الوحدة الرابعة

الإحصاء

Statistics

مشروع الوحدة: ما هي أفضل طريقة لإيجاد وظيفة؟

1 مقدمة المشروع: بعد التخرج يواجه الحاصلون على الإجازات والشهادات الجامعيّة تحدّ جديد هو الانخراط في سوق العمل.

2 الهدف: هو البحث عن فرص عمل من خلال القيام بعدة خطوات ومحاولات متنوعة واستخدام العديد من الوسائل.

3 اللوازم: حاسوب — شبكة الإنترنت.

4 أسئلة حول التطبيق:

a كيف ستختار عينة عشوائية من الموظفين للاستفسار عن الوسيلة التي استخدموها في إيجاد وظيفتهم؟

b ما الخيارات التي اكتشفتها؟ نظّمها في استمارة.

(إرشاد):

- من خلال الأصدقاء والمعارف.
- من خلال الإعلانات في الصحف والمجلات.
- من خلال الوكالات المختصة في الربط بين سوق العمل وطالبي الوظائف.
- من خلال البحث عبر شبكة الإنترنت.
- من خلال التقدم مباشرة لطلب وظيفة من الشركة المختصة أو اعتماد وسيلة أخرى (اذكرها ...).

c حدّد النسب المئوية لكل خيار ممّا سبق.

5 التقرير: اكتب تقريرًا مفصلًا يحدّد النسب التي حصلت عليها من خلال العينة العشوائيّة التي اعتمدتها مكوّنًا جدولًا بالنسب المئوية عن كل وسيلة تمّ استخدامها لإيجاد وظيفة.

القرار: ضمّن تقريرك بعض الاقتراحات والنصائح والاستنتاجات التي نتجت عن تلك الدراسة.

دروس الوحدة

الارتباط والانحدار	اختبارات الفروض الإحصائية	التقدير
4–3	4–2	4–1

الوحدة الرابعة

أين أنت الآن (المعارف السابقة المكتسبة)

- تعلمت مقاييس النزعة المركزية: المتوسط الحسابي — الوسيط — المنوال.
- تعلمت المجتمع الإحصائي.
- تعلمت العينة العشوائية وأنواعها واستخداماتها.

ماذا سوف تتعلم؟

- يُعرّف المعلمة والإحصاءة.
- إيجاد التقدير بنقطة.
- إيجاد التقدير بفترة ثقة.
- استكشاف الفروض الإحصائية.
- يُعرّف الاختبارات الإحصائية ويجريها.
- اتخاذ القرار المناسب.
- يتعرف الارتباط والانحدار.
- يوجد مُعامل ارتباط بيرسون.
- يكتب معادلة خط الانحدار ويتنبأ.

أضف إلى معلوماتك

في الوسائل الإعلامية المرئية والمسموعة والمكتوبة تطالعك نتائج إحصائية تتحدث عن توقعات أحداث معينة تتناول انتخابات نيابية أو رئاسية أو مبيعات أو مباريات... وأكثر ما يستوقفك هو نسبة مئوية معينة مع هامش خطأ محدد والسؤال المهم هو: كيف يتم التقدير وكيف يحتسب هامش الخطأ؟ توفر دروس هذه الوحدة فرصة أمام الطلاب للتعرّف على التقدير وهامش الخطأ والفروض الإحصائية وكيفية احتسابها.

كما يتعرف الطلاب على مفهوم الارتباط والانحدار ويحتسبوا معامل ارتباط بيرسون ثم يكتبوا معادلة خط الانحدار ويتنبؤا نتائج معينة.

المصطلحات الأساسية

المعلمة — الإحصاءة — التقدير — التقدير بنقطة — فترة الثقة — التقدير بفترة الثقة — درجة الثقة — التوزيع الطبيعي — القيمة الحرجة — هامش الخطأ — الخطأ المعياري — خواص التوزيع t — الفروض الإحصائية — المقياس الإحصائي — فرض العدم — الفرض البديل — القرار — الانحدار — المخطط الانتشاري — الارتباط — مُعامل الارتباط الخطي — خواص مُعامل الارتباط — مُعامل ارتباط بيرسون — التنبؤ.

4-1

التقدير

Estimation

سوف تتعلم
- التقدير بنقطة.
- التقدير بفترة الثقة.
- هامش الخطأ.

المفردات والمصطلحات:
- المعلمة Parameter
- الإحصاءة Statistic Function
- تقدير المعلمة Parameter Estimate
- تقدير Estimation
- تقدير بنقطة Point Estimate
- تقدير بفترة الثقة Confidence Interval Estimation
- درجة الثقة (مستوى الثقة) (Level of Confidence) Degree of Confidence
- نسبة الخطأ (مستوى المعنوية) Percentage of error (Significance Level)
- القيمة الحرجة Critical Value
- هامش الخطأ Margin of Error
- درجات الحرية Degree of Freedom

دعنا نفكر ونتناقش

متوسط عدد الرحلات الجويّة المغادرة يوميًا خلال شهر يونيو من أحد المطارات هو 75 رحلة.
هل يمكن استخدام هذه العيّنة لتقدير متوسّط عدد الرحلات μ خلال أشهر السنة؟ لماذا؟
وما هي أفضل وسيلة للتقدير لنقترب من الحقيقة؟

سبق لنا في الصف الحادي عشر تعريف المجتمع الإحصائي والعينة العشوائية والأسباب التي تؤدي إلى أخذ العينات لدراسة المجتمع بدلًا من الحصر الشامل، وذلك لتقدير المتوسط (الوسط) الحسابي للمجتمع μ أو الانحراف المعياري له σ.

ويعتبر الوسط الحسابي للمجتمع μ والانحراف المعياري للمجتمع σ من معالم المجتمع، وعادة ما تكون هذه المعالم مجهولة.

ولتقدير هذه المعالم نلجأ إلى سحب عينة عشوائية منه، ثم نحسب المتوسط الحسابي للعينة $\overline{x}$ أو الانحراف المعياري S والذي يعتمد على قيم العينة ولا يعتمد على معالم المجتمع.

المعلمة (Parameter): هي ثابت يصف المجتمع أو يصف توزيع المجتمع كالمتوسط الحسابي μ أو الانحراف المعياري σ.

الإحصاءة (Statistic Function): هو اقتران تتعين قيمته من العينة كالمتوسط الحسابي $\overline{x}$ أو الانحراف المعياري S.

تقدير المعلمة (Parameter Estimate): هو إحصاءة تعتمد على قيم العينة وتعكس قيمة قريبة لمعلمة المجتمع ككلّ وتوزيعه.

في هذا الدرس سوف تتعرف طريقتين تساعدان على إيجاد قيم تقديرية لبعض معالم مجتمع معين:
- **طريقة أولى:** التقدير بنقطة.
- **طريقة ثانية:** التقدير بفترة الثقة.

تذكر:
الاقتران هو قيمة تربط مفردات معينة وتنتج منها.

التقدير بنقطة　Point Estimate

التقدير بنقطة هي قيمة وحيدة محسوبة من العينة تستخدم لتقدير معلمة مجهولة من معالم المجتمع.

فمثلًا المتوسط الحسابي للعينة العشوائية $\overline{x}$ يستخدم كتقدير بنقطة للمتوسط الحسابي للمجتمع μ، وكذلك الانحراف المعياري للعينة S يستخدم كتقدير بنقطة للانحراف المعياري للمجتمع σ.

التقدير بفترة الثقة　Confidence Interval Estimation

علمنا مما سبق أن لكل مجتمع معالم منها المتوسط الحسابي μ والانحراف المعياري σ، ودرسنا كيفية إيجاد التقدير بنقطة لتلك المعالم. وعلمنا أن التقدير بنقطة لإحدى معالم المجتمع هو قيمة وحيدة محسوبة من العينة، ولذلك فإن احتمال الخطأ في التقدير بنقطة يكون كبيرًا.

وبناء عليه فإنه من الأفضل إيجاد فترة معينة يتوقع أن تقع معلمة المجتمع داخلها بنسبة معينة أو باحتمال معين. إن مثل هذه الفترة تسمى فترة الثقة.

فترة الثقة　Confidence Interval

هي فترة طرفاها متغيران عشوائيان (أي أنها فترة عشوائية) تحوي إحدى معالم المجتمع بنسبة معينة تسمى درجة الثقة (مستوى الثقة).

التقدير بفترة الثقة

هو إيجاد فترة معينة يتوقع أن تقع معلمة المجتمع داخلها بنسبة معينة أو احتمال معيّن.

درجة الثقة (مستوى الثقة)

Degree of Confidence (Level of Confidence)

إن درجة الثقة أو مستوى الثقة هو احتمال $(1 - \alpha)$ أن تكون فترة الثقة تحوي القيمة الحقيقية لمعلمة المجتمع قيد الدراسة، وعادة يعبّر عنها كنسبة مئوية.

أما α فهي نسبة الخطأ في التقدير وتسمى **مستوى المعنوية** أو مستوى الدلالة.

فمثلًا:

- إذا كانت $\alpha = 0.05$ حينها تكون درجة الثقة $1 - \alpha = 0.95$ أي 95%
- إذا كان مستوى الثقة 90% فإن مستوى المعنوية $\alpha = 10\%$
- أيضًا إذا كان مستوى الثقة 99% فإن مستوى المعنوية $\alpha = 1\%$

ومن هذه الخيارات الثلاثة، يعتبر مستوى الثقة 95% هو الأكثر انتشارًا لأنه يؤمن التوازن الأنسب بين الدقة الموضحة من خلال طول فترة الثقة والدقة الموضحة من خلال مستوى الثقة.

تذكر:

المتوسط الحسابي لعينة = $\dfrac{\text{مجموع قيم البيانات}}{\text{عدد البيانات}}$

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

الانحراف المعياري لعينة:

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

تذكر:

المتوسط الحسابي للمجتمع

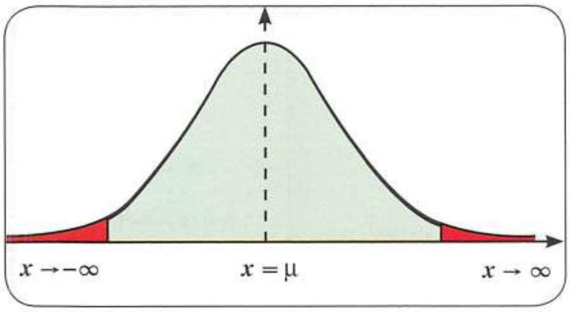
$$\mu = \frac{\Sigma x_i}{N}$$

الانحراف المعياري للمجتمع:

$$\sigma = \sqrt{\frac{\Sigma(x_i - \overline{x})^2}{N}}$$

التوزيع الطبيعي Normal Distribution

تعرفنا فيما سبق على بيان منحنى التوزيع الطبيعي، وعلمنا من خواص التوزيع الطبيعي ما يلي:

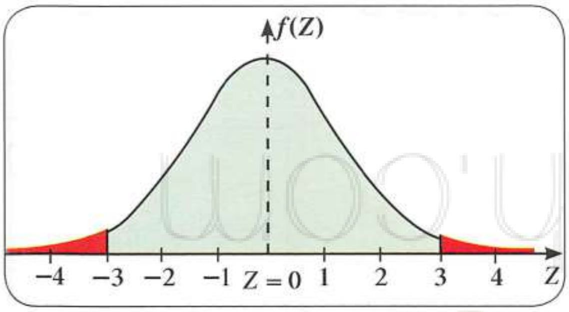


شكل (1)

- المتوسط الحسابي = الوسيط = المنوال.
- يكون بيان المنحنى على شكل ناقوس (جرس) متماثل حول محوره $(x = \mu)$.
- يمتد المنحنى من طرفيه إلى ∞ وإلى $-\infty$ (لا يقطع المحور الأفقي).
- المساحة تحت المنحنى تساوي الواحد الصحيح (وحدة مساحة).

