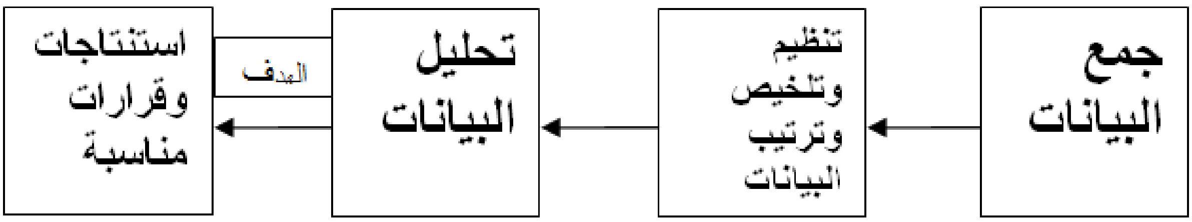
<u>الفصل الأول</u>

مدخل الى علم الإحصاء

<u>علم الإحصاء</u>: هو ذلك العلم الذي يبحث في جمع البيانات وتنظيمها وتلخيصها وعرضها ثم تحليلها، واستخلاص النتائج التي تساعد في التنبؤ والاستنتاج واتخاذ القرارات المناسبة، ويمكن توضيح مراحل العملية الإحصائية بالاتي:



<u>فروع علم الإحصاء</u>

الإحصاء الوصفي: الإحصاء الوصفي هو الفرع المتعلق بطرق جمع وتبويب وتنظيم وعرض وتلخيص المعلومات والبيانات ووصف توزيع البيانات وذلك باستخدام جداول تكرارية أو رسوم بيانية وكذلك إيجاد بعض المقاييس العددية أو الوصفية التي تصف توزيع البيانات.

الإحصاء الاستدلالي: الإحصاء الاستدلالي هو الفرع المتعلق بطرق اتخاذ القرارات حول المجتمع تحت الدراسة وذلك عن طريق دراسة العينة. وهذا يهدف إلى إيجاد تقديرات لمعالم مجهولة أو الإجابة عن بعض الأسئلة البحثية أو التحقق من بعض الفروض المسبقة حول هذه المعالم المجهولة.

<u>البيانات</u>: مجموعة القيم التي يتم جمعها من مفردات المجتمع أو العينة لخاصية معينة أو متغير معين.

بيانات نوعية (وصفية)

وصف الظواهر بشكل **<u>غير رقمي</u>** مثل الطول يمكن وصفه بالصفات طويل، قصير، أو مثل التحصيل الدراسى يمكن وصفه بأنه منخفض أو عالى كذلك الجنس ، المستوى التعليمي. ولا يمكن اجراء العمليات الحسابية عليها.

بيانات كمية

وصف الظواهر بشكل **<u>رقمي</u>** أو البيانات التي نحصل عليها على شكل أعداد مثل طول الطلاب يمكن أن يأخذ قيما، وكذلك الوزن، أو درجات تحصيل الطلاب ، مستوى السكر في الدم .

<u>أنواع البيانات الكمية</u>

البيانات الكمية المنفصلة

- القيم التي لا تحتوى على أجزاء (أعداد طبيعية) مثل عدد الطالبات في برنامج التمريض 70 طالبة، عدد السكان.
- البيانات التي يمكن عدها
- تكون منفصلة عن بعضها.
- مثل عدد غرف المنزل ، عدد أفراد الأسرة

البيانات الكمية المتصلة

- القيم التي تحتوى على أجزاء (أعداد عشرية) مثل طول الشخص 1.8 م، أو وزنه 34.5 كم، وغيرها من البيانات والقيم التي تأخذ أجزاء عشرية.
- البيانات التي لا يمكن عدها .

أنواع البيانات النوعية

الإسمي: أبسط أنواع المقاييس وهي تضع البيانات ضمن فئات محددة مثل الجنس : ذكر/أنثى ، اللون : أبيض/أسود ،الحالة الاجتماعية: متزوج/ غير متزوج.

الترتيبي: يمكننا من ترتيب البيانات تصاعدياً أو تنازلياً. مثال (متوسط- ممتاز - جيد جداً.....الخ) (الأول - الثاني - الثالث) .

المتغير الإحصائي: هو مقدار له خصائص رقمية (كمية) وغير رقمية (وصفية) تتغير قيمته من عنصر إلى آخر من عناصر المجتمع أو العينة.

المجتمع: هو مجموعة من المفردات التي تشترك في صفات وخصائص محددة .ومجتمع الد راسة هو الذي يشمل جميع مفردات الدراسة، مثل: مجتمع من الطلبة، مجتمع من الأسر، مجتمع من الفئران.

العينة: بأنها جزء من المجتمع تختار بحيث تمثل جميع صفات المجتمع .

الفصل الثاني

عرض البيانات الإحصائية

الهدف منها هو تصنيف و تفسير البيانات والمعطيات وعرضها بصورة يمكن الاستفادة منها في وصف الظاهرة محل الدر ا سة، وهناك طريقتين لعرض البيانات هما **العرض الجدولي** و**العرض البياني**.

العرض الجدولي

وتختلف الجداول الإحصائية باختلاف نوع البيانات من ناحية والغرض من الد ا رسة من ناحية أخرى لذلك يتم التمييز بين الحالات الموالية:

- **الجداول التكرارية البسيطة**

حيث يتم عرض البيانات من خلال تفريغها في جداول نهائية يحتوي كل منها على عمودين (سطرين). يبين العمود الأول قيم الظاهرة أو المتغير المدروس (x_i)، وتكون هذه القيم على شكل قيم نقطية أو شكل مجالات، أما العمود الثاني فيحتوي على تكرارات (n_i)هذه القيم أو المجالات.

مثال

x_i	n_i
0	4
1	7
2	11
3	8
4	5
5	3

الجدول : توزيع 38 أسرة حسب عدد الأطفال

التكرار النسبي والمئوي

التكرار المطلق

يمثل التكرار العادي للبيانات، و يرمز له بالرمز n_i.

التكرار النسبي

يستعمل للتعبير عن الأهمية النسبية لتكرار كل متغير بالنسبة لإجمالي التكرارات، يرمز له بالرمز f_i، وهو يحسب بالعلاقة التالية:

$$f_i = \frac{n_i}{n},$$

التكرار المتجمع الصاعد المطلق

يرمز له بالرمز $N_i^{\uparrow}$، ويتم الحصول عليه بالطريقة التالية:

$$N_1^{\uparrow} = n_1$$

$$N_2^{\uparrow} = n_1 + n_2$$

$$N_3^{\uparrow} = n_1 + n_2 + n_3$$
$$\vdots$$
$$N_i^{\uparrow} = n_1 + n_2 + \cdots + n_i$$

التكرار المتجمع الصاعد النسبي

يرمز له بالرمز $F_i^{\uparrow}$، ويتم ايجادة بنفس طريقة الحصول على التكرار المتجمع الصاعد المطلق لكن بإستخدام التكرار النسبي بدلا من التكرار المطلق.

مثال

$\boldsymbol{x_i}$	$\boldsymbol{n_i}$	f_i	$F_i^{\uparrow}$	$F_i^{\downarrow}$
13	18	0,18	0,18	1
15	27	0,27	0,35	0,82
20	24	0,24	0,69	0,55
22	12	0,12	0,81	0,31
23	19	0,19	1	0,19
$\sum$	100	1	/	/

العرض البياني

يمكن عرضها البيانات بيانيا بإستخدام عدة طرق نذكر منها :

طريقة الأعمدة أو المستطيلات

هذه الطريقة بوضع المسمّيات على محور أفقي أو عمودي ورسم مستطيل على كل مسمّى بحيث يكون ارتفاع كل مستطيل ممثلّا للقيمة المقابلة لذلك المسمّى وذلك باستعمال مقياس رسم مناسب.

الطريقة الدائرية

استعمال هذه الطريقة يكون بتقسيم الكل إلى اجزائه، فيمثل المجموع الكلي بدائرة كاملة ويمثل كل جزء بقطاع دائرة وتعطى قياس زاويته بالعلاقة التالية:

$$\text{زاوية قطاع الدائرة لقيمة معينة} = \frac{\text{القيمة نفسها}}{\text{مجموع القيم}} \times 360^0$$

المدرج التكراري

عبارة عن مستطيلات متجاورة يخصص كل مستطيل لإحدى الفئات، حيث تتناسب مساحة المستطيلات مع تكرارت الفئة، يخصص المحور الأفقي للفئات، أما المحور العمودي فيخصص للتكرارات المقابلة لها.

الفصل الثالث

مقاييس النزعة المركزية

تعريف

تسمى مقاييس الترعة المركزية بمقاييس الموضع أو المتوسطات، وهى القيم التى تتركز القيم حولها ومن هذه المقاييس، الوسط الحسابي ، والمنوال ، والوسيط.

