

جامعة زيان عاشور - الجلفة -

كلية العلوم الإنسانية و الإجتماعية

قسم علم النفس و الفلسفة

الأستاذ : حربي سليم

محاضرات مقياس السنة الثالثة ليسانس علم النفس

المدرسي

مقياس : تحليل البيانات

المحاضرة (01)

مفاهيم عامة حول تحليل البيانات

تحليل البيانات أو المعطيات بالإنجليزية (**Data analysis**) هو عملية الفحص والتدقيق للبيانات، وتنظيمها لتكون أكثر دقة، واعداد تشكيلها، وتخزينها أيضا لنحصل ونستنبط في النهاية معلومات يمكن على أساسها اتخاذ وتحديد القرارات ، ولتحليل البيانات طرق عديدة تختلف باختلاف المجال المستخدمة فيه .

01 - البيانات :

هي مجموعة من الحروف أو الكلمات أو الأرقام أو الرموز أو الصور التي تكون خام وأولية، ومتعلقة بموضوع معين، والبيانات وحدات فردية للمعلومات، ويتم في بعض الأحيان تحويل البيانات إلى معلومات عندما يتم معالجتها في مرحلة ما بعد التحليل، ويتم استخدام البيانات في البحث العلمي والتمويل وإدارة الأعمال و الحوكمة وفي جميع أشكال النشاط التنظيمي البشري، ويتم قياس البيانات وجمعها والإبلاغ عنها وتحليلها من أجل استعمالها بشكل ناجح، ويجب تصحيح البيانات الخام لإزالة القيم المتطرفة أو أخطاء إدخال البيانات وتحدث معالجة البيانات بشكل شائع على مراحل، و يتم وصف البيانات بأنها اللفظ الجديد للاقتصاد الرقمي

02 - تحليل البيانات :

هي عملية يتم فيها أخذ البيانات وتجميعها من مصادر مختلفة، ثم إجراء معالجة وتقييم على تلك البيانات باستخدام تفكير تحليلي ومنطقي لكل مكون من تلك البيانات، لإنتاج رؤى ونتائج للأعمال ودعم عملية صنع واتخاذ القرار وتحديد الأولويات، ويمكن أن يعني تحليل البيانات أخذ الأرقام والمقاييس وغيرها من القياسات المختلفة ودراستها، وتحويلها إلى صورة ذات مغزى يمكن أن تستند إليها القرارات، وفي العديد من الحالات، يتضمن ذلك إخراج المخططات وعمل الرسوم البيانية وغيرها من أساليب التحليل لتحقيق الغايات منها وهذا التحليل هو مجرد خطوة من ضمن العديد من الخطوات التي يجب إكمالها عند إجراء تجربة بحثية، وهي جزء من عملية أكبر لاستخلاص ذكاء الأعمال، وفي إجابة سؤال: "ما هو تحليل البيانات"، تتضمن هذه العملية الخطوات الآتية:

2 - 1 - تحديد الأهداف : يجب أن تبدأ أي دراسة بمجموعة من أهداف العمل المحددة

بوضوح، حيث تعتمد القرارات المتخذة في بقية العملية على مدى وضوح أهداف الدراسة **2**

2 - 2 - طرح الأسئلة : هي عملية إجراء محاولة لطرح سؤال في مجال المشكلة .

2 - 3 - جمع البيانات : يجب جمع البيانات ذات الصلة بالسؤال من المصادر المناسبة .

2 - 4 - معالجة البيانات : حيث أنه يتم جمع البيانات الخام في عدة أشكال مختلفة، ويجب ترتيب البيانات التي تم جمعها وتحويلها بطريقة يمكن لأدوات تحليل البيانات من استيرادها واستعمالها .

2 - 5 - تحليل البيانات : تتيح لك عملية التحليل استكشاف البيانات، والعثور على أنماط فيها، وطرح أسئلة والإجابة عليها، وهذه هي العملية التي يتم بها استخدام المعطيات التي يتم جمعها في البحث عن طريق التطبيق الصحيح للأساليب الإحصائية .