المستقيم الرأسي $x = \mu$ يقسم المساحة تحت المنحنى إلى منطقتين متماثلتين مساحة كل منهما تساوي نصف (وحدة مساحة) كما في الشكل (1).

منحنى التوزيع الطبيعي المعياري:



شكل (2)

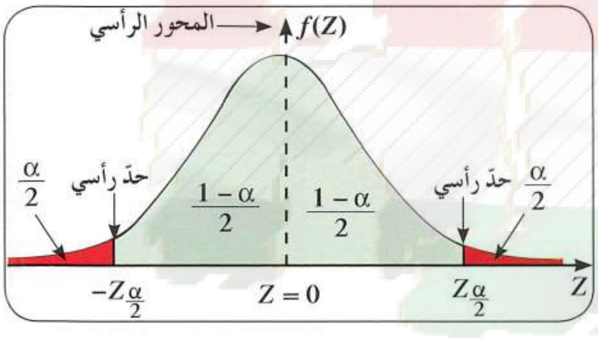
إذا كان المتوسط الحسابي للتوزيع الطبيعي $\mu = 0$ والانحراف المعياري $\sigma = 1$ يسمى التوزيع الطبيعي **بالتوزيع الطبيعي المعياري**.

الشكل المرسوم يمثّل بيان منحنى التوزيع الطبيعي المعياري.

المستقيم $Z = 0$ هو محور التماثل للمنحنى. تأخذ Z قيمًا موجبة وتزداد جهة اليمين بينما تأخذ Z قيمًا سالبة وتنقص جهة اليسار.

القيمة الحرجة $Z_{\frac{\alpha}{2}}$ Critical Value

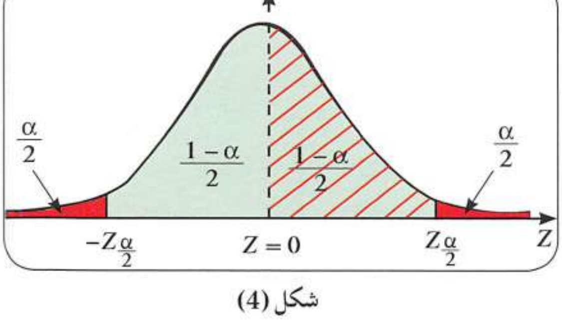
الشكل (3) المرسوم يبيّن منحنى التوزيع الطبيعي المعياري.



شكل (3)

- نعلم أن مساحة المنطقة تحت منحنى التوزيع الطبيعي المعياري تساوي الواحد (وحدة مساحة) والمحور الرأسي يقسم المنطقة تحت المنحنى إلى قسمين متطابقين مساحة كل منهما تساوي $\frac{1}{2}$ وحدة مساحة. ومجموع مساحتي الجزئين باللون الأحمر معًا تساوي α وتكون مساحة كل جزء منهما تساوي $\frac{\alpha}{2}$ وعليه تكون مساحة كل من الجزئين باللون الأخضر على جانبي المحور الرأسي $\frac{1}{2} - \frac{\alpha}{2}$ أي $\frac{1-\alpha}{2}$
- نعبّر عن الحدين الرأسيين بالرمز $Z_{\frac{\alpha}{2}}$ وبالرمز $-Z_{\frac{\alpha}{2}}$ حيث $Z_{\frac{\alpha}{2}}$ يفصل المنطقة التي مساحتها $\frac{\alpha}{2}$ من ذيل الطرف الأيمن عن المنطقة التي مساحتها $\frac{1-\alpha}{2}$ من المستقيم $Z = 0$، بينما $-Z_{\frac{\alpha}{2}}$ يفصل المنطقة التي مساحتها $\frac{\alpha}{2}$ من ذيل الطرف الأيسر عن المنطقة التي مساحتها $\frac{1-\alpha}{2}$ من المستقيم $Z = 0$.

ملاحظة: $Z_{\frac{\alpha}{2}} = Z_{\frac{1-\alpha}{2}} \quad , \quad -Z_{\frac{\alpha}{2}} = -Z_{\frac{1-\alpha}{2}}$



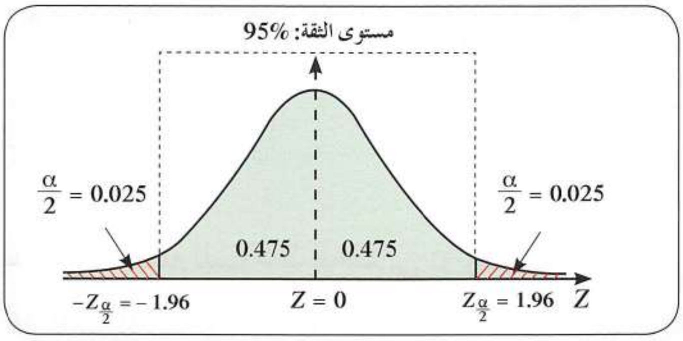
شكل (4)

إيجاد القيمة الحرجة من جدول التوزيع الطبيعي المعياري:

لإيجاد قيمة $Z_{\frac{\alpha}{2}}$ المناظرة للمساحة تحت المنحنى نحسب المساحة $\frac{1-\alpha}{2}$ التي تقع على يسار $Z_{\frac{\alpha}{2}}$ ويمين الصفر أي في الفترة $[0, Z_{\frac{\alpha}{2}}]$ ثم نكشف عنها في الجدول المرفق في نهاية الوحدة صفحة 194 حيث العمود الأول قيم Z ابتداءً من 0.0 وحتى 3.1 وأكثر. والصف الأول يمثل الأجزاء من المئة لقيم Z، ومنه يمكن تحديد قيمة $Z_{\frac{\alpha}{2}}$ وذلك بجمع قيمتي الصف والعمود لـ Z.

مثال (1)

أوجد القيمة الحرجة $Z_{\frac{\alpha}{2}}$ المناظرة لمستوى ثقة 95% باستخدام جدول التوزيع الطبيعي المعياري.



جدول التوزيع الطبيعي المعياري (Z)

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817

الحل:

$\because$ مستوى الثقة هو 95%

$$\therefore\ 1-\alpha = 0.95 \qquad \therefore\ \frac{1-\alpha}{2} = \frac{0.95}{2} = 0.475$$

نأخذ جدول التوزيع الطبيعي المعياري (Z) صفحة 194

نبحث في الجدول عن 0.4750 فنجدها على التقاطع الأفقي/العمودي للعددين على الترتيب: 1.9 ، 0.06

وبالتالي القيمة الحرجة هي:

$$Z_{\frac{\alpha}{2}} = 1.9 + 0.06$$

$$\therefore\ Z_{\frac{\alpha}{2}} = 1.96$$

حاول أن تحل

1. أوجد القيمة الحرجة $Z_{\frac{\alpha}{2}}$ المناظرة لمستوى ثقة 97% باستخدام جدول التوزيع الطبيعي المعياري.

هامش الخطأ Margin of Error

أوّلًا: الخطأ بالتقدير بنقطة Point Estimation Error

علمنا فيما سبق أنه يمكن استخدام المتوسط الحسابي للعينة $\overline{x}$ كتقدير بنقطة للمتوسط الحسابي μ للمجتمع.

ومن المتوقع أن تكون قيمة المتوسط الحسابي للعينة $\overline{x}$ غير مساوية لقيمة المتوسط الحسابي μ للمجتمع.

تسمى **القيمة المطلقة للفرق بين القيمتين السابقتين بالخطأ المعياري** وتساوي $\frac{\sigma}{\sqrt{n}}$ حيث **σ الانحراف المعياري للمجتمع**، **n عدد قيم العينة** (أو حجم العينة).

ثانيًا: الخطأ بالتقدير بفترة Interval Estimation Error

والآن نتعرض للخطأ بالتقدير بفترة فعندما نستخدم عينة لتقدير المتوسط الحسابي μ لمجتمع، يكون الخطأ في التقدير هو القيمة المطلقة للفرق بين المتوسط الحسابي للعينة $\overline{x}$، والمتوسط الحسابي μ للمجتمع.

هامش الخطأ E

عند استخدام بيانات عيّنة لتقدير المتوسط الحسابي μ لمجتمع، يكون هامش الخطأ، يرمز إليه بـ E، القيمة العظمى الأكثر ترجيحًا عند درجة ثقة $(1-\alpha)$ للفرق بين المتوسط الحسابي $\overline{x}$ للعيّنة والمتوسط الحسابي μ للمجتمع.

يسمى أيضًا هامش الخطأ الأكبر في التقدير، ويمكن إيجاده بأخذ ناتج ضرب القيمة الحرجة $Z_{\frac{\alpha}{2}}$ والخطأ المعياري $\frac{\sigma}{\sqrt{n}}$.

$$E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

عند درجة ثقة $(1-\alpha)$.

وحتى يكون الخطأ في التقدير أقل ما يمكن يجب أن تتحقق المتباينة:

$$|\overline{x}-\mu| < E$$

أي أن: $|\mu-\overline{x}| < E$

$$-E < \mu-\overline{x} < E$$

$$\overline{x}-E < \mu < \overline{x}+E$$

وعليه تكون فترة الثقة هي: $(\overline{x}-E \ , \ \overline{x}+E)$

التقدير بفترة الثقة للمتوسط الحسابي μ للمجتمع الإحصائي

Confidence Interval Estimation for the Mean Value μ of Statistical Population

أوّلاً: إذا كان التباين σ^2 للمجتمع معلوم

إذا أخذت عينة عشوائية حجمها n من مجتمع طبيعي $(\mu \ , \ \sigma^2)$ حيث تباينه σ^2 معلوم وحجم العينة $n > 30$ أو $n \leq 30$

فإن تقدير فترة الثقة للمتوسط الحسابي μ هو:

$(\overline{x}-E \ , \ \overline{x}+E)$ عند درجة ثقة $(1-\alpha)$

حيث $\overline{x}$ المتوسط الحسابي للعينة، E هامش الخطأ.

وتسمى القيمتان $\overline{x}-E \ , \ \overline{x}+E$ **طرفي فترة الثقة.**

ملاحظة: عند إيجاد فترة الثقة سنكتفي بدرجة الثقة 95% والتي تناظرها القيمة الحرجة $Z_{\frac{\alpha}{2}} = 1.96$
(من جدول التوزيع الطبيعي المعياري).

الخطوات المتبعة لإيجاد فترة الثقة للمتوسط الحسابي μ

إذا كانت σ^2 معلومة و $n > 30$ أو $n \leq 30$.

1. نوجد القيمة الحرجة $Z_{\frac{\alpha}{2}}$ المناظرة لدرجة ثقة 95% وهي 1.96
2. نوجد هامش الخطأ $E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$، حيث σ هي الانحراف المعياري للمجتمع.
3. نوجد فترة الثقة $(\overline{x}-E \ , \ \overline{x}+E)$.