المتوسط الحسابي

تعريف: لتكن $x_1x_2, \dots, x_m$ مجموعة من القيم، تكراراتها $n_1, n_2, \dots, n_m$ على الترتيب، فإن المتوسط الحسابي لهذة القيم يعطى بالعلاقة التالية :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} n_i \times x_i = \frac{1}{n}(n_1 \times x_1 + \cdots + n_m \times x_m).$$

المنوال

تعريف: يعرف المنوال بأنه القيمة الأكثر تكرارا أو الأكثر شيوعا في العينة ويرمز له بالرمز M_0.

مثال: أوجد المنوال للبيانات التالية:

5 9 5 14 7 14 10 5 8

الوسيط

تعريف: يعرف الوسيط بأنه القيمة التتي تتوسط البيانات بعد ترتيبها تصاعديا أو تنازليا ويرمز له بالرمز M_e.

طريقة ايجاد الوسيط

- في حالة عدد القيم فردي: رتبة الوسيط هي القيمة التي تقابل $\frac{n+1}{2}$ أي

$$\mathrm{M_e} = \chi_{\left(\frac{n+1}{2}\right)}.$$

- في حالة عدد القيم زوجي: رتبة الوسيط هي متوسط القيمة التي تقابل $\frac{n}{2}$ و القيمة التي تقابل $\frac{n}{2}+1$ أي

$$\mathrm{M_e} = \frac{1}{2}\left(\chi_{\left(\frac{n}{2}\right)} + \chi_{\left(\frac{n}{2}+1\right)}\right).$$

مثال: أوجد الوسيط للبيانات التالية:

- 8 5 10 14 7 14 5 9 4.
- 17 10 21 31 21 13 11 18 7 9 7 3.

الفصل الرابع

مقاييس التشتت

تعريف

يقصد بالتشتت مدى تباعد قيم المتغير الإحصائي عن بعضها البعض أو عن القيمة المركزية.

المدى

تعريف: يعرف المدى بأنه الفرق بين أكبر و أصغر قيمة، ويرمز له بالرمز E.

التباين

تعريف: لتكن $x_1 x_2, \dots, x_m$ مجموعة من القيم، تكراراتها $n_1, n_2, \dots, n_m$ على الترتيب، فإن التباين لهذة القيم يعطى بالعلاقة التالية :

$$v_x = \frac{1}{n}\sum_{i=1}^{m} n_i \times (x_i - \bar{x})^2 = \left[\frac{1}{n}\sum_{i=1}^{m} n_i \times (x_i)^2\right] - (\bar{x})^2.$$

الانحراف المعياري

تعريف: يعرف الإنحراف المعياري بأنه جذر التباين.

$$\sigma = \sqrt{v_x}.$$

مثال: لتكن البيانات التالية

11 9 8 7 9 5

- أوجد المدى
- أحسب التباين والإنحراف المعياري.

معامل الإختلاف

يشير معامل التباين إلى المقياس الإحصائي الذي يساعد في قياس تشتت نقاط البيانات المختلفة في سلسلة البيانات حول المتوسط ويتم حسابه بقسمة

$$\mathrm{c.v} = \frac{\delta_e}{\bar{x}} \times 100\%.$$

تمارين

التمرين الأول: لتكن العينة التالية

1. ماهي الصفة المدروسة وما نوعها؟
2. ارسم جدولا احصائيا يمثل النتائج السابقة.
3. مثل بيانيا بواسطة الأعمدة المتغير الإحصائي المدروس .
4. أحسب المتوسط الحسابي.
5. أوجد المنوال و الوسيط.
6. أوجد المدى
7. أحسب التباين و الإنحراف المعياري.
8. احسب معامل الإختلاف

9 14 12 **7** **9** 8

التمرين الثاني: لتكن العينة التالية

4 9 5 14 7 14 10 5 8

1. ماهي الصفة المدروسة وما نوعها؟
2. ارسم جدولا احصائيا يمثل النتائج السابقة.
3. مثل بيانيا بواسطة الأعمدة المتغير الإحصائي المدروس .
4. أحسب المتوسط الحسابي.
5. أوجد المنوال و الوسيط.
6. أوجد المدى
7. أحسب التباين و الإنحراف المعياري.
8. احسب معامل الإختلاف

الفصل الخامس

التقديرات

تعريف

التقدير (Estimation) هو أسلوب إحصائي مبني على نظريات إحصائية، يستخدم لتقدير معلمة ما محل الاهتمام عن طريق استخدام مقاييس العينة.

تعريف

المقصود بالتقدير هو تقدير معالم المجتمع الإحصائي(أو التوزيع الاحتمالي)والتي غالباً ما تك ون مجهولة ويكون المطلوب هو الحصول على تقديرات لها من بيانات العينة.

التقدير بنقطة

تستخدم بيانات العينة لتقدير معلمة المجتمع المجهولة بنقطة واحدة فقط، أي بقيمة واحدة.

التقديربفترة

- تستخدم بيانات العينة لتقدير معلمة المجتمع المجهولة بفترة من القيم.
- يقصد بتقدير فترة ثقة للمتوسط في المجتمع هو إيجاد الحدين الأدنى، والأعلى اللذان يقع بينهما قيمة متوسط المجتمع باحتمال قدره **$(1-\alpha)\%$**

ملاحظة

- فترة الثقة كما سبق بأنها الفترة التي يقع داخلها معلمة المجتمع باحتمال او بمعامل ثقة معينة.
- التقدير بمجال يسمى أيضا" مجال الثقة "لأن هذه الفترات تعتمد في تكوينها الإحصائي على درجات أو مستويات ثقة.

دقة التقدير

كلما تناقص طول مجال الثقة تزداد دقة التقدير.

التقدير بمجال المتوسط الحسابي

الانحراف المعياري للمجتمع $\boldsymbol{\sigma}$ موجود

$$\mathbf{I.c} = \left[\overline{\boldsymbol{x}} - \varepsilon_{\alpha}\frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{N}}}, \overline{\boldsymbol{x}} + \varepsilon_{\alpha}\frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{N}}}\right]$$

الانحراف المعياري للمجتمع $\boldsymbol{\sigma}$ غير موجود و $\mathrm{N} < 30$

$$\mathbf{I.c} = \left[\overline{\boldsymbol{x}} - \mathrm{T}_{(\alpha,\mathrm{N}-1)}\frac{\widehat{\boldsymbol{\sigma}}}{\sqrt{\boldsymbol{N}}}, \overline{\boldsymbol{x}} + \boldsymbol{T}_{(\boldsymbol{\alpha},\boldsymbol{N}-\mathbf{1})}\frac{\widehat{\boldsymbol{\sigma}}}{\sqrt{\boldsymbol{N}}}\right]$$

$\mathrm{T}_{(\alpha,\mathrm{N}-1)} \longrightarrow$ Table de Student

الانحراف المعياري للمجتمع $\boldsymbol{\sigma}$ غير موجود و $\mathrm{N} > 30$

$$\mathbf{I.c} = \left[\overline{\boldsymbol{x}} - \varepsilon_{\alpha}\frac{\widehat{\boldsymbol{\sigma}}}{\sqrt{\boldsymbol{N}}}, \overline{\boldsymbol{x}} + \varepsilon_{\alpha}\frac{\widehat{\boldsymbol{\sigma}}}{\sqrt{\boldsymbol{N}}}\right]$$

التقدير بمجال الثقة تباين المجتمع

المتوسط الحساب للمجتمع $\boldsymbol{\mu}$ غير موجود

$$\mathbf{I.c} = \left[\frac{\sum(x_i - \bar{x})^2}{\chi^2_{1-\frac{\alpha}{2}}(N-1)}, \frac{\sum(x_i - \bar{x})^2}{\chi^2_{\frac{\alpha}{2}}(N-1)}\right],$$

حيث

$$\sum(x_i - \bar{x})^2 = N \times (\sigma_e)^2$$

التمرين الأول

البيانت التالية تمثل كمية الهيموغلوبين في الدم لمجموعة من الأشخاص (27 شخص) مقاس ب الغرام/ 100مل:

$$\overline{\boldsymbol{x}} = \mathbf{12}. \qquad \boldsymbol{\delta_e} = \mathbf{1,53}.$$

- علّق على المتوسط الحسابي .
- قدّر بمجال ثقة 95 % المتوسط الحسابي للعينة السابقة.
- قدّر بمجال ثقة تباين المجتمع السابق عند مستوى 95 %.

$\boldsymbol{T}_{(\mathbf{0,05;26})} = \mathbf{2,056.}$	$\chi^2_{0,975}(26) = 41{,}92.$	$\chi^2_{0,025}(26) = 13{,}34.$

الحل

التعليق على المتوسط الحسابي

$$\mathrm{c.v} = \frac{\sigma_e}{\bar{x}} \times 100\% = \frac{1{,}53}{12} \times 100\% = 12{,}75\% < 20\%.$$

اذن العينة متجانسة و $\bar{x}$ دقيق في الوصف.

