=1,2,1,1,2

حيث يشير العدد 1 إلى المريض الذكر والعدد 2 يشير اذا كان أنثى والجنس هو متغير اسمى .

الوسط الحسابي : Arithmetic mean

يعد الوسط الحسابي أحد وأهم مقاييس النزعة المركزية ، ويعرف الوسط الحسابي بأنه القيمة التي نحصل عليها اذا قسمنا مجموع القيم على عددها . اذا رمزنا لقيم المجتمع بالمتغير X حيث لدينا N قيمة أو مفردة يكون الوسط الحسابي للمجتمع ونرمز له بالرمز μ .

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} \qquad (1)$$

حيث $\sum_{i=1}^{N} X_i$ يشير إلى مجموع قيم المجتمع التي عددها N قيمة . واذا رمزنا إلى قيمة العينة في السحب i بـ x_i حيث لدينا i = 1,2,...,n فان الوسط الحسابي للعينه ونرمز له بالرمز $\bar{x}$ وتقرأ x

يساوي :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (2)$$

ويعد متوسط العينه من أفضل المقدرات لمتوسط المجتمع الذي يكون غالباً غير معلوم ، لأن قيم المجتمع غير معلومه في معظم الحالات . وكثيراً ما تستخدم كلمة المتوسط Mean للدلالة على المتوسط الحسابي .

التباين والانحراف المعياري : Variance & Standard Deviation

يعد التباين والانحراف المعياري من أهم مقاييس الانتشار والتشتت Measure of Dispersion التي تقيس مدى انتشار القيم عن بعضها أو عن قيمة معينة . ويعد التباين أحد المقاييس التي تستخدم لقياس مدى ابتعاد القيم عن الوسط الحسابي ، إذ كلما كانت القيم بعيدة عنه كان التباين أكبر، والتباين هو عبارة عن مجموع مربعات انحرافات القيم عن وسطها الحسابي مقسوماً على عددها . ويمكننا التمييز بين تباين المجتمع σ^2 وتباين العينه s^2 .

- تباين المجتمع يساوي :

$$\sum_{i=1}^{N} \frac{(x_i - \mu)^2}{N} \qquad (3)$$

حيث μ الوسط الحسابي للمجتمع و N حجم المجتمع .

- تباين العينه يساوي

$$S^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1} \qquad (4)$$

حيث $\overline{x}$ هو الوسط الحسابي للعينه و n حجم العينه وعندما يكون حجم العينه $n \geq 30$ نضع في المقام n عوضاً عن n-1 . أما الانحراف المعياري فهو عبارة عن الجذر التربيعي للتباين ويكون لدينا :

- الانحراف المعياري للمجتمع σ يساوي :

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i - u)^2}{N}} \qquad (5)$$

- والانحراف المعياري للعينه s يساوي :

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}} \qquad (6)$$

وكثيراً ما نستخدم الصيغة التالية لحساب الانحراف المعياري لليعنه

$$S = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right)} \qquad (7)$$

كما يرمز أحياناً لتباين المجتمع بالرمز V(X) أو Var(X) ولتباين العينه v(x) أو var (x) فيكون الانحراف المعياري للعينه:

$$s = \sqrt{vas(x)}$$

التغاير والارتباط Covariance & Correlation

نفترض أننا نرغب في دراسة العلاقة بين متغيرين عشوائيين x و y لعينه حجمها n وحدة ، فيكون لدينا n زوج من القيم $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ ان متوسط مجموع حاصل ضرب انحرافات القيم عن الوسط الحسابي للمتغيرين هو التغاير أي أن التغاير ولنرمز له بالرمز Cov (x,y) يساوي:

$$Cov(x, y) = \sum_{i=1}^{n} \frac{(x - \overline{x})(y - \overline{y})}{n-1} \qquad (8)$$

ونضع n عوضاً عن n-1 اذا كان حجم العينه كبيراً $n \geq 30$.

ويستخدم التغاير كمقياس نوعي لمدى وجود علاقة بين المتغيرين x و y . عندما يكون التغاير مساوياً للصفر يعني عدم وجود علاقة بين المتغيرين .