2 - 6 - استخلاص النتائج و وضع التنبؤات : هي الخطوة التي يمكن من خلالها استخلاص النتائج من البيانات بعد التحليل الكافي لها، ويمكن بعدها صناعة قرار وتلخيص هذه الاستنتاجات والتنبؤات في تقرير وفق مبدأ البحوث الإستقرائية للخروج بقواعد عامة قابلة للتعميم

و منه نستخلص بأن عملية تفرغ البيانات و وتجميعها و

طريقة عرضها بغرض إستخدام الأساليب الإحصائية المناسبة

هي الخطوة الأساسية لعملية تحليل البيانات

مراجعة عامة حول مقياس الإحصاء الوصفي

أولاً : مقاييس النزعة المركزية

تتمثل هذه المقاييس في مدى تمركز النقاط حول المركز أي ميل البيانات الإحصائية إلى التمركز (التراكم) حول قيمة معينة وكلما إبتعدنا عنها فإن عدد المعلومات يبدأ بالتناقص و بالتالي سنركز من خلال هذه المراجعة على كل من المتوسط الحسابي و الوسيط و المنوال :

01 - المتوسط الحسابي :

هو قيمة تتجمع حولها قيم مجموعة ويمكن من خلالها الحكم على بقية قيم المجموعة فتكون هذه القيمة هي الوسط الحسابي .

يرمز له بـ : \bar{x} وهو القيمة التي إذا أعطيت لجميع مفردات الظاهرة المدروسة كان مجموع قيم المفردات يساوي مجموع القيم الأصلية.

أ/ في حالة البيانات غير المبوبة :

هي الحالة التي تكون فيها عدد القيم قليل ويحسب المتوسط الحسابي في هاته الحالة :

$$\bar{x} = \frac{\sum x}{n} = \frac{\text{مجموع القيم الفردية}}{\text{عدد الافراد}}$$

حيث \bar{x} المتوسط الحسابي، x : هو القيم المتحصل عليها، n : حجم العينة

مثال: اخذ طالب في السنة الثانية لقسم علم النفس و الفلسفة العلامات التالية :

$$\bar{x} = \frac{14+15+16+10+15}{5} = 14.15.16.10.15$$

ب/ في حالة البيانات المبوبة :

$$= \frac{\sum (xi.f)}{n} \bar{x}$$

\bar{x} : المتوسط الحسابي / xi : مركز الفئة / f : التكرار / n : حجم العينة

مثال2: اليك الفئات التالية:

الفئات	التكرار f	مركز الفئة xi	xi.f
60-70	5	65	325
70-80	15	75	1125
80-90	20	85	1700
90-100	30	95	2850
100-110	15	105	1575
110-120	10	115	1150
120-130	5	125	625
المجموع Σ	100		9350

$$\bar{x} = \frac{\Sigma(xi.f)}{n} = \frac{\Sigma(9350)}{100} = 93.5$$

الوسيط ME: Median

عبارة القيم التي تقع في منتصف القيم، أي القيم التي عدد القيم التي قبلها يكون مساوي العدد لعدد القيم التي بعدها بعد ترتيب القيم تصاعديا او تنازليا ويرمز له بالرمز Me

أ/ حالة البيانات غير المبوبة :

- ترتيب البيانات تصاعديا أو تنازليا حسب قيمها.
- اذا عدد المفردات n عدد فردي فان قيمة الوسيط تساوي $Me = \frac{n+1}{2}$
- اذا عدد المفردات n عدد زوجي فان قيمة الوسيط تساوي مجموع القيمتين المتوسطتين قسمة 2

مثال1: لدينا القيم التالية: 50.80.70.60.40.100.90

الحل: نرتب القيم تصاعديا: 40.50.60.70.80.90.100

الوسيط هو القيمة الوسطى 70 أو $Me = \frac{7+1}{2} = 4$

بمعنى هو الرتبة الرابعة من القيم الموجودة.

مثال2: لدينا القيم التالية: 100.40.50.80.60.70

الحل: نرتب القيم تصاعديا : 100.80.70.60.50.40

الوسيط هو مجموع القيمتين المتوسطتينقسمة 2 لاحظ 65.5
 $Me = \frac{60+70}{2} = 65.5$

ب/ حالة البيانات المبوبة:

$$Me = A + \frac{\frac{\sum f}{2} - Mn - 1.k}{E}$$

حيث:

$Me = L + \frac{C_1 - C_2}{C_3 - C_2} \cdot K$ L : بداية الفئة الوسيطة C ₁ : ترتيب الوسيط C ₂ : التكرار السابق C ₃ : التكرار اللاحق K : طول الفئة $C_1 = \frac{\sum f}{n}$	Me : الوسيط A : الحد الأدنى للفئة Mn-1 : تكرارات قبل الفئة الوسيطة K : طول الفئة تكرار الفئة الوسيطة
---	--

مثال: لدينا مجموعة البيانات التالية ونريد إيجاد الوسيط

الفئات	التكرار f	ت ت ص	ت ت ن
90-100	5	5	80
100-110	9	14	75
110-120	16	30	66
120-130	25	55	50
130-140	13	68	25
140-150	7	75	12
150-160	3	78	5
160-170	2	80	2
∑	80		