التقدير بمجال الثقة المتوسط الحسابي

- $\mathrm{N} = 27 < 30$ و غير موجود σ
- $\alpha = 0{,}05$

$$Ic = \left[\bar{x} - T_{(\alpha;N-1)}\frac{\hat{\sigma}}{\sqrt{N}}, \bar{x} - T_{(\alpha;N-1)}\frac{\hat{\sigma}}{\sqrt{N}}\right]$$

$$\hat{\sigma} = \sqrt{\frac{N}{N-1}} \times \sigma_e = \sqrt{\frac{27}{26}} \times 1{,}53 = 1{,}01 \times 1{,}53 = 1{,}54.$$

$$T_{(\alpha;N-1)} = T_{(0,05;5)} = \mathbf{2,056}.$$

اذن

$$Ic = \left[12 - \mathbf{2,056}\frac{1{,}54}{\sqrt{27}}; 12 + \mathbf{2,056}\frac{1{,}54}{\sqrt{27}}\right]$$
$$= [12 - (\mathbf{2,056} \times 0{,}29); 12 + (\mathbf{2,056} \times 0{,}29)]$$
$$= [12 - 0{,}59; 12 + 0{,}59]$$
$$= [11{,}41; 12{,}49].$$

التقدير بمجال الثقة تباين المجتمع

- غير موجود μ
- $\alpha = 0.05$

$$Ic = \left[\frac{\sum(x_i - \bar{x})^2}{\chi^2_{1-\frac{\alpha}{2}}(N-1)}, \frac{\sum(x_i - \bar{x})^2}{\chi^2_{\frac{\alpha}{2}}(N-1)}\right]$$

$$\sum(x_i - \bar{x})^2 = N \times (\sigma_e)^2 = 27 \times (1{,}54)^2 = 64{,}03.$$

$$\chi^2_{0,975}(26) = 41,92. \qquad \chi^2_{0,025}(26) = 13,34.$$

اذن

$$Ic = \left[\frac{64,03}{41,62}; \frac{64,03}{13,34}\right] = [0,93; 38,39].$$

التمرين الثاني

أراد أحد الأطباء تقدير المدة اللازمة لإجراء نوع معين من العمليات الجراحية ، فقام بقياس الزمن اللازم (بالساعات) لإجراء مجموعة عشوائية من العمليات الجراحية، فتحصل على النتائج التالية:

10 6 8 9 7,5 5,5 4

1. أحسب المتوسط الحسابي لهذه العينة و الانحراف المعياري لها، ثم علق على المتوسط الحسابي.
2. قدّر بمجال الثقة تباين المجتمع السابق عند مستوى ثقة **99 %**.
3. قدّر بمجال ثقة **99 %** المتوسط الحسابي للعينة السابقة.

$\boldsymbol{T_{(0,05;26)} = 2,056.}$	$\chi^2_{0,995}(6) = 18,55.$	$\chi^2_{0,005}(6) = 0,68$

الحل

x_i	4	5,5	6	7,5	8	9	10	
n_i	1	1	1	1	1	1	1	N=7

.1

المتوسط الحسابي

$$\bar{x} = \frac{1}{N}\sum n_i x_i = \frac{1}{7}(4 + 5,5 + 6 + 7,5 + 8 + 9 + 10) = \frac{1}{7}(50) = 7,14.$$

الإنحراف المعياري

$$\sigma_e = \sqrt{v_x}$$

$$v_x = \left[\frac{1}{N}\left(\sum (x_i)^2\right)\right] - (\bar{x})^2$$

$$= \left[\frac{1}{7}((4)^2 + (5{,}5)^2 + (6)^2 + (7{,}5)^2 + (8)^2 + (9)^2 + (10)^2)\right] - (7{,}14)^2$$

$$= \left(\frac{383{,}5}{7}\right) - (7{,}14)^2 = 54{,}78 - 50{,}97 = 3{,}81.$$

$$\sigma_e = \sqrt{v_x} = \sqrt{3{,}81} = 1{,}95.$$

التعليق على تشتت العينة

$$c.v = \frac{\sigma_e}{\bar{x}} \times 100\% = \frac{1{,}95}{7{,}14} \times 100\% = 27{,}31\% > 20\%.$$

اذن العينة غير متجانسة و $\bar{x}$ غير دقيق في الوصف.

2. التقدير بمجال الثقة تباين المجتمع

- μ غير موجود
- $\alpha = 0.01$

$$Ic = \left[\frac{\sum (x_i - \bar{x})^2}{\chi^2_{1-\frac{\alpha}{2}}(N-1)}, \frac{\sum (x_i - \bar{x})^2}{\chi^2_{\frac{\alpha}{2}}(N-1)}\right].$$

$$\sum (x_i - \bar{x})^2 = N \times (\sigma_e)^2 = 7 \times (1{,}95)^2 = 26{,}6.$$

$\chi^2_{0{,}995}(6) = 18{,}55.$ $\chi^2_{0{,}005}(6) = 0{,}68.$

اذن

$$Ic = \left[\frac{26{,}6}{18{,}55}; \frac{26{,}6}{0{,}68}\right] = [1{,}43; 39{,}11].$$

3. التقدير بمجال الثقة المتوسط الحسابي

- $N = 7 < 30$ و σ غير موجود
- $\alpha = 0{,}01$

$$Ic = \left[\bar{x} - T_{(\alpha;N-1)}\frac{\hat{\sigma}}{\sqrt{N}}, \bar{x} - T_{(\alpha;N-1)}\frac{\hat{\sigma}}{\sqrt{N}}\right]$$

$$\hat{\sigma} = \sqrt{\frac{N}{N-1}} \times \sigma_e = \sqrt{\frac{7}{6}} \times 1{,}95 = 1{,}08 \times 1{,}95 = 2{,}1.$$

$$T_{(\alpha;N-1)} = T_{(0{,}01;5)} = 3{,}707.$$

اذن

$$Ic = \left[7{,}14 - 3{,}707\frac{2{,}1}{\sqrt{7}}; 7{,}14 + 3{,}707\frac{2{,}1}{\sqrt{7}}\right]$$

$$= [7{,}14 - (3{,}707 \times 0{,}79); 7{,}14 + (3{,}707 \times 0{,}79)]$$

$$= [7{,}14 - 2{,}92; 7{,}14 + 2{,}92]$$

$$= [4{,}22; 10{,}06].$$

التمرين الثالث

مجتمع يتبع توزيعا طبيعيا انحرافه المعياري مجهول. سحبنا منه عيّنة عشوائية حجمها 30 شخصا فكان المتوسط الحسابي لهذه العينة يقدر ب 48 مع انحراف معياري قدره 11.

- علّق على المتوسط الحسابي .
- قدّر بمجال ثقة 95 % المتوسط الحسابي للعينة السابقة.
- قدّر بمجال ثقة تباين المجتمع السابق عند مستوى 95 %.

$\varepsilon_\alpha = \varepsilon_{0{,}05} = \mathbf{1{,}96}.$	$\chi^2_{0{,}975}(29) = 45{,}72.$	$\chi^2_{0{,}025}(29) = 16{,}05.$

الحل

التعليق على المتوسط الحسابي

$$c.v = \frac{\sigma_e}{\bar{x}} \times 100\% = \frac{11}{48} \times 100\% = 22{,}91\% > 20\%.$$

اذن العينة غير متجانسة و $\bar{x}$ غير دقيق في الوصف.

التقدير بمجال الثقة المتوسط الحسابي

- $N = 30 \geq 30$ و غير موجود σ
- $\alpha = 0{,}05$

$$Ic = \left[\bar{x} - \varepsilon_\alpha \frac{\hat{\sigma}}{\sqrt{N}}, \bar{x} + \varepsilon_\alpha \frac{\hat{\sigma}}{\sqrt{N}}\right]$$

$$\hat{\sigma} = \sqrt{\frac{N}{N-1}} \times \sigma_e = \sqrt{\frac{30}{29}} \times 11 = 1{,}01 \times 11 = 11{,}11.$$

$$\varepsilon_\alpha = \varepsilon_\alpha = \mathbf{1,96}.$$

اذن

$$\begin{aligned} Ic &= \left[48 - \mathbf{1,96}\frac{11{,}11}{\sqrt{30}}; 48 + \mathbf{1,96}\frac{11{,}11}{\sqrt{30}}\right] \\ &= [48 - (\mathbf{1,96} \times 2{,}03); 48 + (\mathbf{1,96} \times 2{,}03)] \\ &= [48 - 3{,}97; 48 + 3{,}97] \\ &= [44{,}03; 51{,}97]. \end{aligned}$$