ومن الصعب استخدام التغاير كمقياس لدرجة قوة العلاقة بين المتغيرين لأن قيمته تعتمد على نوع المقياس المستخدم ، لذا من الصعب تحديد ما اذا كان التغاير كبيراً من نظرة سريعة ، لذا يستخدم معامل الارتباط كمقياس درجة قوة العلاقة بين متغيرين وكثيراً ما تستخدم الصيغة التالية لاستخراج التغاير بين متغيرين :

$$\operatorname{cov}(x, y) = \frac{\sum x_i y_i - n\overline{x}\,\overline{y}}{n-1} \qquad (9)$$

ونستخدم معامل الارتباط لقياس درجة قوة الارتباط الخطي بين متغيرين ولنرمز له بالرمز γ ويساوي :

$$r = \frac{\operatorname{cov}(x, y)}{S_x S_y} \qquad (10)$$

حيث S_y , S_x هما الانحراف المعياري للمتغيرين X و y على التوالي : وتتراوح قيمة معامل الارتباط بين -1 و 1 أي $-1 \le r \le 1$ حيث يساوي -1 عندما يكون الارتباط بين المتغيرين y,x تاماً ولكن سالباً ويساوي +1 عندما يكون الارتباط تاماً ولكن موجباً ويساوي الصفر عندما يكون الارتباط الخطي معدوماً . أي لايوجد ارتباط بين المتغيرين . ويمكننا استخدام احدى الصيغتين التاليتين لحساب معامل الارتباط :

$$r = \sum_{i=1}^{n} \frac{(x_i - \overline{x})(y_i - \overline{y})}{(n-1) S_x S_y} \qquad (11)$$

$$r = \sum_{i=1}^{n} \frac{x_i y_i - n\overline{x}\,\overline{y}}{(n-1) S_x S_y} \qquad (12)$$

حيث

$$S_x = \sqrt{\sum_{i=1}^{n} \frac{x_i^2 - n\overline{x}^2}{n-1}}$$

$$S_y = \sqrt{\sum_{i=1}^{n} \frac{y_i^2 - n\overline{y}}{n-1}}$$

ان الصيغ السابقة لاستخراج معامل التغاير ومعامل الارتباط من بيانات العينه هي مقدرات للمعالم المقابلة لها في المجتمع.

معلمة المجتمع : pop. parameter

عند دراسة متغير عشوائي X فان دالة كثافة احتماله تعتمد على مقياس أو عدة مقاييس (ثوابت) كالوسط الحسابي والتباين . وإن معرفة هذه المقاييس تحدد الخصائص الأساسية للمتغير موضوع الدراسة وتسمى الثوابت التي تعتمد عليها دالة كثافة الاحتمال بمعالم المجتمع .

ان معلمة المجتمع تعبير عددي يلخص خصائص جميع قيم المجتمع اذا كانت غير خاضعة للأخطاء. ويتم حساب معالم المجتمع عند استخدام أسلوب الحصر الشامل بشكل تام دقيق أي عندما لا تقع أخطاء . ويعد الوسط الحسابي للمجتمع μ وتباينه σ^2 من أهم معالم المجتمع حيث :

$$\mu = \frac{1}{N} \sum_{i=1}^{N} X_i , \qquad \sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (X_i - \mu)^2$$

الاحصائية العينه : Asample Statistic

غالباً ما تكون معالم المجتمع مجهولة حيث نقوم بتقديرها من بيانات عينه تمثل المجتمع . ان احصائية العينه هي مقدر لمعلمه المجتمع يتم حسابها من بيانات العينه التي تمثل هذا المجتمع . ويعد الوسط الحسابي للعينه $\overline{x}$ وتباين العينه S^2 من احصائيات العينه حيث :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

الاحتمال : Probability

كثيراً ما نستخدم مفهوم الاحتمال في حياتنا اليومية كأن نقول ان احتمال نجاح الطالب في مادة الرياضيات 50% أو 70 % ويتراوح الاحتمال بين الصفر والواحد . إذ كلما كان الحدث أكثر وقوعاً كان الاحتمال أقرب إلى الواحد ، وكلما كان الحدث أقل وقوعاً كان الاحتمال أقرب إلى الصفر . إن احتمال وقوع الحدث الاكيد يساوي الواحد واحتمال عدم وقوعه يساوي الصفر . لنرمز إلى احتمال حدوث الحدث E بالرمز P(E) واحتمال عدم حدوثه بالرمز q(E) حيث : q(E)=1-P(E) وتستخدم كلمة نجاح للاشارة إلى وقوع الحدث وكلمة فشل لعدم وقوعه . وللوصول إلى تعريف دقيق للاحتمال ، لا بد من من تعريف التجربة والحدث . وتعرف التجربة An Experiment بأنها " عملية تجرى تحت ظروف معينه ولا يمكن التنبؤ بنتيجتها بشكل أكيد " وللتجربة نتائج محتملة Possible outcome أما الحدث An event فهو مجموعة النتائج التي لها خصائص محددة في المجموعة الكلية للنتائج Ω.