تفسير فترة الثقة

عند اختيار عينات عشوائية مختلفة متساوية في الحجم (n) وحساب حدود فترة الثقة لكل عينة فإننا نتوقع أن 95% من فترات الثقة هذه تحوي القيمة الحقيقية للمتوسط الحسابي للمجتمع (μ).

فمثلًا عند اختيار 100 عينة عشوائية ذات الحجم نفسه (n) وفي كل مرة نحسب $\overline{x}$ وفترة الثقة فإننا نتوقع أن 95 فترة تحوي μ الحقيقية و5 فترات لا تحويها.

مثال (2)

أجريت دراسة لعينة من الإناث حول معدل النبض لديهن فإذا كان حجم عينة الإناث $n = 40$ والانحراف المعياري لمجتمع الإناث $\sigma = 12.5$ والمتوسط الحسابي للعينة $\overline{x} = 76.3$. باستخدام مستوى ثقة 95%

1. أوجد هامش الخطأ.
2. أوجد فترة الثقة للمتوسط الحسابي للمجتمع الإحصائي μ.
3. فسّر فترة الثقة.

الحل:

1. $\because$ مستوى الثقة 95%

$\therefore$ القيمة الحرجة: $Z_{\frac{\alpha}{2}} = 1.96$

نلاحظ أن σ معلومة $\quad \therefore \; E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

$\overline{x} = 76.3 \quad , \quad \sigma = 12.5 \quad , \quad \because \; n = 40$

هامش الخطأ:

$$E = 1.96 \times \frac{12.5}{\sqrt{40}}$$

$$E \approx 3.87379$$

$\therefore$ هامش الخطأ ≈ 3.8738

2. فترة الثقة هي:

$$(\overline{x} - E \; , \; \overline{x} + E)$$

$$= (76.3 - 3.8738 \; , \; 76.3 + 3.8738)$$

$$= (72.4262 \; , \; 80.1738)$$

3. عند اختيار 100 عينة عشوائية ذات الحجم نفسه $(n = 40)$ وحساب حدود فترة الثقة لكل عينة فإننا نتوقّع أن 95 فترة تحوي القيمة الحقيقية للمتوسط الحسابي للمجتمع μ.

حاول أن تحل

2. من المثال (2) إذا أجريت الدراسة على عينة أخرى من الإناث حجمها 25 والانحراف المعياري لمجتمع الإناث $\sigma = 3.6$ والمتوسط الحسابي للعينة $\overline{x} = 18.4$

باستخدام مستوى ثقة 95%

1. أوجد هامش الخطأ.
2. أوجد فترة الثقة للمتوسط الحسابي للمجتمع الإحصائي μ.
3. فسّر فترة الثقة.

ثانيًا: إذا كان التباين σ^2 للمجتمع غير معلوم وحجم العينة $n > 30$

الخطوات المتبعة لإيجاد فترة الثقة للمتوسط الحسابي μ

1. نوجد القيمة الحرجة $Z_{\frac{\alpha}{2}}$ المناظرة لدرجة ثقة 95% وهي 1.96
2. نوجد هامش الخطأ $E = Z_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$ ، حيث S هي الانحراف المعياري للعينة.
3. نوجد فترة الثقة $(\overline{x} - E \; , \; \overline{x} + E)$.

مثال (3)

عينة عشوائية حجمها 36، فإذا كان المتوسط الحسابي للعينة 60 وتباينها 16، باستخدام مستوى ثقة 95%:

1. أوجد هامش الخطأ.
2. أوجد فترة الثقة للمتوسط الحسابي للمجتمع الإحصائي μ.
3. فسّر فترة الثقة.

الحل:

حجم العينة: $n = 36$, المتوسط الحسابي: $\overline{x} = 60$

التباين: $S^2 = 16$ ، الانحراف المعياري: $S = 4$

1. $\because$ مستوى الثقة 95%

$$\therefore \; Z_{\frac{\alpha}{2}} = 1.96$$

$\because$ σ^2 غير معلوم ، $n > 30$

$$E = Z_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$$
$$= 1.96 \times \frac{4}{\sqrt{36}}$$
$$= 1.3066$$

$\therefore$ هامش الخطأ ≈ 1.3067

2. فترة الثقة هي:

$$(\overline{x} - E \;,\; \overline{x} + E)$$
$$= (60 - 1.3067 \;,\; 60 + 1.3067)$$
$$= (58.6933 \;,\; 61.3067)$$

3. عند اختيار 100 عينة عشوائية ذات الحجم نفسه ($n = 36$) وحساب حدود فترة الثقة لكل عينة فإننا نتوقّع أن 95 فترة تحوي القيمة الحقيقية للمتوسط الحسابي للمجتمع μ.

حاول أن تحل

3. أخذت عينة عشوائية من مجتمع طبيعي حجمها $n = 81$ ومتوسطها الحسابي $\overline{x} = 50$، وانحرافها المعياري $S = 9$، باستخدام مستوى ثقة 95%.
   1. أوجد هامش الخطأ.
   2. أوجد فترة الثقة للمتوسط الحسابي للمجتمع الإحصائي μ.
   3. فسّر فترة الثقة.

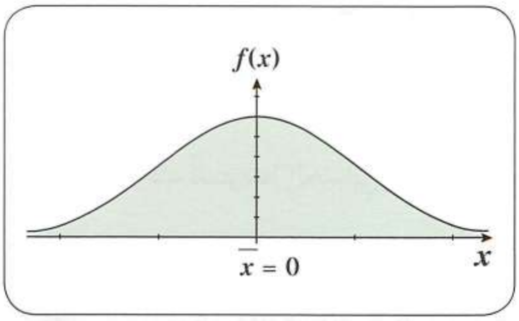
ثالثًا: إذا كان التباين σ^2 للمجتمع غير معلوم وحجم العينة $n \leqslant 30$

إذا أخذت عينة عشوائية حجمها n من مجتمع طبيعي تباينه σ^2 غير معلوم وحجم العينة $n \leqslant 30$ فإن توزيع العينة لا يؤول إلى التوزيع الطبيعي وفي هذه الحالة يلزمنا استخدام توزيع آخر هو توزيع t **للعينات الصغيرة** التي حجمها $n \leqslant 30$ ويكون تقدير فترة الثقة $(1 - \alpha)$ للمتوسط الحسابي μ هو $(\overline{x} - E \;,\; \overline{x} + E)$.

حيث $\overline{x}$ المتوسط الحسابي للعينة، E هامش الخطأ.

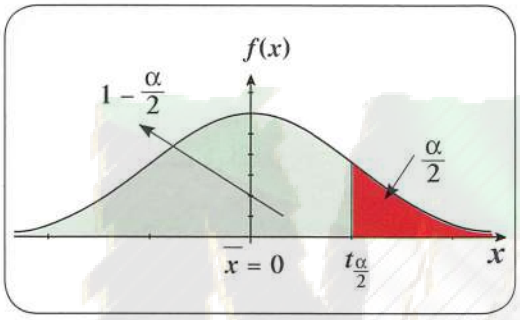
خواص التوزيع t — Properties of t Distribution

1. توزيع متماثل حول متوسطه الحسابي والذي يساوي صفرًا، ويمتد إلى ∞ من جهة اليمين وإلى $-\infty$ من جهة اليسار ويزداد قربًا من الصفر في الجهتين.



2. انحرافه المعياري أكبر من الواحد.
3. يعتمد هذا التوزيع على درجات الحرية والتي تساوي (حجم العينة – 1) أي $(n-1)$.
4. التوزيع t يشبه التوزيع الطبيعي إلّا أن قمته أكثر انخفاضًا من التوزيع الطبيعي.
5. كلما زادت درجات الحرية اقترب هذا التوزيع من التوزيع الطبيعي ويقترب انحرافه المعياري إلى الواحد الصحيح.

إيجاد القيمة الحرجة من جدول توزيع t



لإيجاد القيمة الحرجة من جدول توزيع t حيث يبيّن العمود الأول قيم درجات الحرية $(n-1)$ وتبدأ من 1 إلى 30 وأكثر والصف الأول يمثّل قيم $\frac{\alpha}{2}$ ومنه يمكن تحديد $t_{\frac{\alpha}{2}}$. لاحظ أن:

$$t_{\frac{\alpha}{2}} = t_{1-\frac{\alpha}{2}}$$

هامش الخطأ للمتوسط الحسابي μ للمجتمع الإحصائي (في حالة σ^2 غير معلوم، $n \leq 30$)

Margin of Error for Mean Value of Statistical Population where σ^2 is not known and $n \leq 30$

$$E = t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$$

حيث S الانحراف المعياري للعينة

فترة الثقة للمتوسط الحسابي μ للمجتمع الإحصائي (في حالة σ^2 غير معلوم، $n \leq 30$)

Confidence Interval for Mean Value of Statistical Population where σ^2 is not known and and $n \leq 30$

$$(\overline{x}-E\ ,\ \overline{x}+E)$$

الخطوات المتبعة لإيجاد فترة الثقة للمتوسط الحساب μ إذا كانت σ^2 غير معلومة، $n \leq 30$:

1. نوجد درجات الحرية $(n-1)$.
2. نوجد القيمة الحرجة $t_{\frac{\alpha}{2}}$ المناظرة لدرجة ثقة 95% من جدول توزيع t.
3. نوجد هامش الخطأ $E = t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$
4. نوجد فترة الثقة $(\overline{x}-E\ ,\ \overline{x}+E)$.

مثال (4)

أخذت عينة عشوائية من مجتمع طبيعي حجمها $n = 25$، فإذا كان الانحراف المعياري للعينة (S) يساوي 10 ومتوسطها الحسابي $(\overline{x})$ يساوي 15، استخدم مستوى ثقة 95% لإيجاد:

1 هامش الخطأ.

2 فترة الثقة للمتوسط الحسابي للمجتمع الإحصائي μ.

الحل:

1 $\because$ σ^2 غير معلوم، $n \leq 30$

$\therefore$ نستخدم توزيع t.

$$\because \quad n = 25$$

$\therefore$ درجات الحرية: $n - 1 = 25 - 1 = 24$

$\because$ مستوى الثقة: $1 - \alpha = 95\%$

$$\therefore \quad 1 - \alpha = 0.95 \Longrightarrow \alpha = 0.050$$

$$\therefore \quad \frac{\alpha}{2} = 0.025$$

من جدول توزيع t تكون قيمة $t_{\frac{\alpha}{2}} = t_{0.025}$ مناظرة للعدد 2.064

هامش الخطأ $E = t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$

$$= 2.064 \times \frac{10}{\sqrt{25}}$$

$\therefore$ هامش الخطأ $= 4.128$

2 فترة الثقة: $(\overline{x} - E \ , \ \overline{x} + E)$

$$= (15 - 4.128 \ , \ 15 + 4.128)$$

$$= (10.872 \ , \ 19.128)$$

حاول أن تحل

4 أوجد فترة ثقة 95% للمتوسط الحسابي للمجتمع الإحصائي μ علمًا أن العينة أخذت من مجتمع طبيعي.

إذا كان لدينا $n = 13 \quad , \quad S = 0.3 \quad , \quad \overline{x} = 8.4$

4–2

اختبارات الفروض الإحصائية

Statistical Hypotheses Testing

سوف تتعلم

- القيمة الحرجة.
- مستوى المعنوية.
- درجة المعنوية.
- الفروض.
- اختبار الفروض.
- فرض العدم.
- الفرض البديل.