التقدير بمجال الثقة تباين المجتمع

- غير موجود μ
- $\alpha = 0.05$

$$Ic = \left[\frac{\sum(x_i - \bar{x})^2}{\chi^2_{1-\frac{\alpha}{2}}(N-1)}, \frac{\sum(x_i - \bar{x})^2}{\chi^2_{\frac{\alpha}{2}}(N-1)}\right]$$

$$\sum(x_i - \bar{x})^2 = N \times (\sigma_e)^2 = 30 \times (11)^2 = 3630.$$

$$\chi^2_{0{,}975}(29) = 45{,}72. \qquad \chi^2_{0{,}025}(29) = 16{,}05.$$

اذن

$$Ic = \left[\frac{3630}{45,72}; \frac{3630}{16,05}\right] = [79,39; 226,16].$$

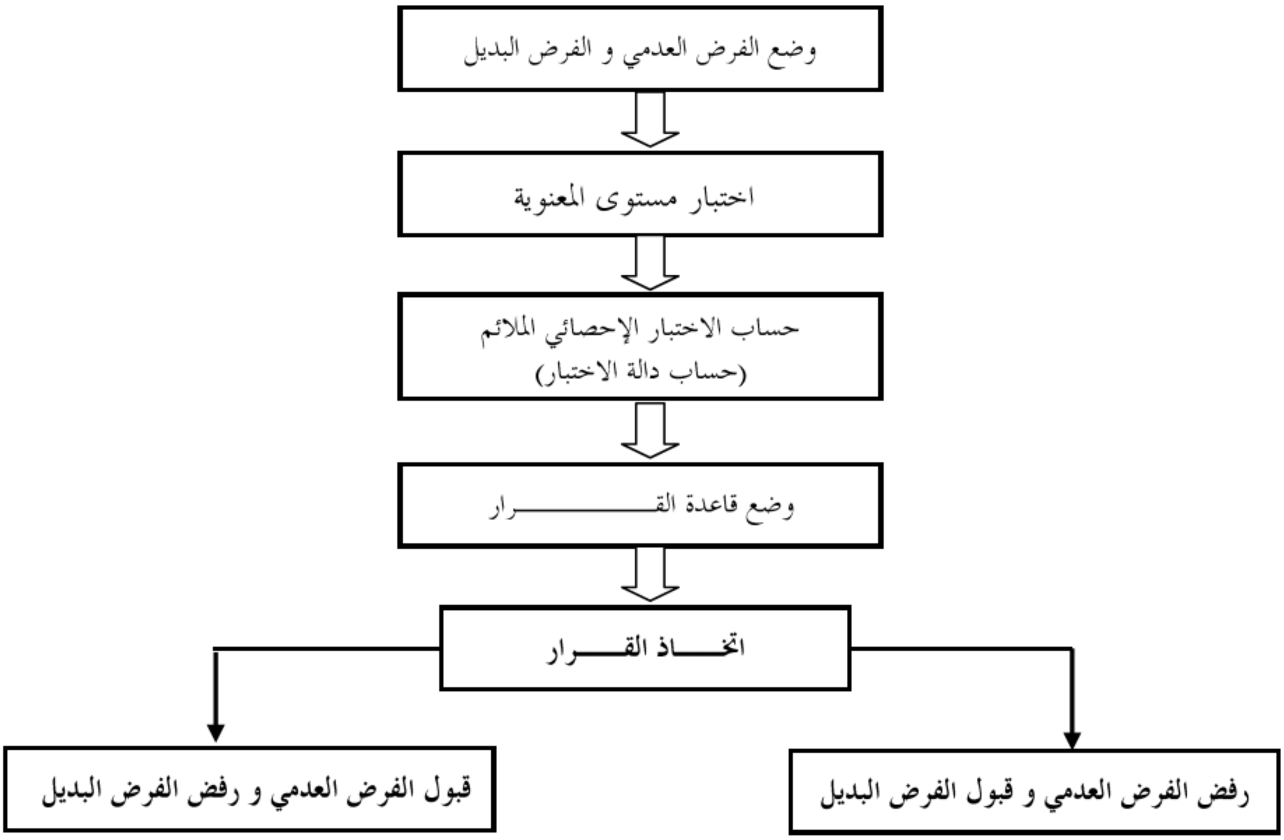
الفصل السادس

اختبار الفرضيات (عينة)

تمهيد

يعتبر إختبار الفروض أحد المواضيع الرئيسية للإستدلال الإحصائي، ويستهدف الوصول إلى القرار بشأن معلمة المجتمع من خلال قبول او رفض تقديرها المعتمد على معطيات العينة المسحوبة من ذلك المجتمع.

خطوات إختبا رات الفروض



صياغة الفرضيات الإحصائية:

الفرضية الإحصائية عبارة عن أدعاء (قد يكون صائبا أو خطأ) حول معلمة أو أكثر لمجتمع أو لمجموعة من المجتمعات.

الفرضية العدمية

الفرضية العدمية هي" الفرضية الاساسية المراد اختبارها ويرمز لها عادة بالرمز H_0.

الفرضية البديلة

في اختبارات الفرضيات يتحتم وضع فرضية أخرى غير الفرضية العدمية المراد اختبارها تسمى الفرضية البديلة .

وهذه الفرضية " هي التي ستقبل في حالة رفض الفرضية العدمية" ويرمز لها عادة بالرمز H_1.

مستوى المعنوية α و درجة الثقة $1-\alpha$

- إن القرار الذي سوف نتخذه بناء على الاختبار الإحصائي لا يمكن اعتباره صحيح 100 % فهناك مقدار من الخطأ لأن المعلومات التي نتخذ قرارنا بناء عليها بيانات مأخوذة من عينة وليس من المجتمع الأصلي .
- في اختبار فرض معين، فإن مقدار ثقتنا في القرار المتخذ بالرفض أو القبول يسمى بدرجة الثقة ويرمز له بالرمز $(1-\alpha)$ كما وأن مقدار عدم الثقة أو مقدار الخطأ يسمى بمستوى المعنوية ويرمز له بالرمز α.

ملاحظة

- عادة يحدد الباحث مستوى المعنوية أو درجة الثقة قبل البدء في عملية الاختبار.
- إذا استخدمنا مستوى معنوية $\alpha = 0.05$ مثلا فهذا يعني انه في المتوسط من بين كل 100 حالة يكون في 95 حالة منها قرارنا سليم وفي الخمس الحالات الباقية قرارنا خطأ

إحصاء الاختبار(دالة الاختبار)

إحصاء الاختبار هو اقتران إحصائي يساعدنا على اتخاذ قرار حول فرضية إحصائية معينة، ويتم حساب قيمته من بيانات العينة، وبالتالي فهو عبارة عن متغير عشوائي تتغير قيمته بتغير بيانات العينة الإحصائية التي نأخذها من المجتمع الإحصائي.

المتوسط الحسابي

الانحراف المعياري للمجتمع σ موجود

$$\boldsymbol{\varepsilon}_{\mathbf{cal}} = \frac{|\boldsymbol{\mu} - \overline{\boldsymbol{X}}|}{\frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{N}}}} \quad \longrightarrow \quad \varepsilon_{\alpha}$$

الانحراف المعياري للمجتمع $\boldsymbol{\sigma}$ غير موجود و $\mathrm{N} < 30$

$$\mathbf{T_{cal}} = \frac{|\boldsymbol{\mu} - \overline{\boldsymbol{X}}|}{\frac{\widehat{\boldsymbol{\sigma}}}{\sqrt{\boldsymbol{N}}}} \qquad \mathrm{T}_{(\alpha,\mathrm{N}-1)} \longrightarrow \text{Table de Student}$$

الانحراف المعياري للمجتمع $\boldsymbol{\sigma}$ غير موجود و $\mathrm{N} > 30$

$$\boldsymbol{\varepsilon}_{\mathbf{cal}} = \frac{|\boldsymbol{\mu} - \overline{\boldsymbol{X}}|}{\frac{\widehat{\boldsymbol{\sigma}}}{\sqrt{\boldsymbol{N}}}} \quad \longrightarrow \varepsilon_{\alpha}$$

<u>النسبة المئوية</u>

$$\boldsymbol{\varepsilon}_{\mathbf{cal}} = \frac{|\boldsymbol{P} - \boldsymbol{p_0}|}{\sqrt{\frac{\boldsymbol{p_0} \times \boldsymbol{q_0}}{\boldsymbol{N}}}} \qquad \varepsilon_{\alpha}$$

- **المقارنة واتخاذ القرار**

المقارنة بين القيمة الجدولية و القيمة الحسابية لدالة الاختبار

التمرين الاول

عند الأشخاص العاديين درجة الحرارة تتبع قانونا طبيعيا ب $\mu = 37$ وانحرافه المعياري مجهول .

قمنا بقياس درجة الحرارة لعيّنة عشوائية من 28 شخص فوجدنا أن متوسط درجة الحرارة عندهم $\bar{x} =$ 38.5 بإنحراف معياري $\sigma_e = 6.25$.

$$T_{(0,01;27)} = 2,771$$

- هل تعتبر هذه العينة طبيعية عند مستوى ثقة 99% ؟

الحل

$$\mu = 37. \quad N = 28. \quad \bar{x} = 38,5. \quad \sigma_e = 6.25\,.$$

- هل تعتبر هذه العينة طبيعية عند مستوى ثقة 99% ؟

- $\begin{cases} H_0: \ \mu = \bar{x} \\ H_1: \mu \neq \bar{x} \end{cases}$
- $\alpha = 0,01.$
- اختبار الفرضيات : المقارنة بين متوسطين
- $N < 30$ و غير موجود σ

$$T_{cal} = \frac{|\mu - \bar{x}|}{\frac{\hat{\sigma}}{\sqrt{N}}}.$$

$$\hat{\sigma} = \sqrt{\frac{N}{N-1}} \times \sigma_e = \sqrt{\frac{28}{27}} \times 6,25 = 1,01 \times 6,25 = 6,31.$$

اذن

$$\varepsilon_{cal} = \frac{|37 - \ 38,5|}{\frac{6,31}{\sqrt{28}}} = \frac{1,5}{1,19} = 1,26$$

لدينا

$$T_{cal} = 1,26 < T_{(0,01;27)} = 2,771$$

اذن H_0 مقبول ومنه العينة طبيعية عندى مستوى ثقة 99%

التمرين الثاني

ضغط الدم الإنقباضي عند الإنسان العادي يتبع توزيعا طبيعيا يقدرب 88 مم/زئبق وانحراف المعياري مجهول .