إذا رمزنا إلى عدد النتائج المحتملة بـــ $N(\Omega)$ وعدد النتائج (الحالات) المواتيه (التي نحصل عليها نتيجة الحدث E) بـ n(E) يكون احتمال حدوث الحدث E ولنرمز له بالرمز P(E) مساوياً لعدد الحالات المواتيه مقسوماً على عدد الحالات الممكنه وذلك عندما يكون جميع النتائج الممكنه في Ω الفرصة نفسها في الحدوث أي أن :

$$P(E) = \frac{n(E)}{N(\Omega)} = \frac{n}{N}$$

ولتوضيح المفاهيم السابقة ، نورد المثال الآتي . اذا أردنا استخراج احتمال اختيار موظف لديه شهادة ماجستير من موظفي احدى الجهات البالغ عددهم 200 موظف اذا كان عدد الذين لديهم ماجستير في هذه الجهة هو 10 موظفين ، فنجد من هذا المثال أن التجربة هي اختيار الموظف للتعرف على مؤهله ، ولدينا عدة حوادث $E_1, E_2, \ldots$ حسب مؤهلات الموظفين حيث E_1 ترمز للحدث اذا كان المؤهل هو الماجستير و E_2 اذا كان مؤهله بكالريوس وهكذا ويكون عدد الحالات الممكنه N=200 وعدد الحالات المواتيه n=10 وبالتالي يكون احتمال الحصول على موظف أختير عشوائياً ومؤهله ماجستير $P(E_1)$ يساوي :

$$P(E_1) = \frac{n(E_1)}{N(\Omega)} = \frac{n}{N} = \frac{10}{200} = .05$$

أي 5% . أما احتمال اختيار موظف مؤهله ليس بشهادة ماجستير فيساوي

$$q(E_1) = 1 - P(E_1) = 1 - .05 = .95$$

أي يساوي 95%

التوقع : Expectation

اذا كان لدينا متغير عشوائي X يمثل عدد أفراد الأسرة لموظفي إحدى الادارات حيث $X: x_1, x_2, \ldots, x_n$ وكانت دالة احتمال أن يكون عدد أفراد الأسرة x_i هو $f(x_i)$ فان التوقع (ويسمى أحياناً التوقع الرياضي أو القيمة المتوقعة) ولنرمز له بالرمز E(X) يساوي :

$$E(X) = \sum_{i=1}^{n} f(x_i)x_i \qquad (14)$$

وتوجد صيغة أخرى للتوقع اذا كان المتغير العشوائي متصلاً باستخدام التكامل :

$$E(X)=\int_{-\infty}^{\infty} x_i f(x_i)dx \qquad (15)$$

حيث f(x) هي دالة كثافة الاحتمال للمتغير العشوائي X . ان القيمة المتوقعة هي الوسط الحسابي للمجتمع μ لأننا اذا استبدلنا في صيغة التوقع $f(x_i)$ بالتكرارات النسبية $\frac{f_i}{n}$ حيث $n=\sum f_i$ فان التوقع يصبح $\sum f_i x_i / n$ أي الوسط الحسابي للعينه التي حجمها n . ان التكرارات النسبية f_i/n تقترب من الاحتمالات $f(x_i)$ كلما زادت قيمة n ويؤدي ذلك إلى تفسير E(X) كقيمة تمثل متوسط المجتمع الذي سحبت منه العينه .

مثال :

فيما يأتي توزيع موظفي احدى الجهات حسب عدد أفراد أسرهم والاحتمالات المقابلة لحجم الأسرة للموظف:

عدد أفراد الأسرة X	صفر	1	2	3	4	5
الاحتمال $f(x_i)$	05.	05.	2.	35.	30.	.05

ما هو متوسط أفراد الأسرة للموظفين (أي القيمة المتوقعة) ؟

الحل :

$$E(X)=\sum_{i=1}^{n} x_i f(x_i)$$

$= 0x.05+1x.05+2x.2+3x.35+4x.3+5x.05$

$= 2.95 = 3$

أي متوسط عدد أفراد الأسرة لمجتمع الموظفين تقريباً يساوي ثلاثة .

أهم التوزيعات الاحتمالية

عند دراستنا للتوزيعات الاحتمالية ، نميز بين نوعين من هذه التوزيعات :

- التوزيعات الاحتمالية للمتغيرات المتقطعة .
- التوزيعات الاحتمالية للمتغيرات المتصلة .

ويعـــد توزيع ذي الحدين مــن أهم التوزيعات الاحتمالية للمتغيرات المتقطعة ، كما يعتبر التوزيع الطبيعي وتوزيع ستيودنت (t) من أهم التوزيعات الاحتمالية المتصلة ولهذه التوزيعات أهمية خاصة عند دراسة العينات لاستخدامها عند تقدير معالم المجتمع وسنقوم بدراسة هذه التوزيعات باختصار .

توزيع ذي الحدين : Binomial distribution

عندما نجري تجربة ما ، فانه يقع الحدث ، نستخدم كلمة نجاح (كلمة نجاح تستخدم للاشارة إلى وقوع الحدث)، وعندما لا يقع الحدث نستخدم كلمة فشل . وعندما نجري التجربة n مرة نستخدم متغيراً عشوائياً X يمثل العدد الكلي لمرات النجاح التي حصلنا عليها أي عدد مرات وقوع الحدث (النجاح) عند تكرار التجربة n مرة . ويسمى المتغير من هذا النوع متغير بحدين .