المفردات والمصطلحات:

- الفرض الإحصائي Statistic Hypothesis
- المقياس الإحصائي Statistical Scale
- اختبارات الفروض الإحصائية Statistical Hypotheses Testing
- فرض العدم Null Hypothesis
- الفرض البديل Alternative Hypothesis

دعنا نفكر ونتناقش

ينتج مصنع نوعًا معينًا من المعلبات مسجّل على العلبة أن الوزن الصافي 200 g.

فإذا تمّ أخذ عينة حجمها 100 علبة وتمّ حساب المتوسط الحسابي لأوزان هذه العينة فوجد أنه 197.3 g، فهل يمكن الحكم على هذا المصنع بأنه يقوم بغش تجاري؟

ما هي حيثيات هذا الحكم؟

نحن نعلم أنه في كثير من الأحيان وفي مواقف معينة نحتاج إلى اتخاذ قرار بناء على معلومات محددة وحيثيات معقولة لها مبررها، لذلك دعت الضرورة إلى دراسة ما يسمى بالفرض الإحصائي واختبارات الفروض الإحصائية.

تعريف: الفرض الإحصائي Statistic Hypothesis

هو ادعاء معيّن مبني على حيثيات معقولة حول معلمة من معالم المجتمع مثل المتوسط الحسابي μ أو الانحراف المعياري σ.

تعريف: المقياس الإحصائي

هو قيمة وحيدة محسوبة من العينة تحت شروط معينة.

تعريف: اختبارات الفروض الإحصائية (اختبار المعنوية)

هي طريقة معيارية لاختبار ادعاء ما حول معلمة من معالم المجتمع.

ملاحظة: سنكتفي في هذا الموضوع بدراسة معلمة واحدة من معالم المجتمع وهي المتوسط الحسابي μ.

إليك بعض الأمثلة عن الفروض التي يمكن اختبارها من خلال الطرق التي سنطوّرها في هذا الدرس. على سبيل المثال:

- **في إدارة الأعمال:** تدّعي إحدى الصحف في مقال لها أنّ معظم الموظّفين يجدون عملًا عن طريق وكالات التوظيف.
- **في الطب:** يدّعي باحثون في الطبّ أنّ متوسّط درجة حرارة جسم أي بالغ معافى ليست 37°C.

■ **في سلامة الطيران المدنيّ:** تدّعي إدارة الطيران المدني في الكويت أن متوسط وزن المسافر (مع حقائبه) يتعدّى الوزن المسموح منذ عشرين سنة والبالغ 84 kg

فرض العدم والفرض البديل — Null and Alternative Hypothesis

- فرض العدم (H_0): يفيد بأنّ قيمة معلمة المجتمع (مثل المتوسّط الحسابي μ) تساوي قيمة مزعومة.

نختبر فرض العدم مباشرة أي نفترض بأنّه صحيح ونتوصّل إلى خلاصة برفض أو عدم رفض H_0.

- الفرض البديل (H_1): يفيد بأن للمعلمة قيمة تختلف نوعًا ما عن فرض العدم (H_0).
- يضم الشكل الرمزي للفرض البديل أحد هذه الرموز: $<$ أو $>$ أو $\neq$

وستقتصر دراستنا على الحالة $(\neq)$. فمثلًا: $H_0: \mu = 98.6 \quad , \quad H_1: \mu \neq 98.6$

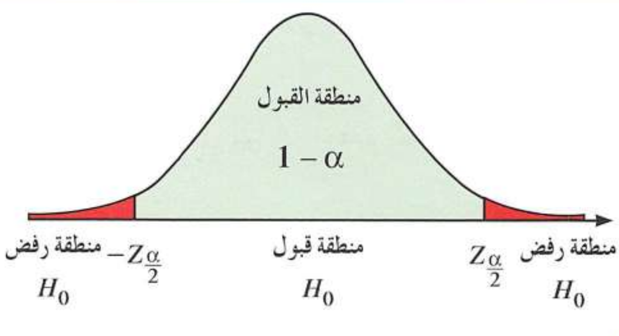
الخطوات المتبعة لإجراء اختبار الفروض الإحصائية:

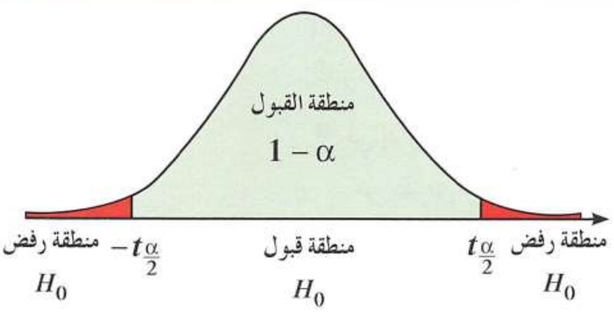
1. صياغة الفروض الإحصائية (فرض العدم H_0 والفرض البديل H_1).
2. التحقق من الانحراف المعياري σ للمجتمع (معلوم أم غير معلوم) وتحديد حجم العينة (n) ومن ثم إيجاد المقياس الإحصائي للاختبار (Z أو t)، (مسترشدًا بالجدول التالي):

الانحراف المعياري (σ)	المقياس الإحصائي (Z أو t)	حجم العينة (n)
معلوم	$Z = \dfrac{\overline{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$	لا يشترط حجم معين للعينة
غير معلوم	$Z = \dfrac{\overline{x} - \mu}{\dfrac{S}{\sqrt{n}}}$	$n > 30$
	$t = \dfrac{\overline{x} - \mu}{\dfrac{S}{\sqrt{n}}}$	$n \leq 30$

3. تحديد مستوى المعنوية α وحساب القيمة الجدولية $Z_{\frac{\alpha}{2}}$ من جدول التوزيع الطبيعي المعياري أو القيمة الجدولية $t_{\frac{\alpha}{2}}$ من جدول t ذي درجات حرية.
4. تحديد منطقة القبول: $(-Z_{\frac{\alpha}{2}} , Z_{\frac{\alpha}{2}})$ أو $(-t_{\frac{\alpha}{2}} , t_{\frac{\alpha}{2}})$ كما هو موضّح بالشكل.
5. اتخاذ القرار الإحصائي (قبول فرض العدم) أو (رفض فرض العدم وقبول الفرض البديل).

ملاحظة: ستقتصر دراستنا على مستوى ثقة 95%.





إذا كان الانحراف المعياري σ لمجتمع معلوم

مثال (1)

تزعم شركة أن متوسط رواتب موظفيها يساوي 4 000 دينار كويتي. إذا أخذت عينة من 25 موظفًا، ووجد أن متوسط رواتب العينة هو 3 950 دينارًا كويتيًا فإذا علمت أن الانحراف المعياري للمجتمع (دينارًا) $\sigma = 125$

وضّح كيفية إجراء الاختبار الإحصائي بمستوى ثقة 95%

الحل:

1 صياغة الفروض

$H_0 : \mu = 4\,000$ مقابل $H_1 : \mu \neq 4\,000$

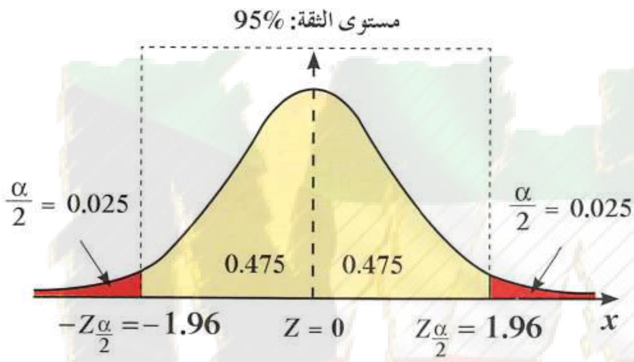
2 $\because$ $\sigma = 125$ (معلومة)

$\therefore$ نستخدم المقياس الإحصائي Z $\quad Z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

$\because \; n = 25 \quad , \quad \bar{x} = 3\,950$

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

$$\therefore \; Z = \frac{3\,950 - 4\,000}{\frac{125}{\sqrt{25}}} = -2$$



3 $\because$ مستوى الثقة 95%

$$\therefore \; \alpha = 0.05 \longrightarrow \frac{\alpha}{2} = 0.025$$

$$\therefore \; Z_{\frac{\alpha}{2}} = 1.96$$

4 منطقة القبول هي $(-1.96 \; , \; 1.96)$

5 اتخاذ القرار الإحصائي: $\because \; -2 \notin (-1.96 \; , \; 1.96)$

$\therefore$ القرار: نرفض فرض العدم $\mu = 4\,000$ ونقبل الفرض البديل $\mu \neq 4\,000$

حاول أن تحل

1 بيّنت الدراسة أن المتوسط الحسابي لقوة تحمل أسلاك معدنية هو $\mu = 1\,800$ kg مع انحراف معياري $\sigma = 150$ kg

ويؤكد الأخصائيون في المصنع المنتج لهذه الاسلاك أن بإمكانهم زيادة قوة تحمل هذه الأسلاك، وتأكيدًا على ذلك تمّ اختبار عينة من 40 سلكًا فتبيّن أن متوسط قوة تحمل هذه الأسلاك يساوي 1 840 kg

هل يمكن قبول مثل هذا الفرض بمستوى معنوية $\alpha = 0.05$؟

إذا كان الانحراف المعياري σ لمجتمع غير معلوم، $n > 30$

مثال (2)

إذا كانت $S = 1.79$, $\bar{x} = 37.2$, $n = 80$

اختبر الفرض بأن $\mu = 37$ عند مستوى معنوية $\alpha = 0.05$

الحل:

$$S = 1.79 \quad , \quad \bar{x} = 37.2 \quad , \quad n = 80$$

1 صياغة الفروض:

$H_0 : \mu = 37$ مقابل $H_1 : \mu \neq 37$

2 $\because$ σ غير معلومة ، $n > 30$

$\therefore$ نستخدم المقياس الإحصائي Z:

$$Z = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

$$Z = \frac{37.2 - 37}{\frac{1.79}{\sqrt{80}}} = 0.999$$

3 تحديد مستوى المعنوية α:

$$\because \ \alpha = 0.05 \longrightarrow \frac{\alpha}{2} = 0.025$$

$$\therefore \ Z_{0.025} = 1.96$$

4 منطقة القبول هي $(-1.96 \ , \ 1.96)$

5 اتخاذ القرار الإحصائي:

$$\because \ 0.999 \in (-1.96 \ , \ 1.96)$$

$\therefore$ القرار بقبول فرض العدم $\mu = 37$

حاول أن تحل

2 متوسط العمر بالساعات لعينة من 100 مصباح كهربائي مصنعة في أحد المصانع $\bar{x} = 1\,570$ بانحراف معياري $S = 120$.

يقول صاحب المصنع إنّ متوسط العمر بالساعات $\mu = 1\,600$ للمصابيح المصنعة في المصنع.