أخذنا عيّنة عشوائية حجمها 100 شخصا من مجتمع معين، و قسنا الضغط الدموي الإنقباضي لكل فرد من هذه العيّنة، فوجد أن المتوسط الحسابي هو 90 مم/ زئبق بانحراف معياري قدره 10 مم/ زئبق.

$$\varepsilon_\alpha = \varepsilon_{0,05} = 1,96$$

- هل تعتبر هذه العينة طبيعية للمجتمع عند مستوى ثقة **95 %**؟

الحل

$$\mu = 88. \quad N = 100. \quad \bar{x} = 90. \quad \sigma_e = 10\,.$$

- $\begin{cases} H_0:\ \mu = \bar{x} \\ H_1: \mu \neq \bar{x} \end{cases}$
- $\alpha = 0,01.$
- اختبار الفرضيات : المقارنة بين متوسطين
- σ غير موجود و $\mathrm{N} > 30$

$$\varepsilon_{cal} = \frac{|\mu - \ \bar{x}|}{\frac{\hat{\sigma}}{\sqrt{N}}}.$$

$$\hat{\sigma} = \sqrt{\frac{N}{N-1}} \times \sigma_e = \sqrt{\frac{100}{99}} \times 10 = 1,005 \times 10 = 10,05.$$

اذن

$$\varepsilon_{cal} = \frac{|88 - \ 90|}{\frac{10,05}{\sqrt{100}}} = \frac{2}{1,005} = 1,99$$

لدينا

$$\varepsilon_{cal} = 1,99 > \varepsilon_{0,05} = 1,96$$

اذن H_0 مقبول ومنه العينة غير طبيعية عندى مستوى ثقة 95%

التمرين الثالث

نسبة الإصابة بمرض التوحّد عند الأطفال في مجتمع معيَن تتبع توزيعا طبيعيا بنسبة 1**%.**

أخذنا عيّنة عشوائية حجمها من 500 طفل فتبين أن 30 منهم مصابون بهذا المرض.

$$\varepsilon_\alpha = \varepsilon_{0,01} = 2{,}58$$

- هل تعتبر هذه العينة ممثلة للمجتمع عند مستوى ثقة 99% ؟

الحل

$$P = 1\% = 0{,}01. \quad N = 500. \quad k = 30.$$

- $\begin{cases} H_0: P = p_0 \\ H_1: \ P \neq p_0 \end{cases}$
- $\alpha = 0{,}01.$
- اختبار الفرضيات : المقارنة بين نسبتين
-

$$\boldsymbol{\varepsilon}_{\mathbf{cal}} = \frac{|\boldsymbol{P} - p_0|}{\sqrt{\frac{p_0 \times \boldsymbol{q_0}}{\boldsymbol{N}}}}$$

$$p_0 = \frac{k}{N} = \frac{30}{500} = 0{,}06.$$

$$\boldsymbol{q_0} = \mathbf{1} - p_0 = \mathbf{1} - 0{,}06 = 0{,}94$$

اذن

$$\boldsymbol{\varepsilon}_{\mathbf{cal}} = \frac{|0{,}01 - 0{,}06|}{\sqrt{\frac{0{,}06 \times 0{,}94}{\mathbf{500}}}} = \mathbf{5}$$

لدينا

$$\varepsilon_{cal} = 5 > \varepsilon_{0,01} = 2{,}58$$

اذن H_0 مرفوض ومنه العينة غير ممثلة للمجتمع عندى مستوى ثقة 99%

الفصل السابع

اختبار الفرضيات (عينتين)

اختبار الفرضيات الفرقية

اختبار الفرضيات الفرقية

اختبار " ت " الفروق بين مجموعتين مستقلتين وغير متساويتين.

اختبار " ت " الفروق بين مجموعتين مستقلتين و متساويتين.

اختبار " ت " الفروق بين مجموعتين مرتبطتين.

اختبار " ف " الفروق لأكثر من مجموعتين تحليل التباين.

اختبار الفروق في التكرارات كا2

اختبار ت

المقصود بالفرضيات الإحصائية تلك الفرضيات التي تتعلق بالمجتمع الإحصائي المسحوبة منه العينة، أو توزيع هذا المجتمع أو معالمه كالوسط الحسابي أو النسبة في المجتمع.

اختبار" ت"

يعتبر اختبار' ت 'احد أشكال اختبار الفرضيات ، وواحد من الاختبارات العديدة المستخدمة لهذا الغرض، وترجع نشأته إلى العالم وليام سيلي جوست سنة 1908 حيث نسب الاختبار للاسم المستعار " ستيودنت "اي الطالب.

شروط استخدام اختبار" ت"

- **مدى تجانس العينة :** ويقصد بتجانس العينات أي مدى انتسابها لأصل واحد بمعنى لهم نفس الخصائص.
- **الفرق بين حجم عينتي البحث :** أو بما يعرف بشرط التقارب ، يجب أن يكون حجم العينتين متقاربا لان
- الحجم يؤثر على مستوى الدلالة لاختبار" ت".
- **حجم العينة :** يجب أن يزيد حجم كل من العينيتين عن 5 أفراد.

اختبار " ت " الفروق بين مجموعتين مستقلتين وغير متساويتين

$$N_1 \neq N_2$$

<u>المتوسط الحسابى</u>

الانحراف المعياري للمجتمع $\boldsymbol{\sigma}$ غير موجود و $N_1 < 30$ او $N_2 < 30$

$$\mathbf{T_{cal}} = \boldsymbol{T_{cal}} = \frac{|\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2|}{\hat{\boldsymbol{\sigma}}\sqrt{\frac{\mathbf{1}}{\boldsymbol{N_1}} + \frac{\mathbf{1}}{\boldsymbol{N_2}}}}, \quad \mathrm{T}_{(\alpha, N_1 + N_2 - 2)} \longrightarrow \text{Table de Student}$$

حيث

$$\hat{\boldsymbol{\sigma}} = \left[\frac{\boldsymbol{N_1 . \sigma_{e1}^2 + N_2 . \sigma_{e2}^2}}{\boldsymbol{N_1 + N_2 - 2}}\right]^{\frac{1}{2}}$$

الانحراف المعياري للمجتمع $\boldsymbol{\sigma}$ غير موجود و $N_1 > 30$ او $N_2 > 30$

$$\boldsymbol{\varepsilon}_{\mathbf{cal}} = \frac{|\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2|}{\sqrt{\frac{\left(\widehat{\boldsymbol{\sigma}_{e_1}}\right)^2}{\boldsymbol{N}_1} + \frac{\left(\widehat{\boldsymbol{\sigma}_{e_2}}\right)^2}{\boldsymbol{N}_2}}} \longrightarrow \varepsilon_\alpha$$

التمرين الاول

أجريت دراسة للمقارنة بين متوسطي أعمار سكان المدن والقرى علما أنهما يخضعان لتوزيع طبيعي. أخذنا عينة عشوائية من سكان المدن حجمها 14 شخص فوجدنا أن متوسط العمر لديهم يقدرب 63 سنة بإنحراف معياري قدره 7.

كما أخذنا عينة أخرى من سكان القرى حجمها 15 شخص فوجدنا أن متوسط العمر لهذه العينة يقدرب 60 سنة مع إنحراف معياري قدره 9.

- هل هنالك اختلاف معنوي بين متوسط أعمار سكان المدن والقرى عند مستوى ثقة **95 %**؟

الحل

$$N_1 = 14. \quad \bar{x}_1 = 63. \quad \sigma_{e_1} = 7\,.$$

$$N_2 = 15. \quad \bar{x}_2 = 60. \quad \sigma_{e_2} = 9\,.$$

هل هنالك اختلاف معنوي بين متوسط أعمار سكان المدن والقرى عند مستوى ثقة 99% ؟

- $\begin{cases} H_0: \mu_1 = \mu_2, \\ H_1: \mu_1 \neq \mu_2. \end{cases}$
- $\alpha = 0{,}01.$
- اختبار التجانس : المقارنة بين متوسطين حسابيين
- $\mathrm{N}_1 < 30,\ \mathrm{N}_2 < 30.$

$$T_{cal} = \frac{|\,\bar{x}_1 - \ \bar{x}_2|}{\hat{\sigma}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}.$$

لدينا

$$\hat{\sigma}^2 = \frac{N_1.\sigma_{e1}^2 + N_2.\sigma_{e2}^2}{N_1 + N_2 - 2} = 73{,}3.$$

$$\Rightarrow\ \hat{\sigma} = \sqrt{73{,}3} = 8{,}56.$$

ومنه

$$T_{cal} = \frac{|63 - 60|}{\hat{\sigma}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} = 0,94.$$

$$T_{(\alpha, N_1+N_2-2)} = T_{(0,005;27)} = 2,052.$$

لدينا

$$T_{cal} = 0,94 < T_{(0,005;27)} = 2,052.$$

اذن H_0 مقبول عند مستوى ثقة 95%، ومنه لايوجد اختلاف معنوي بين متوسط أعمار سكان المدن والقرى.