وعندما نقوم باعداد جدول يحتوي على المتغير العشوائي X والاحتمالات المقابلة لكل قيمة ($f(x_i)$) f نحصل على ما يسمى جدول توزيع المتغير العشوائي .

ان الصيغة المستخدمة لحساب الاحتمالات للقيم الممكنه للمتغير العشوائي ، والتي تسمى دالة الاحتمال لتوزيع ذي الحدين ولنرمز له بالرمز f(x) وذلك عندما تكون نتائج التجربة في المحاولات المختلفة مستقلة عن بعضها البعض ونجد أن :

$$f(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x} \qquad (16)$$

حيث :

P احتمال حدوث الحدث في المحاولة الواحدة للتجربة .

q احتمال عدم حدوثه حيث p+q=1

n عدد مرات تكرار التجربة .

x عدد مرات النجاح التي سنحصل عليها

n! تقرأ مضروب n وهي عبارة عن حاصل ضرب كل الاعداد الصحيحة من 1 إلى n ، مثلاً !4 تساوي 1×2×3×4 كما ان مضروب الصفر !O يساوي واحد .

وعند استخدام توزيع ذي الحدين ، فان الوسط الحسابي لمتغير ذي الحدين يساوي :

$$\mu = np \qquad (17)$$

وتباينه يساوي :

$$\sigma^2 = npq \qquad (18)$$

التوزيع الطبيعي : Normal Distribution

يعد التوزيع الطبيعي (أو المعتاد) أحد الأمثلة المهمة للتوزيع الاحتمالي للمتغير المتصل ويستخدم هذا التوزيع كثيراً في مجال العينات . ويتصف هذا التوزيع بعدة خصائص :

- المتغير العشوائي المتصل x يأخذ قيماً من $-\infty$ إلى $+\infty$.
- ان شكل منحنى التوزيع الطبيعي يشبه الجرس .
- ان قمة المنحنى تقع عند متوسط المجتمع μ والمنحنى متماثل حول μ اذ كل طرف هو صورة مطابقة للطرف الآخر .
- يعتمد التوزيع الطبيعي على معلمتين هما متوسط المجتمع μ وتباين المجتمع σ^2 لذا يشار إلى هذا التوزيع بالرمز $N(\mu,\sigma^2)$ حيث تشير N إلى Normal .
- ان مركز التوزيع يعتمد على μ وشكله يعتمد على الانحراف المعياري σ .

ان دالة كثافة الاحتمال للتوزيع الطبيعي هي :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}} (x-\mu)^2 / \sigma^2 \qquad (19)$$

حيث $-\infty \langle X \langle \infty$

μ الوسط الحسابي للمجتمع

σ الانحراف المعياري للمجتمع

e قيمة ثابتة تساوي تقريباً 2.71828

$\prod$ قيمة ثابتة تساوي تقريباً 3.14159

f(x) دالة كثافة الاحتمال

وهناك ما يسمى التوزيع الطبيعي المعياري Standard Normal distribution وهو توزيع طبيعي متوسطه $\mu = 0$ وتباينه 1 ويرمز لهذا التوزيع N(0,1) ويستخدم الرمز Z للاشارة إلى المتغير العشوائي الذي له توزيع طبيعي ويتم حساب احتمالات أي متغير له توزيع طبيعي من احتمالات منحنى التوزيع الطبيعي المعياري وفقاً للعينه :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad (20)$$

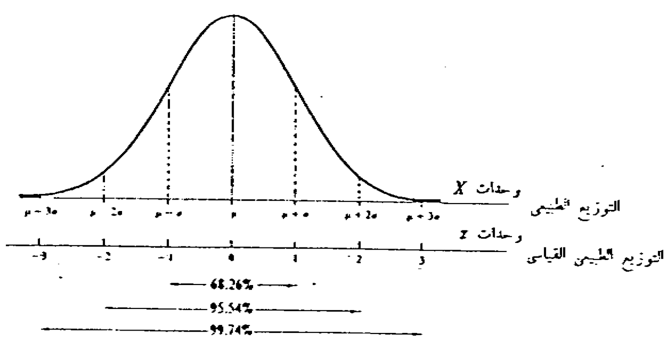
حيث

$$Z = \frac{x-\mu}{\sigma}$$

والشكل أدناه يوضح المنحنى الطبيعي المعياري ، حيث نلاحظ أن المساحة الواقعة بين $Z = \pm 1$ 68.27% وبين $Z = \pm 2$ هي 95.45% وبين $Z = \pm 3$ هي 99.73% وذلك من المساحة الكلية التي تساوي واحد وهناك جداول توضح المساحة تحت المنحنى الطبيعي المحصورة بين الاحداثي Z=0 وأية قيمة موجبة لــ Z ومن هذا الجدول فان المساحة بين أية نقطتين يمكن حسابها باستخدام تماثل المنحنى حول Z=0 .