1/ نحدد التكرار التجميعي الصاعد

$$2/ \text{ نحدد ترتيب الوسيط} = \frac{\sum f}{n} = 40$$

3/ نحدد فئة الوسيط و هي 120-130

$$Me = 120 + \frac{(40-30)10}{25} = 120+4 = 124 = A + \frac{\frac{\sum f}{2} - Mn - 1.k}{E} \text{ : حساب الوسيط:}$$

المونال MO : Mode

أ/ حالة البيانات غير المبوبة: في هاته الحالة لا يستدعي البحث عن المونال أي عملية حسابية ويكفي أن ننظر إلى القيم أن عددها قليل و نستخرج القيمة التي تكررت أكثر من غيرها.

مثال: لدينا مجموع القيم التالية: 4.5.6.7.5.8.5.6

نلاحظ أن Mo يساوي 5 لأنه القيمة الأكثر تكرارا.

ب/ حالة البيانات المبوبة: نستخدم المعادلة التالية:

$$Mo = A + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot K$$

حيث : A : الحد الأدنى للفئة المونالية

Δ_1 : الفرق بين تكرار الفئة المونالية والفئة التي قبلها.

Δ_2 : الفرق بين تكرار الفئة المونالية والفئة التي بعدها.

K: طول الفئة المونالية.

أولاً: تحديد الفئة المونالية و هي التي تقابل أكبر تكرار عندما يكون طول الفئة ثابتا

أو الفئة التي تقابل أكبر تكرار معدل عندما تكون أطوال الفئات غير متساوية

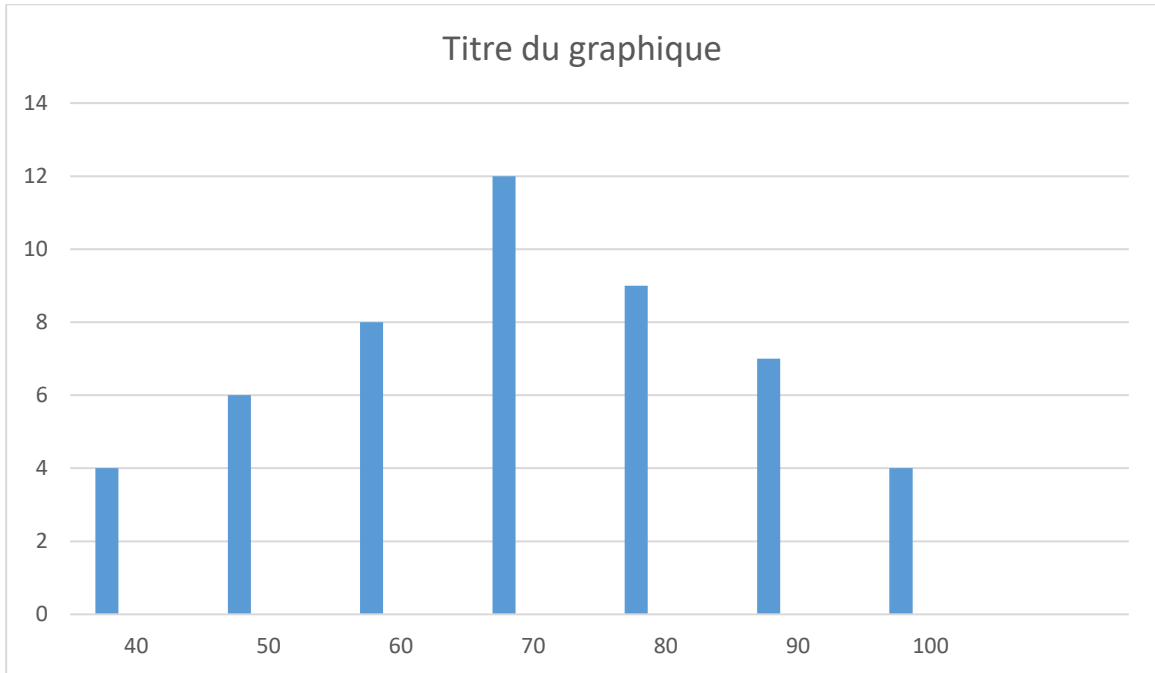
$$\text{تكرار معدل} = \frac{\text{تكرار فعلي} \cdot \text{طول الفئة المنتظمة}}{\text{طول الفئة غير المنتظمة}}$$

ثانياً: وضع مدرج تكراري قصد تحديد علامة المنوال

من عيوب المنوال انه غير دقيق لأنه يمكن وجود أكثر من منوال لنفس البيانات

الفئات	f
30-40	4
40-50	6
50-60	8
60-70	12
70-80	9
80-90	7
90-100	4
Σ	50

$$M_o = 60 + \frac{4}{4+3} \cdot 10 = 65.71$$



المحاضرة (02)