التمرين الثاني

في دراسة للمقارنة بين طريقة التعلم المبرمج و طريقة المحاضرة من حيث تأ ثيرهما على تحصيل الطلبة لإحدى المواد المُدرسة بالإعتماد على نتائج الإختبار المتحصل عليها في تلك المادة.

إخترنا عينة عشوائية مؤلفة من 30 طالب فوجدنا أن متوسط نتائج الإختبار لديهم تقدرب 85 درجة بإنحراف معياري قدره 8 درجات. كما أخذنا عينة ثانية مكونة من 32 طالب فوجدنا أن متوسط نتائج الإختبار لديهم تقدرب 90 درجة مع إنحراف معياري قدره 10 درجات.

- هل هنالك اختلاف بين المتوسط الحسابي للعينتين السابقتين عند مستوى ثقة 95 % ، علما أن إحدى العينتين درست عن طريق التعليم المبرمج أما الأخرى عن طريق المحاضرة؟

الحل

$$N_1 = 30. \quad \bar{x}_1 = 85. \quad \sigma_{e_1} = 8\,.$$

$$N_2 = 32. \quad \bar{x}_2 = 90. \quad \sigma_{e_2} = 10\,.$$

هل هنالك اختلاف بين المتوسط الحسابي للعينتين السابقتين عند مستوى ثقة 95 % ؟

- $\begin{cases} H_0: \mu_1 = \mu_2, \\ H_1: \mu_1 \neq \mu_2. \end{cases}$
- $\alpha = 0,05.$
- اختبار ت : المقارنة بين متوسطين حسابيين
- $\mathrm{N}_1 > 30,\ \mathrm{N}_2 > 30.$

$$\varepsilon_{cal} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\hat{\sigma}_1^2}{N_1} + \frac{\hat{\sigma}_2^2}{N_2}}}.$$

لدينا

$$\hat{\sigma}_1 = \sqrt{\frac{N_1}{N_1 - 1}} \times \sigma_{e_1} = \sqrt{\frac{30}{29}} \times 8 = 1{,}01 \times 8 = 8{,}08 \Rightarrow \hat{\sigma}_1^2 = 65{,}28.$$

$$\hat{\sigma}_2 = \sqrt{\frac{N_2}{N_2 - 1}} \times \sigma_{e_2} = \sqrt{\frac{32}{31}} \times 10 = 1{,}01 \times 8 = 10{,}1 \Rightarrow \hat{\sigma}_2^2 = 102{,}01.$$

ومنه

$$\varepsilon_{cal} = \frac{|85 - 90|}{\sqrt{\frac{65{,}28}{30} + \frac{102{,}01}{32}}} = \frac{5}{\sqrt{2{,}17 + 3{,}18}} = 2{,}31.$$

لدينا

$$\varepsilon_{cal} = 2.31 > \varepsilon_{0{,}05} = 1{,}96.$$

اذن H_0 مرفوض عند مستوى ثقة 95%، ومنه يوجد اختلاف بين المتوسط الحسابي للعينتينل السابقتين.

اختبار " ت " الفروق بين مجموعتين مستقلتين و متساويتين

$$N_1 = N_2 = N$$

<u>المتوسط الحسابي</u>

الانحراف المعياري للمجتمع $\boldsymbol{\sigma}$ غير موجود

$$\mathbf{T_{cal}} = \frac{|\bar{\boldsymbol{X}}_{\mathbf{1}} - \bar{\boldsymbol{X}}_{\mathbf{2}}|}{\sqrt{\frac{\left(\widehat{\boldsymbol{\sigma}_{\boldsymbol{e}_{\mathbf{1}}}}\right)^{\mathbf{2}} + \left(\widehat{\boldsymbol{\sigma}_{\boldsymbol{e}_{\mathbf{2}}}}\right)^{\mathbf{2}}}{N-1}}}, \quad \mathrm{T}_{(\alpha,2\mathrm{N}-2)} \longrightarrow \text{Table de Student}$$

اختبار " ت " الفروق بين مجموعتين مرتبطتين

يستخدم هذا الاختبار في مقارنة متوسط عينتين مرتبطتين (مثل ان يكون الافراد في العينة الاولى هم نفسهم في العينة الثانية)

$$\boldsymbol{T_{cal}} = \frac{\sum \boldsymbol{d}}{\sqrt{\frac{\boldsymbol{N}\sum \boldsymbol{d} - (\sum \boldsymbol{d})^2}{\boldsymbol{N-1}}}}, \quad \mathrm{T}_{(\alpha, N-1)} \longrightarrow \text{Table de Student}$$

حيث:

- d: الفرق بين القياس القبلي والبعدي ؛
- N: *عدد افراد العينة*.

تمرين

أراد باحث تجريب فعالية دواء لمعالجة حالة الاكتئاب الحاد، فاختار عينة من 8 أفراد مكتئبين، وقام بقياس درجة الاكتئاب لديهم قبل تجريب الدواء، ثم بعد تناولهم للدواء بمدة معينة، وافترض أنه لا يوجد اختلاف في درجة الاكتئاب سواء قبل تناول الدواء أو بعده .وهذه بيانات القياسين:

الافراد	1	2	3	4	5	6	7	8
القياس القبلي	8	7	5	4	6	7	8	3
القياس البعدي	8	9	4	8	8	10	10	5

- هل توجد فرق بين القياسين القبلي والبعدي عند مستوى الدلالة 0.05 ؟

الحل

القياس القبلي	8	7	5	4	6	7	8	3	
القياس البعدي	8	9	4	8	8	10	10	5	
d	0	-2	1	-4	-2	-3	-2	-2	-14
d^2	0	4	1	16	4	9	4	4	42

$$\mathbf{N} = 8$$

- $\begin{cases} H_0: \mu_1 = \mu_2, \\ H_1: \mu_1 \neq \mu_{2.} \end{cases}$
- $\alpha = 0{,}05.$
- اختبار ت : المقارنة بين متوسط عينتين مرتبطتين
-

$$\boldsymbol{T_{cal}} = \frac{\sum \boldsymbol{d}}{\sqrt{\frac{\boldsymbol{N}\sum \boldsymbol{d}^2 - (\sum \boldsymbol{d})^2}{\boldsymbol{N-1}}}} = \frac{-14}{\sqrt{\frac{(\boldsymbol{8} * 42) - (-14)^2}{\boldsymbol{7}}}} = -3{,}13$$

لدينا

$$\boldsymbol{T_{cal}} = -3{,}13 < T_{(0{,}05;7)} = 2{,}365.$$

اذن H_0 مقبول عند مستوى ثقة 95%، ومنه لايوجد بين القياسين القبلي والبعدي عند مستوى الدلالة 0.05

اختبار الفرضيات الارتباطية

تناولنا في الأبواب السابقة طرق دراسة متغير واحد لأي ظاهرة محل الدراسة و تلخيص البيانات في جداول توزيعات تكرارية وكيفية عرضها بيانيا. كذلك دراسة بعض المقاييس العددية التي تساعد على معرفة بعض خصائص التوزيعات التكرارية، ومنها مقاييس النزعة المركزية ومقاييس التشتت والالتواء والتفرطح.

سوف نتناول الآن دراسة البيانات التى يكون لأفرادها متغيران يتغيران معا فى وقت واحد، وذلك لمعرفة نوع العلاقة التى تربط بينهما.

تعريف: الارتباط هو تعيين طبيعة وقوة العلاقة بين متغيرين X و Y. وتقاس العلاقة بمؤشر إحصائي يدعى معامل الارتباط ويرمز له بالرمز (r) وهناك عدة أنواع من معاملات الارتباط منها البسيط والجزئي ومعامل ارتباط الرتب ومعامل ارتباط فاي ... الخ.

الارتباط الموجب الطردي

فإن المتغير الآخر يتبعه بنفس الاتجاه.y أو xإذا تغير أحد المتغيرين

الارتباط السالب العكسي

إذا تغير أحد المتغيرين x أو y فإن المتغير الآخر يتبعه بالاتجاه المضاد.

خصائص معامل الارتباط

- تتراوح قيمة معاملة الارتباط بين -1 و +1 $(+1 \geq r \geq -1)$، إذ يمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن (± 1).
- تكون قيمته تساوي صفراً عندما يكون المتغيران مستقلان عن بعضهما تماما، ويكون مساوٍ للواحد الصحيح عندما يكون الارتباط تاما.
- تكون قيمته موجبة عندما يكون الارتباط بين المتغيرين طرديا ويكون قويا عندما يكون معامل الارتباط قريبا من الواحد الصحيح $(+1)$ وضعيفا عندما يكون قريبا من الصفر.
- تكون قيمته سالبة عندما يكون الارتباط بين المتغيرين عكسيا ويكون قويا عندما يكون المقدار السالب قريبا من (-1) ، وضعيفا عندما يكون المقدار السالب قريبا من الصفر.