<u>شكل رقم (1)</u>

<u>منحنى التوزيع الطبيعي</u>



توزيع ستيودنت : Student Dist

يستخدم التوزيع الطبيعي للاستدلال على متوسط المجتمع عندما يكون تباين المجتمع σ^2 غير معلوماً ، أو تكون العينه كبيرة بشكل كاف ، لنتمكن من الاستعاضة عن هذا التباين بتقديره من العينه S^2 . ولكن عندما يكون تباين المجتمع غير معلوم وحجم العينه صغيراً (تكون العينه صغيرة اذا كان حجمها أقل من 30 أي عندما يكون n<30) نستخدم متغيراً جديد يسمى متغير توزيع t أو ستيودنت t وصيغته :

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} \qquad (21)$$

ويشبه هذا المتغير الطبيعي المعياري Z باستثناء القيم الصغيرة جداً للعدد n وتختلف عنه في استخدامنا الانحراف المعياري للعينه s وهذه ميزة تساعدنا على تقدير معالم المجتمع خاصة اذا كان حجم العينه صغيراً .

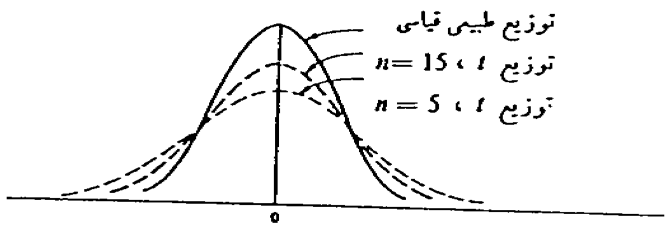
وعند اختيار عدد كبير من العينات ، حجم كل منها n وحدة مجتمع طبيعي ، نحصل على عدد كبير من قيم t ويمكننا الحصول على التوزيع الاحتمالي لـــt والذي دالة كثافته الاحتمالية :

$$f(x) = \frac{Y_0}{(1+\frac{t}{n-1})^{n/2}} \qquad (22)$$

حيث Y_0 مقدار ثابت يجعل المساحة تحت المنحنى مساويه للواحد و n-1 هو عدد درجات الحرية . ان المتغير t يتبع توزيع t اذا كان توزيع المجتمع طبيعياً . كذلك نجد أن هذا التوزيع يكون قريباً جداً من التوزيع الطبيعي عندما يكون حجم العينه كبيراً. وهناك جداول لتوزيع t توضح الاحتمالات لقيمة t . بمستويات متعددة ودرجات حرية n-1 متعددة أيضاً ويوضح الشكل أدناه منحنى توزيع ستيودنت لدرجات حرية 5 و 15 حيث يلاحظ اقترابه من منحنى التوزيع الطبيعي بازدياد حجم العينه n .

شكل رقم (2)

التوزيع الطبيعي ومنحنى توزيع t



تقدير معالم المجتمع : Estimation of pop. Parameters

عندما نقوم بدارسة ظاهرة معينه من بيانات المجتمع نحصل على معلمتي المجتمع σ^2,μ ، ولكن في كثير من الحالات ، نجد أن هاتين المعلمتين غالباً ما تكونان مجهولتين ، فنقوم بتقديرهما من بيانات عينه يتم اختيارها عشوائياً لتمثيل المجتمع تمثيلاً حقيقياً . وسنقوم بدراسة أهم الموضوعات المتعلقة بتقدير معالم المجتمع للتمييز بين مفهومي التقدير والمقدر وخواص المقدر وأنواع التقدير .

التقدير والمقدر : Estimate & Estimator

عندما نسحب عينه ما مفرداتها $x_1, x_2, \ldots, x_n$ ونقوم بتقدير ثوابت دالة كثافة الاحتمال باستخدام هذه المفردات فان القيمة المقدرة لكل ثابت تسمى تقديراً . أما الصيغة (المعادلة) التي تستخدم للوصول إلى التقدير فتسمى مقدراً وهو عبارة عن الدالة التي تعتمد على المفردات ، بينما التقدير عبارة قيمة الدالة عند وضع قيم المشاهدات فيها .