اختبر صحة الفرض $\mu = 1\,600$ مقابل الفرض $\mu \neq 1\,600$ وباختيار مستوى معنوية $\alpha = 0.05$

إذا كان الانحراف المعياري σ لمجتمع غير معلوم، $n \leq 30$

مثال (3)

يعتقد مدير شركة دراسات إحصائية أن متوسط الإنفاق الشهري على الطعام في منازل مدينة معينة يساوي 290 دينارًا كويتيًا.
فإذا أخذت عينة عشوائية من 10 منازل تبيّن أن متوسطها الحسابي (دينارًا) $\overline{x} = 283$ وانحرافها المعياري (دينارًا) $S = 32$.
فهل يمكن الاعتماد على هذه العينة لتأكيد ما افترضه؟
استخدم مستوى ثقة 95% (علمًا بأن المجتمع يتبع توزيعًا طبيعيًا).

الحل:

$n = 10 \quad , \quad \overline{x} = 283 \quad , \quad S = 32$

1 صياغة الفروض

$H_0 : \mu = 290$ مقابل $H_1 : \mu \neq 290$

2 σ غير معلومة ، $n = 10 \quad , \quad 10 < 30$

$\therefore$ نستخدم المقياس الإحصائي t :

$$t = \frac{\overline{x} - \mu}{\frac{S}{\sqrt{n}}}$$

$$t = \frac{283 - 290}{\frac{32}{\sqrt{10}}}$$

$$t \approx -0.6917$$

3 $\because \; n = 10$ درجات الحرية: $n - 1 = 10 - 1 = 9$

مستوى الثقة 95% $\quad 1 - \alpha = 0.95$

$$\alpha = 0.05 \longrightarrow \frac{\alpha}{2} = 0.025$$

من جدول توزيع t $\quad \therefore \; t_{0.025} = 2.262$

4 منطقة القبول هي $(-2.262 \; , \; 2.262)$

5 اتخاذ القرار الإحصائي: $\because \; -0.6917 \in (-2.262 \; , \; 2.262)$

$\therefore$ القرار بقبول فرض العدم $\mu = 290$

حاول أن تحل

3 في المثال (3)، إذا أجريت دراسة إحصائية أخرى على المدينة ذاتها وتبيّن من خلالها أن $S = 5 \; , \; \overline{x} = 296$ لعينة من 10 منازل مع استخدام درجة الثقة نفسها.

فهل يبقى افتراض المدير عند الشركة صحيحًا أم لا؟ وضّح إجابتك.

الارتباط والانحدار

Correlation and Regression

سوف تتعلم
- مفهوم الارتباط وأنواعه.
- رسم مخطط الانتشار.
- إيجاد مُعامل الارتباط الخطي.
- خواص مُعامل الارتباط.
- إيجاد مُعامل ارتباط بيرسون.
- مفهوم الانحدار.
- إيجاد معادلة خط الانحدار.
- تنبؤ قيمة أحد المتغيرين.
- التقدير باستخدام معادلة خط الانحدار.
- التنبؤ.
- إيجاد مقدار الخطأ.

المفردات والمصطلحات:
- الارتباط Correlation
- ارتباط طردي Positive Correlation
- ارتباط عكسي Negative Correlation
- مُعامل الارتباط الخطي Linear Correlation Coefficient
- الانحدار Regression
- معادلة خط الانحدار Regression Line Equation
- التنبؤ Prediction
- مقدار الخطأ Error Value

دعنا نفكر ونتناقش

هل تساءلت يومًا: كيف تحسب العلاقة بين الطول والوزن؟
ما الذي يربط بين التدخين والإصابة بمرض السرطان؟
كيف نجد رابطًا بين وزن سيارة واستهلاكها للوقود؟
كيف يتغيّر سعر الذهب مع تغير قيمة الدولار الأمريكي؟
وما هي أفضل وسيلة للتقدير لنقترب من الحقيقة؟

أوّلًا: الارتباط Correlation

من دراستنا السابقة تمّ عرض بعض المقاييس الإحصائية مثل مقاييس النزعة المركزية (المتوسط الحسابي – الوسيط – المنوال) ومقاييس التشتت (المدى – التباين – الانحراف المعياري).
نلاحظ أن هذه المقاييس كانت تصف شكل البيانات التي تمّ جمعها من ظاهرة إحصائية واحدة أي من متغير واحد والذي يمكن الحصول عليه من العينة. بينما يقابلنا في حياتنا العملية مواقف كثيرة تتضمن متغيرين (ظاهرتين) أو أكثر ويكون تساؤلنا: هل هناك علاقة بين هذه المتغيرات؟ وما هو شكل هذه العلاقة؟ وأيضًا كيف يمكن التنبؤ بقيمة أحد هذين المتغيرين إذا علم قيمة المتغير الآخر؟ وكثيرًا ما يرى الباحثون ضرورة دراسة العلاقة بين متغيرين (ظاهرتين) كما يتضح من الأمثلة التالية:

- الطول والوزن.
- التدخين والإصابة بمرض السرطان.
- وزن سيارة واستهلاكها للوقود.
- الإنفاق والدخل.
- سعر السلعة والكمية المعروضة منها.
- العمر وضغط الدم.

والأمثلة في هذا المجال كثيرة ومتعددة ولدراسة العلاقة بين هذه الظواهر ندرس ما يسمى الارتباط.

تعريف

الارتباط هو العلاقة بين متغيرين.

سنرمز للمتغير الأول بالرمز « x» وهو المتغير الذي يتم تحديده من قبل الباحث القائم بالدراسة ويسمى «**بالمتغير المستقل**».

ونرمز للمتغير الثاني بالرمز « y » وهذا المتغير غير مستقل بذاته لأن نتيجته مرتبطة بالمتغير المستقل ولذلك يسمى «**بالمتغير التابع**».

تذكر:
مخطط الانتشار هو تمثيل بياني لعدد من الأزواج المرتبة (x , y) يستخدم لوصف العلاقة بين المتغيرين.

أنواع الارتباط

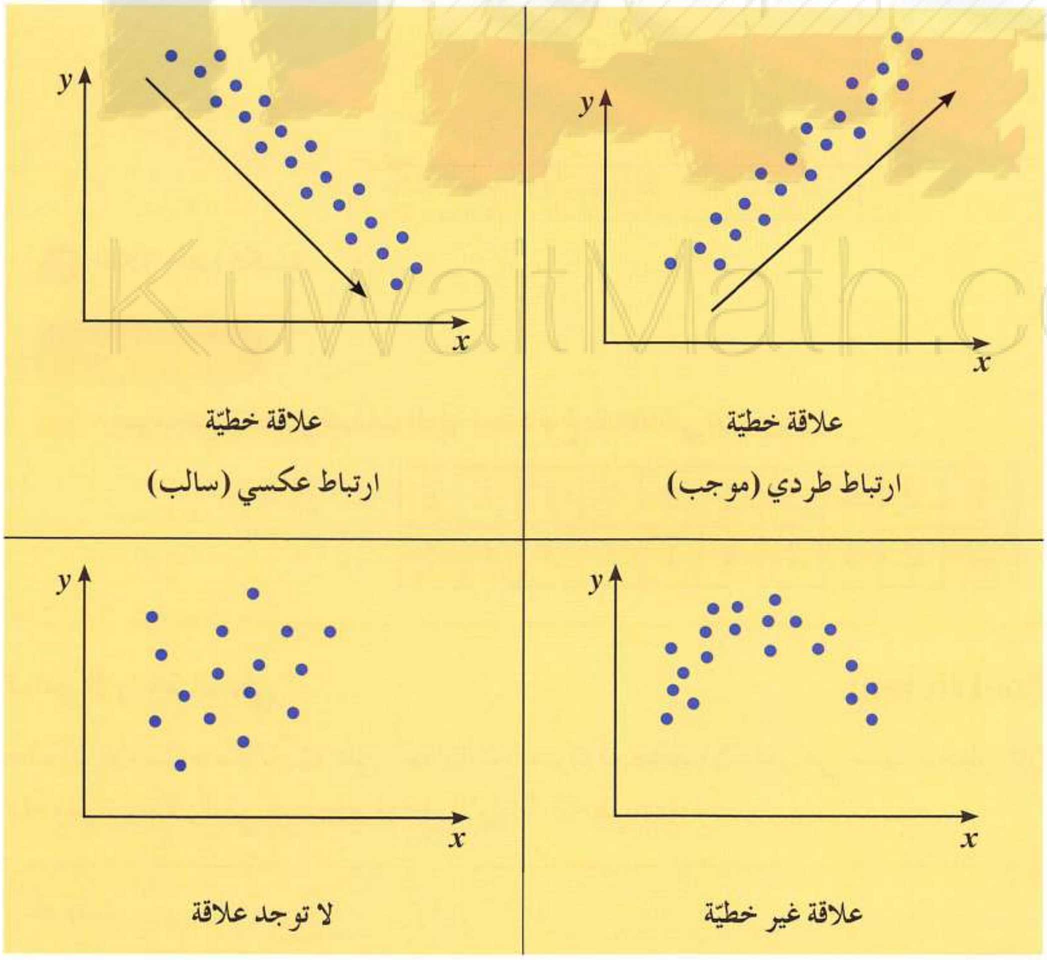
1 ارتباط طردي (موجب)

هو علاقة بين متغيرين y , x بحيث إذا تغير المتغير المستقل (x) فإن المتغير التابع (y) يتبعه في نفس الاتجاه.

2 ارتباط عكسي (سالب)

هو علاقة بين متغيرين y , x بحيث إذا تغير المتغير المستقل (x) فإن المتغير التابع (y) يتبعه في الاتجاه المضاد.

بعض مخططات الانتشار التي توضّح أنواع الارتباط



علاقة خطيّة
ارتباط طردي (موجب)

علاقة خطيّة
ارتباط عكسي (سالب)

علاقة غير خطيّة

لا توجد علاقة

مثال (1)

البيانات التالية تبيّن العلاقة بين عمر الشخص وعدد ساعات التمرينات الرياضية التي يقوم بها:

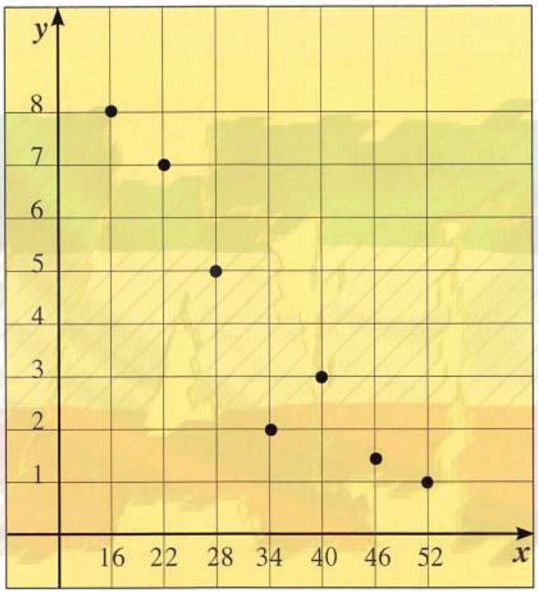
العمر (x)	16	22	28	34	40	46	52
عدد ساعات التمرينات (y)	8	7	5	2	3	$1\frac{1}{2}$	1

a ارسم مخطط الانتشار.

b حدّد نوع العلاقة.

الحل:

a



b العلاقة خطيّة عكسية.