انواع معاملات الارتباط

- **معامل ارتباط بيرسون**
- **معامل ارتباط سبيرمان**
- **معامل ارتباط كرامر**

معامل ارتباط بيرسون

لقياس درجة الارتباط بين **المتغيرات الكمية** والتي ليا توزيع طبيعي. ويمكن قياس قيمة معامل بيرسون $\mathrm{r_p}$ بين متغيرين بالعلاقة التالية:

$$\mathrm{r_p} = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}},$$

حيث:

- $\sum xy$: مجموع حاصل ضرب قيم x في قيم y؛
- $\sum x$: مجموع قيم المتغير x؛
- $\sum y$: مجموع قيم المتغير y؛
- $\sum x^2$: مجموع مربعات قيم المتغير x؛
- $\sum y^2$: مجموع مربعات قيم المتغير y؛
- n: هو عدد أزواج القيم لكل من x و y.

مثال: الجدول التالى يوضح ساعات الدراسة لمجموعة مكونة من ستة طلاب فى مادة الرياضيات (x) وعلامات الدراسة المحصل عليها في هذه المادة (y) :

x	2	3	4	5	6	7
y	8	8	10	12	14	16

هل هناك علاقة بين ساعات الدراسة والدرجات التي حصل عليها الطلبة؟

الحل:

$$r_p = \frac{(6 * 336) - (27 * 68)}{\sqrt{\left((6 * 139) - (27)^2\right)\left((6 * 824) - (68)^2\right)}}$$

$$= \frac{180}{\sqrt{(105)(320)}} = \frac{180}{\sqrt{33600}} = \frac{180}{183{,}3} = 0{,}98.$$

تحليل التباين

تمهيد

يعتبر أسلوب تحليل التباين أحد الأساليب الإحصائية التي تستخدم في كثير من المجالات التطبيقية في البحوث العلمية .ويرجع استخدام تحليل التباين إلى العالم الإحصائي فيشر، و يستخدم تحليل التباين لدراسة الفروق لاكثر من مجموعتين.

تعريف

تحليل التباين الاحادي يرمز له اختصار بـANOVA ، هو اختبار معلمي يستخدم للمقارنة بين المتوسطات أو التوصل إلى قرار يتعلق بوجود أو عدم وجود فروق بين متوسطات الأداء عند المجموعات.

أنواع تحليل التباين

يتوقف نوع تحليل التباين على عدد المتغيرات المستقلة.

- تحليل التباين الاحادي
- تحليل التباين الثنائي
- تحليل التباين الثلاثي

الإفتراضات

- تباين المجتمعات تكون متساوية
- ان تكون المجتمعات المسحوبة منها العينات متجانسة
- ان تكون البيانات مستقلة وتتبع التوزيع الطبيعي

ملاحظة

اختبار Levens يحدد ما اذا كان التباين متساويا ام لا.

الفرضيات

$$\begin{cases} \mathbf{H_0}: \ \mu_1 = \mu_2 = \mu_3 = \cdots \\ \mathbf{H_1}: \text{على الأقل اثنان غير متساويين} \end{cases}$$

الطريقة النظرية

يعتمد على خمس مراحل:

- حساب مجحوع المربعات الكلي
- حساب مجموع المربعات بين المجموعات
- حساب مجموع المربعات داخل المجموعات
- انشاء جدول تحليل التباين
- المقارنة واتخاذ القرار

الطريقة العملية

لتكن لدنا مجموعة من المجموعات (العينات)

- نرمز ب K عدد المجموعات.
- نرمز ب N عدد عناصر المجموعات الكلي

$$N = N_1 + N_2 + \cdots + N_k$$

- نرمز ب S_i مجموع عناصر المجموعة (العينة) رقم i

$$S_i = \sum x_i$$

- نرمز ب S_i^2 مجموع مربعات عناصر المجموعة (العينة) رقم i

$$S_i^2 = \sum x_i^2$$

مجحوع المربعات الكلي

$$S = \sum S_i^2 - \frac{(\sum S_i)^2}{N} = S_1^2 + S_2^2 + S_3^2 + \cdots - \frac{(S_1 + S_2 + S_3 + \cdots)^2}{N}$$

مجموع المربعات بين المجموعات

$$\beta = \sum \frac{[S_i]^2}{N_i} - \frac{(\sum S_i)^2}{N} = \frac{[S_1]^2}{N_1} + \frac{[S_2]^2}{N_2} + \frac{[S_3]^2}{N_3} + \cdots - \frac{(S_1 + S_2 + S_3 + \cdots)^2}{N}$$

مجموع المربعات داخل المجموعات

$$W = S - \beta$$

انشاء جدول تحليل التباين

مصدر التباين S.V	مجموع المربعات S.S	درجة الحرية D.f	متوسط المربعات S.M	القيمة الفائية المحسوبة F
بين المجموعات	β	$\mathbf{k-1}$	$\mathbf{R=\frac{\beta}{k-1}}$	$\mathbf{F=\frac{R}{H}}$
داخل المجموعات	$\mathbf{W}$	$\mathrm{N-k}$	$\mathbf{H=\frac{W}{N-k}}$	
الكلي	S	$\mathrm{N-1}$		

المقارنة واتخاذ القرار

المقارنة بين القيمة الفائية المحسوبة F والقيمة الجدولية $\mathbf{F}(\boldsymbol{\alpha}, \mathrm{N-k}, \mathbf{k-1})$

التمرين الأول

لتكن البيانات التالية

العينة الأولى	8	3	7	6	9	$N_1 = 5$
العينة الثانية	11	7	10	3	12	$N_2 = 5$
العينة الثالثة	7	2	3	1	7	$N_3 = 5$

هل هنالك فرق بين المتوسطالت الحسابية للعينات السابقة عند مستوى الدلالة 0,05؟

$N = N_1 + N_2 + N_3 = 15. \quad k = 3.$

$$S_1 = \sum x_i = 8 + 3 + 7 + 6 + 9 = 33.$$

$$S_2 = \sum x_i = 11 + 7 + 10 + 3 + 12 = 43.$$

$$S_3 = \sum x_i = 7 + 2 + 3 + 1 + 7 = 20.$$

$$S_1^2 = \sum x_i = 8^2 + 3^2 + 7^2 + 6^2 + 9^2 = 239$$

$$S_2^2 = \sum x_i^2 = 11^2 + 7^2 + 10^2 + 3^2 + 12^2 = 423.$$

$$S_3^2 = \sum x_i = 7^2 + 2^2 + +3^2 + 1^2 + 7^2 = 112.$$

الفرضيات

$$\begin{cases} \mathbf{H_0}: \mu_1 = \mu_2 = \mu_3 \\ \mathbf{H_1}: \text{على الأقل اثنان غير متساويين} \end{cases}$$

مستوى المعنوية

$\boldsymbol{\alpha = 0,05.}$

نوع الإختبار: المقارنة بين المتوسطات الحسابية (تحليل التباين في اتجاه واحد).

حساب احصاء الاختبار

حساب مجحوع المربعات الكلي

$$S = \sum S_i^2 - \frac{(\sum S_i)^2}{N} = S_1^2 + S_2^2 + S_3^2 - \frac{(S_1 + S_2 + S_3)^2}{N}$$

$$= 239 + 423 + 112 - \frac{(33 + 43 + 20)^2}{15}$$

$$= 774 - 614{,}4 = 159{,}6.$$

مجموعات مربعات بين المجموعات

$$\beta = \sum \frac{[S_i]^2}{N_i} - \frac{(\sum S_i)^2}{N} = \frac{(S_1)^2}{N_1} + \frac{(S_2)^2}{N_2} + \frac{(S_3)^2}{N_3} - - \frac{(S_1 + S_2 + S_3)^2}{N}$$

$$= \frac{(33)^2}{5} + \frac{(43)^2}{5} + \frac{(20)^2}{5} - \frac{(33 + 43 + 20)^2}{15}$$

$$= 667{,}6 - 614{,}4 = 53{,}2.$$

مجموع مربعات داخل المجموعات

$$W = S - \beta = 159{,}6 - 53{,}2 = 106{,}4.$$

انشاء جدول تحليل التباين

مصدر التباين S.V	مجموع المربعات S.S	درجة الحرية D.f	متوسط المربعات S.M	القيمة الفائية المحسوبة F
بين المجموعات	$\beta = 53{,}2$	$\mathbf{k - 1 = 2}$	$\mathbf{R = \frac{\beta}{k-1} = 26,6}$	$\mathbf{F = \frac{R}{H} = 3}$
داخل المجموعات	$\mathbf{w} = 106{,}4$	$N - k = 12$	$\mathbf{H = \frac{W}{N-k} = 8,86}$	
الكلي	$S = 159{,}6$	$N - 1 = 14$		

المقارنة واتخاذ القرار

بما أن قيمة F الحسابية أقل من الجدولية

$$\mathbf{F = 3 < F(0,05, 12, 2) = 3,88}$$

اذن نقبل الفرضية الصفرية (H_0) ومنه لا توجد فرق بين المتوسطات الحسابية للعينات (المجموعات) السابقة.