إن قيمة متوسط العينه $\bar{x}$ هو تقدير المتوسط للمجتمع (معلمه المجتمع) أي أن $(\hat{\mu} = \bar{x})$. أما الدالة المستخدمة لتقدير المتوسط فهي عبارة عن المقدر Estimator أي يساوي :

$$\bar{x} = \hat{\mu} = f(x_1, x_2, \ldots, x_n)$$

وبصيغة أخرى نجد أن المقدر يساوي

$$\hat{\mu} = \bar{x} = \tfrac{1}{n}\sum_{i=1}^{n} x_i$$

المقدر الجيد : Best Estimator

ان المقدر لا يختلف من عينه إلى أخرى إلا اذا تغيرت صيغة هذا المقدر ، بينما يختلف التقدير من عينه لأخرى عند استخدام المقدر نفسه لاختلاف قيم المفردات (المشاهدات) من عينه لأخرى . وقد تكون القيمة المقدرة قريبة جداً من القيمة الحقيقية للمجتمع أو بعيدة عنها ويعد المقدر جيداً اذا كان في المتوسط لعدد كبير من العينات يعطي قيماً قريبة جداً من القيم الحقيقية للمجتمع . والمقدر الأقرب إلى معلمه المجتمع هو المقدر الأفضل Best . توجد عدة خواص للمقدر الجيد تساعدنا على استخدامه لتقدير معالم المجتمع **عندما تكون مجهولة** .

خواص المقدر الجيد

للمقارنة بين المقدرات المختلفة ، توجد خواص معينه عندما تتحقق في المقدر يعد محققاً لصفات الجودة ، وهذه الخواص هي :

- عدم التحيز Unbiasedness
- الاتساق Consistency
- الكفاءة Efficiency
- الكفايه Sufficiency

عدم التحيز

يسمى المقدر $\hat{\theta}$ مقدراً غير متحيز للمعلمه θ اذا كان توقعه يساوي هذه المعلمه أي عندما :

$$E(\hat{\theta}) = \sigma \qquad (23)$$

وذلك لجميع قيم θ في Ω_θ حيث تتضمن Ω_θ جميع θ.

تطبيق :

- الوسط الحسابي لعينه عشوائية سحبت من مجتمع متغيره العشوائي x وتوقعه u يساوي :

$$\hat{\mu} = \bar{x} = \tfrac{1}{n}\sum_{i=1}^{n} x_i$$

هو مقدر غير متحيز لـ μ وذلك لأن :

- تباين عينه عشوائية مسحوبة من مجتمع احصائي متيغرة العشوائي x وتباينه σ^2 يساوي :

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

يعد مقدراً غير متحيز لتباين المجتمع S^2 وذلك لأن :

$$E\left(s^2\right) = \sigma^2 \frac{N}{N-1} = S^2$$

الاتساق

اذا كان $\hat{\theta}$ مقدراً للمعلمه θ محسوباً من مفردات عينه حجمها n فان معنى الاتساق أن يؤول المقدر $\hat{\theta}$ احتمالياً إلى القيمة الحقيقية للمعلمه θ عندما يزداد حجم العينه ويصبح قريباً من اللانهايه أي أن:

$$\lim_{n \to \infty} p\left(\left|\hat{\theta} - \theta\right| > \varepsilon\right) = 0 \qquad (24)$$

عندما $\varepsilon > 0$

ويتم ذلك عندما يتحقق الشرطان الآتيان :

$$\lim_{n \to \infty} E\left(\hat{\theta}\right) \to \theta$$
$$\lim_{n \to \infty} V\left(\hat{\theta}\right) \to 0$$

تطبيق :

سحبت عينه عشوائية عدد مفرداتها n من مجتمع متوسطه μ وتباينه σ^2 ، ان متوسط العينه $\bar{x}$ مقدر متسق للمعلمه μ . لاثبات ذلك نعلم أن :

$$\lim_{n \to \infty} \bar{x} \to \mu$$
$$\lim_{n \to \infty} V\left(\bar{x}\right) = \lim_{n \to \infty} \left(\sigma^2 / n\right) \to 0$$

أي تحقق الشرطان اللازمان لاعتبار $\bar{x}$ مقدراً متسقاً .

الكفاءة

اذا كان لدينا مقدران غير متحيزين للمعلمه θ هما θ_2, θ_1 وكان تباين المقدر الأول أصغر من تباين المقدر الثاني أي

$$V\left(\theta_1\right) < V\left(\theta_2\right)$$

يعد المقدر الأول θ_1 أكفأ من المقدر الثاني θ_2 .

تطبيق :

لمقارنة مدى تمركز الوسط الحسابي $\bar{x}$ مع مدى تمركز الوسيط ME (المقدران غير متحيزين) نلجأ إلى مقارنة تباين المقدرين ونختار المقدر ذا التباين الأصغر .

ان تباين الوسط الحسابي والوسيط للعينات الكبيرة هما

$$V\left(\bar{x}\right) = \frac{\sigma^2}{n}$$
$$V\left(ME\right) = \frac{\pi\sigma^2}{2n}$$

حيث: $\pi = 3.1416$

اذا كان حجم العينه محدداً فان

$$\frac{V(\bar{x})}{V(ME)}=\frac{2}{\pi}=.636$$

وهذا يعني تباين الوسط الحسابي أصغر من تباين الوسيط وبالتالي يكون المقدر $\bar{x}$ أكفأ من المقدر ME .