حاول أن تحل

1 ارسم مخطط الانتشار للبيانات التالية وحدّد نوع العلاقة التي تعبر عنها:

x	2	6	5	2	7	3	4	7	5
y	2	3	1	4	1	5	3	4	5

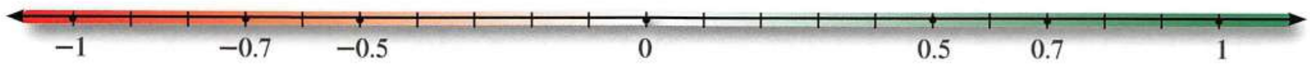
مُعامل الارتباط الخطي Linear Correlation Coefficient

تعلم أن الاستنتاجات المبنيّة على المعاينات البصريّة لمخطط الانتشار هي نسبيّة بامتياز، لذا فنحن بحاجة إلى قياسات أكثر دقة وموضوعية بالتالي نستخدم مُعامل الارتباط الخطّي (r).

تعريف

مُعامل الارتباط الخطّي (r) هو عبارة عن مقياس عددي لقوّة العلاقة بين متغيّرين يمثّلان بيانات كمّية، حيث $-1 \leq r \leq 1$

خواص مُعامل الارتباط (r)



1. $-1 \le r \le 1 \quad , \quad r \in [-1,1]$
2. إذا كانت $r = 1$ يكون الارتباط **طردي (موجب) تام.**
3. إذا كانت $r = -1$ يكون الارتباط **عكسي (سالب) تام.**
4. إذا كانت $r = 0$ ينعدم الارتباط.
5. إذا كانت $r \in [0.7,1)$ يكون الارتباط **طردي (موجب) قوي.**
6. إذا كانت $r \in [0.5,0.7)$ يكون الارتباط **طردي (موجب) متوسط.**
7. إذا كانت $r \in (0,0.5)$ يكون الارتباط **طردي (موجب) ضعيف .**
8. إذا كانت $r \in (-0.5,0)$ يكون الارتباط **عكسي (سالب) ضعيف.**
9. إذا كانت $r \in [-0.5,-0.7)$ يكون الارتباط **عكسي (سالب) متوسط.**
10. إذا كانت $r \in [-0.7,-1)$ يكون الارتباط **عكسي (سالب) قوي.**

مُعامل ارتباط بيرسون r: **Pearson Correlation Coefficient r:**

$$r = \frac{\Sigma(x-\overline{x})(y-\overline{y})}{nS_x \cdot S_y}$$

حيث: (الانحراف المعياري للمتغير x) ...... $= S_x = \sqrt{\frac{\Sigma(x-\overline{x})^2}{n}}$

(الانحراف المعياري للمتغير y) ...... $= S_y = \sqrt{\frac{\Sigma(y-\overline{y})^2}{n}}$

$$r = \frac{\Sigma(x-\overline{x})(y-\overline{y})}{\sqrt{\Sigma(x-\overline{x})^2}\sqrt{\Sigma(y-\overline{y})^2}}$$

مثال (2)

احسب مُعامل الارتباط الخطي للبيانات التالية وحدّد نوعه وقوة الارتباط.

x	1	2	3	4	5
y	3	5	7	9	11

الحل:

مُعامل الارتباط:

$$r = \frac{\Sigma(x-\overline{x})(y-\overline{y})}{\sqrt{\Sigma(x-\overline{x})^2}\sqrt{\Sigma(y-\overline{y})^2}}$$

$$\overline{x} = \frac{\Sigma x}{n} = 3 \quad , \quad \overline{y} = \frac{\Sigma y}{n} = 7$$

	x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
	1	3	-2	-4	4	16	8
	2	5	-1	-2	1	4	2
	3	7	0	0	0	0	0
	4	9	1	2	1	4	2
	5	11	2	4	4	16	8
المجموع	$\Sigma x = 15$	$\Sigma y = 35$			$\Sigma(x-\bar{x})^2 = 10$	$\Sigma(y-\bar{y})^2 = 40$	$\Sigma(x-\bar{x})(y-\bar{y}) = 20$

$$\therefore\ r = \frac{20}{\sqrt{10} \times \sqrt{40}} = 1$$

نوع الارتباط: طردي (موجب) تام.

حاول أن تحل

2 احسب مُعامل الارتباط الخطي للبيانات التالية وحدّد نوعه وقوة الارتباط:

x	1	2	3	4	5
y	1	-1	-4	-6	-5

صيغة أخرى لمعامل ارتباط بيرسون

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\ \sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$$

مثال (3)

احسب مُعامل الارتباط الخطي للمتغيرين التاليين وبيّن نوعه وقوته.

x	1	2	3	4	5	6
y	4	7	8	3	5	5

الحل:

$n = 6$

	x	y	xy	x^2	y^2
	1	4	4	1	16
	2	7	14	4	49
	3	8	24	9	64
	4	3	12	16	9
	5	5	25	25	25
	6	5	30	36	25
المجموع	$\Sigma x = 21$	$\Sigma y = 32$	$\Sigma xy = 109$	$\Sigma x^2 = 91$	$\Sigma y^2 = 188$

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\ \sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$$

$$r = \frac{6 \times 109 - 21 \times 32}{\sqrt{6 \times 91 - (21)^2}\ \sqrt{6 \times 188 - (32)^2}}$$

$$r = \frac{-18}{\sqrt{105} \times \sqrt{104}}$$

$r \approx -0.172$ ارتباط عكسي (سالب) ضعيف

حاول أن تحل

3 احسب مُعامل الارتباط الخطي للبيانات التالية وبيّن نوعه وقوته:

x	1	2	3	4	5	6
y	98	99	75	40	100	150

ثانيًا: الانحدار Regression

سوف نتعلّم وصف العلاقة بين متغيّرين بإيجاد معادلة الخط المستقيم الممثّل لهذه العلاقة.

يسمّى هذا الخط المستقيم **بخط الانحدار**، وتسمّى معادلته **بمعادلة خط الانحدار**.

تعريف

الانحدار هو وصف العلاقة بين متغيرين.

تعريف

معادلة خط الانحدار: هي المعادلة الخطية التي يمكن من خلالها التنبؤ بقيم أحد المتغيرين إذا علمت قيمة المتغير الآخر.

سبق لنا دراسة معادلة الخط المستقيم على الصورة: $y = b_1 x + b_0$ حيث b_1 ترمز إلى ميل هذا المستقيم،

$|b_0|$ ترمز إلى طول الجزء المقطوع من محور الصادات.

في الإحصاء توجد طرق متعددة لإيجاد معادلة خط انحدار مستقيم والتي تساعدنا في التنبؤ بقيمة أحد المتغيرين ومنها الطريقة التالية:

تعريف

$\widehat{y} = b_0 + b_1 x$ ، حيث $|b_0|$ ترمز إلى طول الجزء المقطوع من محور الصادات، b_1 ترمز إلى ميل المستقيم.

حيث: $$b_0 = \overline{y} - b_1 \overline{x} \quad , \quad b_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

حيث: $$\overline{x} = \frac{\Sigma x}{n} \quad , \quad \overline{y} = \frac{\Sigma y}{n}$$

وهذا ما يسمى **بطريقة المربعات الصغرى** التي تتلخص خطواتها فيما يلي:

1. تعيين قيمة b_1
2. تعيين قيمة b_0
3. نكتب معادلة خط الانحدار: $\widehat{y} = b_0 + b_1 x$
4. التنبؤ بقيمة y إذا علمت قيمة x
5. تحديد مقدار الخطأ في التنبؤ

مقدار الخطأ = |القيمة الجدولية − القيمة التي تحقق معادلة الانحدار|

مقدار الخطأ = $|y_x - \widehat{y}_x|$

مثال (4)

باستخدام البيانات التالية لقيم y , x:

x	1	3	5	7	9
y	2	5	9	10	14

أوجد:

a) **معادلة خط الانحدار.**

b) **قيمة y عندما $x = 10$**

c) **مقدار الخطأ عندما $x = 5$**

الحل:

a)

	x	y	xy	x^2
	1	2	2	1
	3	5	15	9
	5	9	45	25
	7	10	70	49
	9	14	126	81
المجموع	$\Sigma x = 25$	$\Sigma y = 40$	$\Sigma xy = 258$	$\Sigma x^2 = 165$

$$n = 5 \ , \ \overline{x} = \frac{\Sigma x}{n} = \frac{25}{5} = 5 \ , \ \overline{y} = \frac{\Sigma y}{n} = \frac{40}{5} = 8$$

$$b_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} = \frac{5 \times 258 - 25 \times 40}{5 \times 165 - 25 \times 25} = 1.45$$

$$b_0 = \overline{y} - b_1 \overline{x} = 8 - 1.45 \times 5 = 0.75$$

∴ معادلة خط الانحدار هي: $\widehat{y} = b_0 + b_1 x = 0.75 + 1.45x$

b) **عندما $x = 10$ فإن:**

$$y = 0.75 + 1.45 \times 10 = 15.25$$

c من الجدول $y = 9$

من المعادلة:

$$\hat{y} = 0.75 + 1.45 \times 5 = 8$$

∴ مقدار الخطأ:

$$|y_5 - \hat{y}_5| = |9 - 8| = 1$$

حاول أن تحل

4 من الجدول التالي:

x	4	5	8	9	10	12
y	2	4	5	8	6	11

أوجد:

a معادلة خط الانحدار.

b قيمة y عندما $x = 10$

c مقدار الخطأ عندما $x = 10$

مثال (5)

سقطت كرة من ارتفاع 50 m، وتمّ تسجيل المسافات (بالمتر) التي قطعتها هذه الكرة كل 0.5 s لمدة ثلاث ثوان.

فأتت النتائج كما يوضح الجدول التالي:

الوقت (x)	0	0.5	1	1.5	2	2.5	3
المسافة (y)	0	1.2	4.9	11	19.5	30.5	44

a أوجد معادلة خط الانحدار.

b قدّر قيمة المسافة y عندما $x = 4$

c أوجد مقدار الخطأ في المسافة عندما $x = 2.5\ s$

الحل:

a

	x	y	xy	x^2
	0	0	0	0
	0.5	1.2	0.6	0.25
	1	4.9	4.9	1
	1.5	11	16.5	2.25
	2	19.5	39	4
	2.5	30.5	76.25	6.25
	3	44	132	9
المجموع	$\Sigma x = 10.5$	$\Sigma y = 111.1$	$\Sigma xy = 269.25$	$\Sigma x^2 = 22.75$

$$n = 7 \quad , \quad \bar{x} = \frac{\Sigma x}{n} = \frac{10.5}{7} = 1.5 \quad , \quad \bar{y} = \frac{\Sigma y}{n} = \frac{111.1}{7} = 15.87$$

$$b_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} = \frac{7 \times 269.25 - 10.5 \times 111.1}{7 \times 22.75 - (10.5)^2} = \frac{718.2}{49}$$

$$b_1 \approx 14.66$$

$b_0 = \overline{y} - b_1\overline{x}$

$= 15.87 - 14.66 \times 1.5$

$= - 6.12$

معادلة خط الانحدار هي: $\widehat{y} = b_0 + b_1 x = - 6.12 + 14.66x$

b التنبؤ: $\because \ \widehat{y} = - 6.12 + 14.66x$

$\therefore$ المسافة y عندما $x = 4$ هي:

$\widehat{y}_4 = - 6.12 + 14.66 \times 4 = 52.52 \text{ m}$

c مقدار الخطأ عند $x = 2.5 \text{ s}$

من المعادلة: $\widehat{y} = - 6.12 + 14.66x$

نجد أن: $\widehat{y}_{2.5} = - 6.12 + 14.66 \times 2.5 = 30.53$

من الجدول عند $x = 2.5 \text{ s}$

نجد أن: $y = 30.5 \text{ s}$

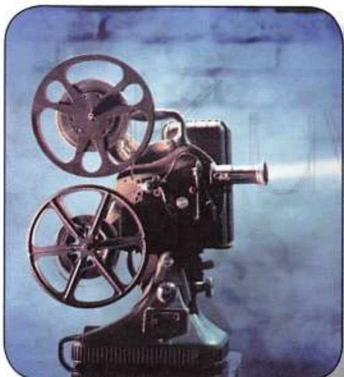
$\therefore$ مقدار الخطأ: $| y_x - \widehat{y}_x |$

$= | 30.5 - 30.53 |$

$= 0.03$

حاول أن تحل

5 في الجدول التالي، المتغير x هو تكلفة إنتاج فيلم سينمائي (بملايين الدولارات) والمتغيّر y هو مردود هذا الفيلم.