التمرين الثاني

تمثل البيانات التالية مجموعة من درجات الطلبة في احدى الامتحانات لإختبار أحد طرق التدريس

العينة الأولى	13	10	9	14	8	12	15	11	6	5	$N_1 = 10$
العينة الثانية	7	13	11	12	8	14	10	9	15	$N_2 = 9$	
العينة الثالثة	10	8	12	11	9	7	8	$N_3 = 7$			

هل هنالك فرق بين المتوسطات الحسابية للعينات السابقة عند مستوى الدلالة 0,05؟

$$N = N_1 + N_2 + N_3 = 26. \qquad k = 3.$$

$$S_1 = \sum x_i = 13 + 10 + 9 + 14 + 8 + 12 + 15 + 11 + 6 + 5 = 103.$$

$$S_2 = \sum x_i = 7 + 13 + 11 + 12 + 8 + 14 + 10 + 9 + 15 = 99$$

$$S_3 = \sum x_i = 10 + 8 + 12 + 11 + 9 + 7 + 8 = 65.$$

$$S_1^2 = \sum x_i = 13^2 + 10^2 + 9^2 + 14^2 + 8^2 + 12^2 + 15^2 + 11^2 + 6^2 + 5^2 = 1161.$$

$$S_2^2 = \sum x_i^2 = 7^2 + 13^2 + 11^2 + 12^2 + 8^2 + 14^2 + 10^2 + 9^2 + 15^2 = 1149.$$

$$S_3^2 = \sum x_i = 10^2 + 8^2 + 12^2 + 11^2 + 9^2 + 7^2 + 8^2 = 623.$$

<u>الفرضيات</u>

$$\begin{cases} \mathbf{H_0}: \ \mu_1 = \mu_2 = \mu_3 \\ \mathbf{H_1}: \text{على الأقل اثنان غير متساويين} \end{cases}$$

<u>مستوى المعنوية</u>

$$\alpha = 0,05.$$

<u>نوع الإختبار</u>: المقارنة بين المتوسطات الحسابية (تحليل التباين في اتجاه واحد).

<u>حساب احصاء الاختبار</u>

حساب مجموع المربعات الكلي

$$S = \sum S_i^2 - \frac{(\sum S_i)^2}{N} = S_1^2 + S_2^2 + S_3^2 - \frac{(S_1 + S_2 + S_3)^2}{N}$$

$$= 1161 + 1149 + 623 - \frac{(103 + 99 + 65)^2}{26}$$

$$= 2933 - 2741{,}88 = 191{,}12.$$

مجموعات مربعات بين المجموعات

$$\beta = \sum \frac{(S_i)^2}{N_i} - \frac{(\sum S_i)^2}{N} = \frac{(S_1)^2}{N_1} + \frac{(S_2)^2}{N_2} + \frac{(S_3)^2}{N_3} - \frac{(S_1 + S_2 + S_3)^2}{N}$$

$$= \frac{(103)^2}{10} + \frac{(99)^2}{9} + \frac{(65)^2}{7} - \frac{(103 + 99 + 65)^2}{26}$$

$$= 1060{,}9 + 1089 + 603{,}57 - 2741{,}88 = 11{,}59.$$

مجموع مربعات داخل المجموعات

$$W = S - \beta = 191{,}12 - 11{,}59 = 179{,}53.$$

انشاء جدول تحليل التباين

مصدر التباين S.V	مجموع المربعات S.S	درجة الحرية D.f	متوسط المربعات S.M	القيمة الفائية المحسوبة F
بين المجموعات	$\beta = 11{,}59$	$k - 1 = 2$	$R = \frac{\beta}{k - 1} = 5,79$	$F = \frac{R}{H} = 0,74$
داخل المجموعات	$W = 179{,}53$	$N - k = 23$	$H = \frac{W}{N - k} = 7,8$	
الكلي	$S = 191{,}12$	$N - 1 = 25$		

المقارنة واتخاذ القرار

بما أن قيمة F الحسابية أقل من الجدولية

$$F = 0,74 < F(0,05,23,2) = 3,422$$

اذن نقبل الفرضية الصفرية (H_0) ومنه لا توجد فرق بين المتوسطات الحسابية للعينات (المجموعات) السابقة.

<u>التمرين الثالث</u>

البيانات التالية تمثل زمن الشفاء (مقدرا بالأيام) لمجموعة من المرضى في احدى المستشفيات

العينة الأولى	13	10	9	14	11	12	15	$N_1 = 7$

العينة الثانية	11	6	9	8	7	5	10	$N_2 = 7$
العينة الثالثة	7	13	11	12	8	14	10	$N_3 = 7$
العينة الرابعة	10	8	7	11	9	7	8	$N_4 = 7$

هل هنالك فرق بين المتوسطالت الحسابية للعينات السابقة عند مستوى الدلالة 0,05؟

$$N = N_1 + N_2 + N_3 + N_4 = 28. \qquad k = 4.$$

$$S_1 = \sum x_i = 13 + 10 + 9 + 14 + 11 + 12 + 15 = 84.$$

$$S_2 = \sum x_i = 11 + 6 + 9 + 8 + 7 + 5 + 10 = 56.$$

$$S_3 = \sum x_i = 7 + 13 + 11 + 12 + 8 + 14 + 10 = 75.$$

$$S_4 = \sum x_i = 10 + 8 + 12 + 11 + 9 + 7 + 8 = 60.$$

$$S_1^2 = \sum x_i = 13^2 + 10^2 + 9^2 + 14^2 + 11^2 + 12^2 + 15^2 = 1036.$$

$$S_2^2 = \sum x_i^2 = 11^2 + 6^2 + 9^2 + 8^2 + 7^2 + 5^2 + 10^2 = 476.$$

$$S_3^2 = \sum x_i = 7^2 + 13^2 + +11^2 + 12^2 + 8^2 + 10^2 + 14^2 = 843.$$

$$S_4^2 = \sum x_i = 10^2 + 8^2 + 7^2 + 11^2 + 9^2 + 7^2 + 8^2 = 528.$$

الفرضيات

$$\begin{cases} \mathbf{H_0}: \boldsymbol{\mu_1 = \mu_2 = \mu_3 = \mu_4} \\ \mathbf{H_1}: \text{على الأقل اثنان غير متساويين} \end{cases}$$

مستوى المعنوية

$$\boldsymbol{\alpha = 0,05.}$$

نوع الإختبار: المقارنة بين المتوسطات الحسابية (تحليل التباين في اتجاه واحد).

حساب احصاء الاختبار

حساب مجحوع المربعات الكلي

$$S = \sum S_i^2 - \frac{(\sum S_i)^2}{N} = S_1^2 + S_2^2 + S_3^2 + S_4^2 - \frac{(S_1 + S_2 + S_3 + S_4)^2}{N}$$

$$= 1036 + 476 + 843 + 528 - \frac{(84 + 56 + 75 + 60)^2}{28}$$

$$= 2833 - 2700{,}89 = 182{,}11.$$

مجموعات مربعات بين المجموعات

$$\beta = \sum \frac{(S_i)^2}{N_i} - \frac{(\sum S_i)^2}{N} = \frac{(S_1)^2}{N_1} + \frac{(S_2)^2}{N_2} + \frac{(S_3)^2}{N_3} + \frac{(S_3)^2}{N_4} - - \frac{(S_1 + S_2 + S_3 + S_4)^2}{N}$$

$$= \frac{(84)^2}{7} + \frac{(56)^2}{7} + \frac{(75)^2}{7} + \frac{(60)^2}{7} - \frac{(84 + 56 + 75 + 60)^2}{28}$$

$$= 2773{,}85 - 2700{,}89 = 72{,}96.$$

مجموع مربعات داخل المجموعات

$$W = S - \beta = 182{,}11 - 72{,}96 = 109{,}15.$$

انشاء جدول تحليل التباين

مصدر التباين S.V	مجموع المربعات S.S	درجة الحرية D.f	متوسط المربعات S.M	القيمة الفائية المحسوبة F
بين المجموعات	$\beta = 72{,}96$	$\mathbf{k - 1 = 3}$	$\mathbf{R = \frac{\beta}{k - 1} = 24,25}$	$\mathbf{F = \frac{R}{H} = 5,34}$
داخل المجموعات	$\mathbf{w} = 109{,}15$	$N - k = 24$	$\mathbf{H = \frac{W}{N - k} = 4,54}$	
الكلي	$S = 182{,}11$	$N - 1 = 27$		

المقارنة واتخاذ القرار

بما أن قيمة $\mathbf{F}$ الحسابية أقل من الجدولية

$$\mathbf{F = 5,34 > F(0,05,24,3) = 3,098}$$

اذن نرفض الفرضية الصفرية $(\mathbf{H_0})$ ونقبل $(\mathbf{H_1})$ ومنه يوجد فرق بين المتوسطات الحسابية للعينات (المجموعات) السابقة.