الكفاية

يسمى المقدر $\hat{\theta}$ مقدراً كافياً للمعلمه θ اذا كان المقدر $\hat{\theta}$ قد امتص جميع المعلومات المتوافرة عن المعلمه θ مجهولة القيمة، بحيث بعد معرفة $\hat{\theta}$ نجد أن المعلومات المتبقية لا تفيد في معرفة θ ويمكننا اثبات كفاية المقدر $\hat{\theta}$ باستخدام طريقة التحليل العاملي .

واذا سحبنا عينه عشوائية حجمها n مفردة $x_1,x_2,\ldots,x_n$ وكانت دالة كثافة احتمال كل من هذه المفردات متشابهة $f(x,\theta)$ فان دالة كثافة الاحتمال المشتركة لهذه القيم العشوائية تساوي :

$$g(x_1,x_2,\ldots,x_n;\theta)=f(x_1,\theta),f(x_2,\theta)\ldots f(x_n,\theta)$$

فاذا استطعنا صياغة هذه الدالة بالشكل

$$g(x_1,x_2,\ldots,x_n;\theta)=h(\hat{\theta},\theta)=k(x_1,x_2,\ldots,x_n)$$

حيث: $k(x_1,x_2,\ldots,x_n)$ دالة لا تحتوي على المعلمة θ فإننا نسمى $\hat{\theta}$ بأنه مقدر كاف للمعلمه θ.

تطبيق :

اذا كان الوسط الحسابي للعينه $\bar{x}$ هو مقدر لتوقع المجتمع a فيمكن اثبات ان هذا المقدر كاف لتوقع المجتمع الطبيعي μ.

نعلم ان احتمال الحصول على هذه العينه (دالة كثافة الاحتمال المشتركة) تساوي

$$g(x_1,x_2,\ldots,x_n;\mu)=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-1/2\sigma^2\sum(x_i-a)^2}$$

وباضافة وطرح $\bar{x}$ نجد أن الطرف الأيمن يساوي :

$$\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \bar{e}\,\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\bar{x}-\bar{x})\ -\frac{n}{2\sigma^2}(\bar{x}-a)^2$$
$$=k(x_1,x_2,\ldots,\bar{x})h(\bar{x},a)$$

أي أن الوسط الحسابي $\bar{x}$ هو مقدر كاف لتوقع التوزيع الطبيعي .

التقدير بنقطة والتقدير بفترة : Point & Interval estimate

كما ذكرنا سابقاً ان أهم الأهداف التي يهتم بها الباحث هي تقدير معالم المجتمع كالوسط الحسابي والانحراف المعياري من بيانات عينه عشوائية ويمكننا التمييز بين نوعين من التقدير :

- التقدير بنقطة Point estimate

يعد التقدير بنقطة النوع الأكثر شيوعاً من أنواع التقدير ، خاصة لدى غير الاحصائيين . والتقدير بنقطة هو تقدير معلمه المجتمع برقم واحد (أو قيمة وحيدة) ، مثلاً الوسط الحسابي للعينه $\bar{x}$ هو تقدير بنقطة لوسط المجتمع μ . كذلك تقدير نسبة المجتمع من بيانات عينه p هو تقدير بنقطة لنسبة المجتمع P .

- التقدير بفترة ثقة Confidence Interval estimate

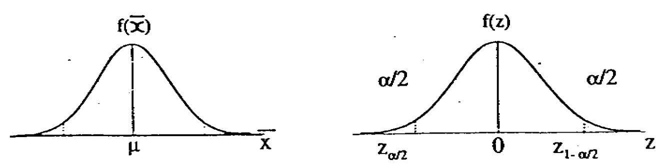
يسمى المدى الذي تقع فيه القيمة الحقيقية لمعلمه مجتمع ما بدرجة ثقة معينه بفترة الثقة . والحد الأعلى والحد الأدنى لهذه الفترة تسمى حدود الثقة Confidence limits . ونستطيع حساب الاحتمالات لفترة الثقة التي تحتوي على القيمة الحقيقية وتكون هذه الاحتمالات صحيحة في حال استخدام المعاينه العشوائية البسيطة . كما أنه لا يمكن حساب حدود الثقة باحتمالات صحيحة من بيانات عينات مسحوبة من مجتمعات **مجهولة التوزيع** . فاذا كان للتقدير توزيع طبيعي وكان الخطأ المعياري للتقدير معروفاً فاننا نستطيع معرفة **احتمال وقوع خطأ** في التقدير أكبر من أي قيمة أخرى . لكن التقدير قد لا يتوزع بصورة طبيعية مما يجعل هذه الاحتمالات غير دقيقة . ولكن اذا كان حجم العينه كبيراً وكان التقدير غير متحيز فاننا نستطيع بمساعدة **جداول التوزيع** الطبيعي **ومعرفة الخطأ المعياري** للتقدير ، حساب فترة الثقة للقيمة الحقيقية لمعلمة المجتمع . اذا كان x متغيراً عشوائياً موزعاً طبيعياً بمتوسط μ وانحراف معياري σ فان القيمة المعيارية :