التكلفة (x)	62	90	50	35	200	100	95
المردود (y)	65	64	48	57	601	146	47

a أوجد معادلة خط الانحدار.

b قدّر مردود فيلم بلغت تكلفته 55 مليون دولار.

c أوجد مقدار الخطأ لفيلم بلغت تكلفته 90 مليون دولار.

المرشد لحل المسائل

نظرًا لأهميّة المياه بالنسبة إلى صحة الإنسان وحياته، قررت مؤسسة تعنى بذلك، القيام بحملة تهدف إلى التأكد من أن كل شخص يستهلك متوسط قدره 2 000 ml يوميًّا من مياه الشرب.

في دراسة سابقة لعينة من 100 شخص، لاحظت المؤسسة أنّ المتوسط الحسابي للاستهلاك: $\overline{x} = 1850$ ml مع انحراف معياري $S = 900$ ml.

وفي دراسة جديدة لعينة من 100 شخص، وبعد القيام بحملتها، لاحظت أن المتوسط الحسابي للاستهلاك: $\overline{x} = 1900$ ml مع انحراف معياري $S = 300$ ml.

اعتقدت المؤسسة أنّ حملتها قد نجحت بما أنّ المتوسط الحسابي للاستهلاك قد ازداد 50 ml وقد اقترب كثيرًا من هدفها وهو 2 000 ml يوميًّا للشخص الواحد.

هل المؤسسة على حق؟ اشرح.

الحل:

وضع يوسف جدولًا ليختبر فرضية الشركة من خلال اختبارات إحصائية مع:

$H_0 : \mu = 2\,000$ مقابل $H_1 : \mu \neq 2\,000$ ، ومستوى الثقة 0.95

	الدراسة السابقة	الدراسة الجديدة
المعايير	$\overline{x} = 1850 \ , \ S = 900 \ , \ n = 100$	$\overline{x} = 1900 \ , \ S = 300 \ , \ n = 100$
القيمة الجدولية	$Z_{\frac{\alpha}{2}} = 1.96$	$Z_{\frac{\alpha}{2}} = 1.96$
قيمة الاختبار الإحصائي	$Z = \dfrac{\overline{x} - \mu}{\frac{S}{\sqrt{n}}} = \dfrac{1850 - 2\,000}{\frac{900}{\sqrt{100}}} = -1.66$	$Z = \dfrac{1900 - 2\,000}{\frac{300}{\sqrt{100}}} = -3.33$
الفترة	$(-1.96 \ , \ 1.96)$	$(-1.96 \ , \ 1.96)$
القرار	$\because \ -1.66 \in (-1.96 \ , \ 1.96)$ قبول $H_0 : \mu = 2\,000$ ml	$\because \ -3.33 \notin (-1.96 \ , \ 1.96)$ رفض H_0 والأخذ بـ: $H_1 : \mu \neq 2\,000$ ml

الاستنتاج:

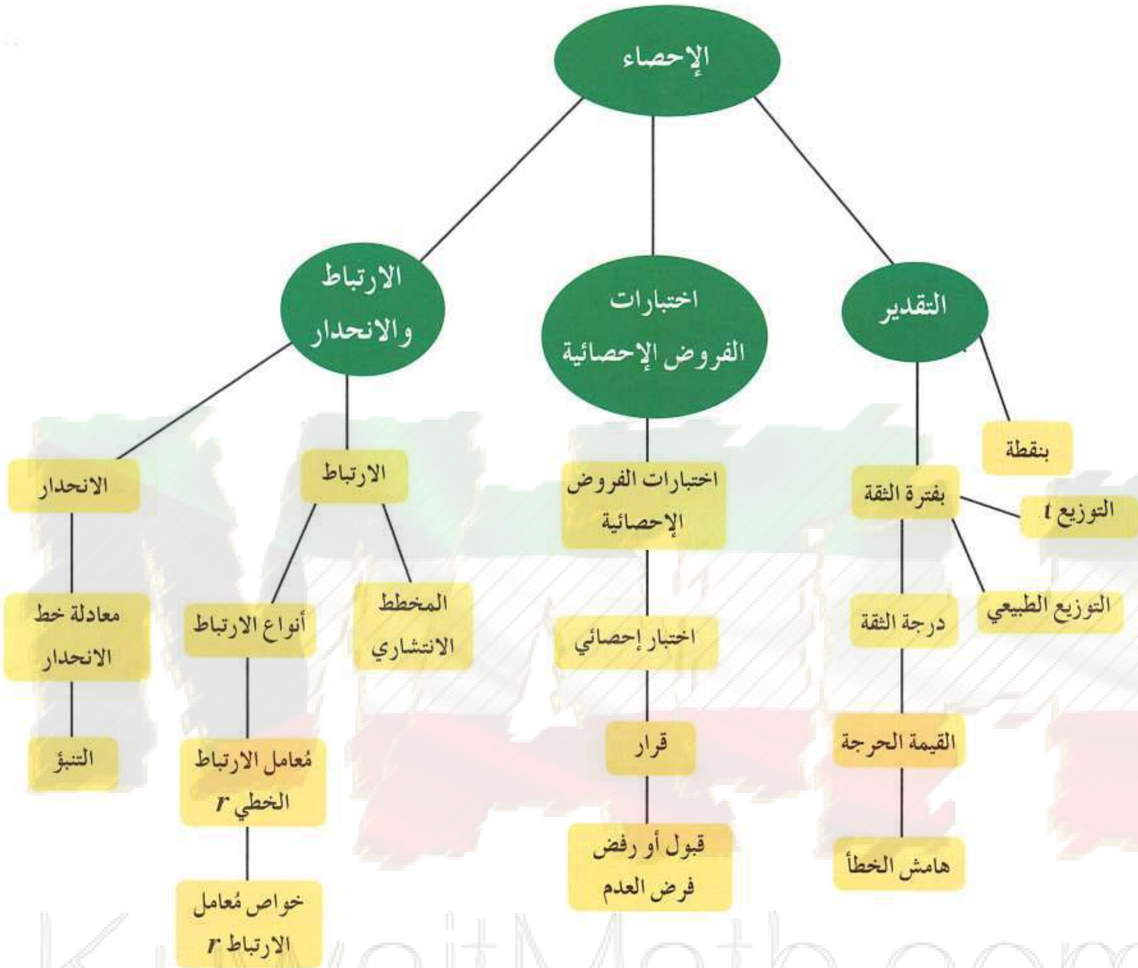
لم تكن الحملة ضرورية، والحصول على قيمة متوسطة أكبر لا يعني الاقتراب من الهدف المنشود.

مسألة إضافية

قامت مؤسسة أخرى بحملة على عينة من 100 شخص تهدف إلى التأكد من أنّ المتوسط الحسابي لاستهلاك كل شخص لمياه الشرب $\mu = 2\,000$ ml يوميًّا. فأتت النتائج على الشكل التالي:

$\overline{x} = 2\,100$ ml , $S = 800$ ml. برأيك، هل كانت حملة هذه المؤسسة ناجحة؟

مخطط تنظيمي للوحدة الرابعة



ملخص

- المعلمة هي ثابت يصف المجتمع أو يصف توزيع المجتمع كالمتوسط الحسابي μ أو الانحراف المعياري σ.
- الإحصاءة هو اقتران تتعيّن قيمته من العينة كالمتوسط الحسابي $\overline{x}$ أو الانحراف المعياري لها S.
- تقدير المعلمة: هو إحصاءة تعتمد على قيم العينة وتعكس قيمة قريبة لمعلمة المجتمع ككل وتوزيعه.
- التقدير بنقطة هي قيمة وحيدة محسوبة من العينة تستخدم لتقدير معلمة مجهولة من معالم المجتمع.
- فترة الثقة هي فترة طرفاها متغيران عشوائيان (أي أنها فترة عشوائية) تحوي إحدى معالم المجتمع بنسبة معينة تسمى درجة الثقة (مستوى الثقة).
- التقدير بفترة الثقة: هو إيجاد فترة معينة يتوقع أن تقع معلمة المجتمع داخلها بنسبة معينة أو احتمال معين.
- α هي نسبة الخطأ في التقدير وتسمى مستوى المعنوية أو مستوى الدلالة.
- $(1-\alpha)$ هي درجة الثقة (مستوى الثقة).
- $Z_{\frac{\alpha}{2}}$ هي القيمة الحرجة المستخرجة من جدول التوزيع الطبيعي المعياري.
- $\overline{x}$ هو المتوسط الحسابي للعينة.

- S هو الانحراف المعياري للعينة.
- $t_{\frac{\alpha}{2}}$ هي القيمة الحرجة المستخرجة من جدول التوزيع t.
- هامش الخطأ $E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ في حالة الانحراف المعياري σ معلوم والتوزيع الطبيعي.
- فترة الثقة هي: $(\overline{x} - E \ , \ \overline{x} + E)$.
- الفرض الإحصائي: هو ادعاء معين مبني على حيثيات معقولة حول معلمة من معالم المجتمع مثل المتوسط الحسابي μ أو الانحراف المعياري σ.
- المقياس الإحصائي هو قيمة وحيدة محسوبة من العينة تحت شروط معينة.
- اختبارات الفروض الإحصائية (اختبار المعنوية) هي طريقة معيارية لاختبار ادعاء ما حول معلمة من معالم المجتمع.
- الارتباط هو العلاقة بين متغيرين.
- ارتباط طردي (موجب): هو علاقة بين متغيرين y , x بحيث إذا تغير المتغير المستقل (x) فإن المتغير التابع (y) يتبعه في نفس الاتجاه.
- ارتباط عكسي (سالب): هو علاقة بين متغيرين y , x بحيث إذا تغير المتغير المستقل (x) فإن المتغير التابع (y) يتبعه في الاتجاه المضاد.
- مُعامل الارتباط الخطي (r) هو عبارة عن مقياس عددي لقوّة العلاقة بين متغيرين يمثلان بيانات كمية حيث $-1 \leq r \leq 1$

خواص مُعامل الارتباط (r)

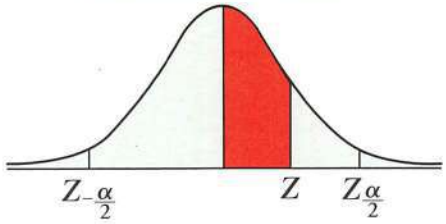
1. إذا كانت $r = 1$ يكون الارتباط **طردي (موجب) تام.**
2. إذا كانت $r = -1$ يكون الارتباط **عكسي (سالب) تام.**
3. إذا كانت $r = 0$ ينعدم الارتباط.
4. إذا كانت $r \in [0.7 , 1)$ يكون الارتباط **طردي (موجب) قوي.**
5. إذا كانت $r \in [0.5 , 0.7)$ يكون الارتباط **طردي (موجب) متوسط.**
6. إذا كانت $r \in (0 , 0.5)$ يكون الارتباط **طردي (موجب) ضعيف.**
7. إذا كانت $r \in (-0.5 , 0)$ يكون الارتباط **عكسي (سالب) ضعيف.**
8. إذا كانت $r \in [-0.5 , -0.7)$ يكون الارتباط **عكسي (سالب) متوسط.**
9. إذا كانت $r \in [-0.7 , -1)$ يكون الارتباط **عكسي (سالب) قوي.**

- $r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\ \sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$ أو $r = \frac{\Sigma(x - \overline{x})(y - \overline{y})}{\sqrt{\Sigma(x - \overline{x})^2}\ \sqrt{\Sigma(y - \overline{y})^2}}$
- الانحدار هو وصف العلاقة بين متغيرين.
- معادلة خط الانحدار: هي المعادلة الخطية التي يمكن من خلالها التنبؤ بقيمة أحد المتغيرين إذا علمت قيمة المتغير الآخر.

$$b_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} \quad \text{حيث} \quad \widehat{y} = b_0 + b_1 x$$

$$b_0 = \overline{y} - b_1 \overline{x} \quad \text{حيث} \quad \overline{y} = \frac{\Sigma y}{n} \quad \text{أو} \quad \overline{x} = \frac{\Sigma x}{n}$$

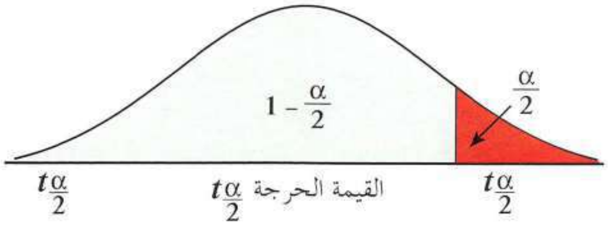
- مقدار الخطأ = | القيمة الجدولية − القيمة من معادلة الانحدار | = $|y_x - \widehat{y}_x|$



جدول التوزيع الطبيعي المعياري (Z)

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.10 وأكثر	0.4999									

ملاحظة: استخدم 0.4999 عندما تزيد قيمة Z عن 3.09



جدول التوزيع t

	$\frac{\alpha}{2}$					
درجات الحرية $(n-1)$	0.005	0.01	0.025	0.05	0.10	0.25
1	63.657	31.821	12.706	6.314	3.078	1.000
2	9.925	6.965	4.303	2.920	1.886	0.816
3	5.841	4.541	3.182	2.353	1.638	0.765
4	4.604	3.747	2.776	2.132	1.533	0.741
5	4.032	3.365	2.571	2.015	1.476	0.727
6	3.707	3.143	2.447	1.943	1.440	0.718
7	3.500	2.998	2.365	1.895	1.415	0.711
8	3.355	2.896	2.306	1.860	1.397	0.706
9	3.250	2.821	2.262	1.833	1.383	0.703
10	3.169	2.764	2.228	1.812	1.372	0.700
11	3.106	2.718	2.201	1.796	1.363	0.697
12	3.054	2.681	2.179	1.782	1.356	0.696
13	3.012	2.650	2.160	1.771	1.350	0.694
14	2.977	2.625	2.145	1.761	1.345	0.692
15	2.947	2.602	2.132	1.753	1.341	0.691
16	2.921	2.584	2.120	1.746	1.337	0.690
17	2.898	2.567	2.110	1.740	1.333	0.689
18	2.878	2.552	2.101	1.734	1.330	0.688
19	2.861	2.540	2.093	1.729	1.328	0.688
20	2.845	2.528	2.086	1.725	1.325	0.687
21	2.831	2.518	2.080	1.721	1.323	0.686
22	2.819	2.508	2.074	1.717	1.321	0.686
23	2.807	2.500	2.069	1.714	1.320	0.685
24	2.797	2.492	2.064	1.711	1.318	0.685
25	2.787	2.485	2.060	1.708	1.316	0.684
26	2.779	2.479	2.056	1.706	1.315	0.684
27	2.771	2.473	2.052	1.703	1.314	0.684
28	2.763	2.467	2.048	1.701	1.313	0.683
29	2.756	2.462	2.045	1.699	1.311	0.683
30 وأكثر	2.575	2.327	1.960	1.645	1.282	0.675

ALGEBRA الجبر

$x^2 - y^2 = (x+y)(x-y)$

$x^3 + y^3 = (x+y)(x^2 - xy + y^2)$

$x^3 - y^3 = (x-y)(x^2 + xy + y^2)$

$(x+y)^2 = x^2 + 2xy + y^2$

$(x-y)^2 = x^2 - 2xy + y^2$

$(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3$

$(x-y)^3 = x^3 - 3x^2y + 3xy^2 - y^3$

$ax^2 + bx + c = 0 \quad , \quad x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

$a > 0$:

$|x| = a \ : \ x = a$ أو $x = -a$

$|x| < a \ : \ -a < x < a$

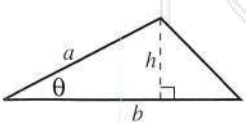
$|x| > a \ : \ x > a$ أو $x < -a$

GEOMETRY الهندسة

Triangle

$A = \dfrac{1}{2}bh$

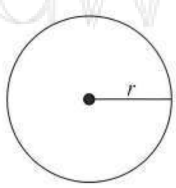
$\quad = \dfrac{1}{2}ab\sin\theta$



Circle

$A = \pi r^2$

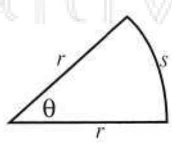
$C = 2\pi r$



Sector of Circle

$A = \dfrac{1}{2}r^2\theta$

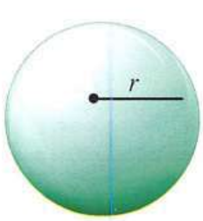
$s = r\theta$ (θ in radians)



Sphere

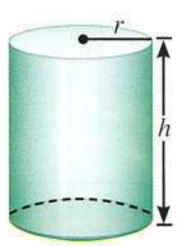
$V = \dfrac{4}{3}\pi r^3$

$A = 4\pi r^2$



Cylinder

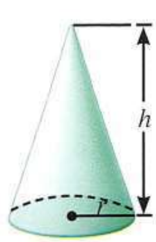
$V = \pi r^2 h$



Cone

$V = \dfrac{1}{3}\pi r^2 h$

$A = \pi r\sqrt{r^2 + h^2}$



TRIGONOMETRY علم المثلثات

$\csc\theta = \dfrac{1}{\sin\theta}$ $\qquad$ $\sec\theta = \dfrac{1}{\cos\theta}$

$\tan\theta = \dfrac{\sin\theta}{\cos\theta}$ $\qquad$ $\cot\theta = \dfrac{\cos\theta}{\sin\theta}$

$\cot\theta = \dfrac{1}{\tan\theta}$ $\qquad$ $\sin^2\theta + \cos^2\theta = 1$

$1 + \tan^2\theta = \sec^2\theta$ $\qquad$ $1 + \cot^2\theta = \csc^2\theta$

$\sin(-\theta) = -\sin\theta$ $\qquad$ $\cos(-\theta) = \cos\theta$

$\tan(-\theta) = -\tan\theta$ $\qquad$ $\sin\left(\dfrac{\pi}{2} - \theta\right) = \cos\theta$

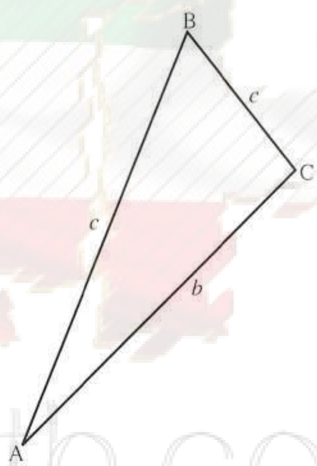
$\cos\left(\dfrac{\pi}{2} - \theta\right) = \sin\theta$ $\qquad$ $\tan\left(\dfrac{\pi}{2} - \theta\right) = \cot\theta$

$$\frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c}$$

$$a^2 = b^2 + c^2 - 2bc\cos A$$

$$b^2 = a^2 + c^2 - 2ac\cos B$$

$$c^2 = a^2 + b^2 - 2ab\cos C$$



$$\sin(x+y) = \sin x\cos y + \cos x\sin y$$

$$\sin(x-y) = \sin x\cos y - \cos x\sin y$$

$$\cos(x+y) = \cos x\cos y - \sin x\sin y$$

$$\cos(x-y) = \cos x\cos y + \sin x\sin y$$

$$\tan(x+y) = \frac{\tan x + \tan y}{1 - \tan x\tan y}$$

$$\tan(x-y) = \frac{\tan x - \tan y}{1 + \tan x\tan y}$$

$$\sin 2x = 2\sin x\cos x$$

$$\cos 2x = \cos^2 x - \sin^2 x = 2\cos^2 x - 1 = 1 - 2\sin^2 x$$

$$\tan 2x = \frac{2\tan x}{1 - \tan^2 x}$$

$\sin^2 x = \dfrac{1 - \cos 2x}{2}$ $\qquad$ $\cos^2 x = \dfrac{1 + \cos 2x}{2}$

الإحصاء Statistics

$Z_{\frac{\alpha}{2}} = Z_{\frac{1-\alpha}{2}}$; $-Z_{\frac{\alpha}{2}} = -Z_{\frac{1-\alpha}{2}}$ (القيمة الحرجة)

$\frac{\sigma}{\sqrt{n}}$ (الخطأ المعياري للمجتمع)

$E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ (هامش الخطأ – توزيع طبيعي)

$(\overline{x} - E\ ,\ \overline{x} + E)$ فترة الثقة للمتوسط الحسابي

$t_{\frac{\alpha}{2}} = t_{1-\frac{\alpha}{2}}$ (التوزيع t)

$E = t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$ (هامش الخطأ – توزيع t الانحراف المعياري σ غير معاوم)

$Z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ (المقياس الإحصائي – توزيع طبيعي)

$Z = \frac{\overline{x} - \mu}{\frac{S}{\sqrt{n}}}$ (المقياس الإحصائي – توزيع طبيعي – الانحراف المعياري σ غير معلوم)

$t = \frac{\overline{x} - \mu}{\frac{S}{\sqrt{n}}}$ (المقياس الاحصائي – توزيع t – الانحراف المعياري σ غير معاوم)

$r = \frac{\sum(x-\overline{x})(y-\overline{y})}{\sqrt{\sum(x-\overline{x})^2}\sqrt{\sum(x-\overline{y})^2}} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}}$ (مُعامل ارتباط بيرسون)

$\widehat{y} = b_1 x + b_0$ (معادلة خط الانحدار)

$b_1 = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$, $b_0 = \overline{y} - b_1\overline{x}$

حيث $\overline{x} = \frac{\sum x}{n}$, $\overline{y} = \frac{\sum y}{n}$

مقدار الخطأ $= |y_x - \widehat{y_x}|$