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \qquad (25)$$

تكون موزعة توزيعاً طبيعياً لمتوسط صفر وانحراف معيار واحد . ان للوسط الحسابي للعينه العشوائية البسيطة $\bar{x}$ المقدر من عينه حجمها n وحدة (مسحوبة من مجتمع له توزيع طبيعي وله متوسط μ وتباينه σ^2/n لذا نجد أن للقيمة المعيارية Z توزيعاً طبيعياً معيارياً .

<u>شكل رقم (3)</u>

<u>منحنى التوزيع الطبيعي ومنحنى التوزيع الطبيعي المعياري</u>



شكل (١) التوزيع الطبيعى

شكل (٢) التوزيع الطبيعى المعيارى

ويمكن القول كما يتضح من الشكل أعلاه أن :

$$p\left(z_{\alpha/2} \le \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \le z_{1-\alpha/2} \right) = 1 - \alpha$$

حيث أن $Z_{\alpha/2}$ هي القيمة التي تسبقها مساحة $\alpha/2$ تحت المنحنى ، أما $Z_{1-\alpha/2}$ هي القيمة التي تسبقها مساحة $1-\alpha/2$ تحت المنحنى ، أما α فهي المساحة المظللة تحت المنحنى خارج فترة الثقة ، و $1-\alpha$ هي درجة أو معامل الثقة ، ويمكن القول أن فترة الثقة للوسط الحسابي :

$$\bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \qquad (26)$$

(قيمة $Z_{\alpha/2}$ سالبه وتستخرج من جداول التوزيع الطبيعي ، أما قيمة $Z_{1-\alpha/2}$ فهي موجبة مثلاً مستوى ثقة 95% نجد أن قيمة $Z_{1-\alpha/2}$ تساوي 1.96 أما $Z_{\alpha/2}$ تساوي (1.96-) .

ان تباين المجتمع غير معلوم في كثير من الحالات ، لذا نستخدم توزيع t (ستيودنت) ويكون الخطأ المعياري للمتوسط هو $\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}}$ وتكون القيمة المعيارية هي :

$$t = \frac{\bar{x} - \mu}{\hat{\sigma}(\bar{x})}$$

موزعة حسب توزيع t بدرجات حرية n-1 وتكون فترة الثقة في حالة السحب مع الاعادة :

$$\bar{x} + t_{(\alpha/2, n-1)} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{(1-\alpha/2, n-1)} \frac{s}{\sqrt{n}} \qquad (27)$$

حيث:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

وتأخذ قيم t من جدول توزيع ستيودنت (t) بمستوى ثقة معين $(1-\alpha)\%$ درجات حريه n-1

أما في حالة السحب مع عدم الاعادة ، تصبح فترة الثقة بعد ادخال معامل تصحيح المجتمع المحدود $\frac{N-n}{N-1}$:

$$\bar{x} + t_{(\alpha/2, n-1)} \frac{s}{\sqrt{n}} \sqrt{1-f} \leq u \leq \bar{x} + t_{(1-\alpha/2, n-1)} \frac{s}{\sqrt{n}} \sqrt{1-f}$$

(يلاحظ أننا استخدمنا $\frac{s}{\sqrt{n}}\sqrt{1-f}$ حيث تم الحصول على هذا المقدار من $\sigma^2 \bar{x} = \frac{\sigma^2}{n}$ بعد ضربها في معامل تصحيح المجتمع المحدود :

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \frac{N-n}{N-1} = \frac{s^2}{n} \left(\frac{N-n}{N} \right)$$

s^2 هو مقدر غير متحيز لــ S^2 عندما يكون تباين المجتمع مجهولاً ، لذا وضعنا الانحراف المعياري للعينه s في صيغة فترة الثقة . وهكذا نلاحظ أنه لاستخراج فترة الثقة لا بد من تقدير الخطأ المعياري $\widehat{\sigma x}$ أو تباين العينه s^2 في حالة عدم معرفة تباين المجتمع σ^2 أو التباين المعدل للمجتمع S^2 .

ويمكننا القول لتوضيح مفهوم حدود الثقة ، لو سحبنا عدداً كبيراً من العينات ذات الحجم n مفردة من المجتمع نفسه وحسبنا حدود الثقة لكل عينة فان 95% (اذا كانت $\alpha = .05$) من هذه الحدود لا بد وان تحتوي على متوسط المجتمع μ .