تابعة لحصة المراجعة العامة حول مقياس الإحصاء الوصفي

ثانياً : مقاييس التشتت و المكانة النسبية

- التشتت هو مدى تقارب أو تباعد البيانات عن بعضها البعض أي عن وسطها الحسابي بحيث
- كلما كانت البيانات قريبة من المتوسط الحسابي تكون بيانات متجانسة .
 - كلما كانت البيانات بعيدة عن المتوسط الحسابي تكون متباعدة ومتشعبة.

01 – المدى :

و هو الفرق بين أكبر قيمة و أقل في التوزيع الكلي للبيانات أي الفرق بين أكبر و أصغر قيمة للتوزيع الاحصائي ويرمز له بالرمز E وهو $E = H - L$

أ - في حالة البيانات غير المبوبة :

مثال: تبين السلسلات الإحصائية التاليتين توزيع اوزان مجموعتين من الرياضيين في مجموعتين مختلفتين حسب اوزانهم:

$$\text{مج 1: } 90.64.58.78.100 / \bar{x}_1 = 78$$

$$\text{مج 2: } 70.78.88.65.89 / \bar{x}_2 = 78$$

نلاحظ هنا ان \bar{x}_1 و \bar{x}_2 متساويين وهذا لا يعني انهما متشابهين وسنرى ذلك من خلال حساب المدى لكلا المجموعتين

$$E_1 = 100 - 58 = 42$$

$$E_2 = 89 - 65 = 24$$

وبالتالي فإن: $E_2 < E_1$

إذن فالتوزيع التكراري الأول أكثر تشتتاً من التوزيع الثاني ومنه فإن المجموعة الثانية أكثر تجانساً من المجموعة الأولى .

ب - فى حالة البيانات المبوبة :

وهو الفرق بين نهاية الفئة الأخيرة ناقص بداية الفئة الأولى

مثال: لدينا اوزان متباينة لمجموعة من المرضى النفسانيين الذين يزاولون العلاج بإحدى المؤسسات الخاصة و المطلوب البحث عن المدى رياضيين. اوجد المدى.

$$E = 100 - 50 = 50$$

الفئات
50-60
60-70
70-80
80-90
90-100

مزايا المدى: سهولة حسابه عددياً / يعطي فكرة سريعة عن التشتت

عيوب المدى: يتأثر بالقيم المتطرفة / يعتمد على حساب قيمتين / لا يمكن حسابه من الجداول التكرارية المفتوحة أحد الطرفين او كلاهما.

02 – التباين: La Variance

التباين لمجموعة من القيم هو مربع الفرق بين كل قيمة في التوزيع و متوسطها الحسابي

على حجم العينة و يرمز له بالرمز: S^2

1/ متوسط فروق انحرافات القيم عن وسطها الحسابي

2/ قيمة التباين لا بد أن تكون موجبة أو تساوي صفر.

3/ كلما اقتربت قيمة التباين من الصفر اقتربت قيمة الانحراف المعياري من الصفر كلما

أصبحت البيانات قريبة من التجانس .

أ - حالة البيانات غير المبوبة :

$$S^2 = \frac{\sum(x^2)}{n} - \bar{x}^2$$

يحسب بالمعادلة التالية:

حيث: x : هي القيم / \bar{x} : المتوسط الحسابي / n : عدد افراد العينة.

مثال: لدينا بعض أجور اللاعبين بالدولار عددهم 5 لاعبين

X	60	90	80	70	50	\sum 350
X ²	3600	8100	6400	4900	2500	\sum 25500

$$\bar{x} = \frac{350}{5} = 70$$

1/ حساب \bar{x} :

$$S^2 = \frac{25500}{5} - 4900 = 200$$

2/ حساب S^2 :

ب - حالة البيانات المبوبة :

$$S^2 = \frac{\sum(f.xi^2)}{\sum f} - \bar{x}^2$$

يحسب بالمعادلة التالية :

مثال: لدينا التوزيع التكراري التالي:

x	f	xi	xi ²	f.xi	f.xi ²
10-20	3	15	225	45	675
20-30	6	25	625	150	3750
30-40	10	35	1225	350	12250
40-50	15	45	2025	675	30375
50-60	8	55	3025	440	24200
60-70	5	65	4225	325	21125
70-80	3	75	5625	225	16875
\sum	50			2210	109250

$$\bar{x} = \frac{\sum(xi.f)}{n} = \frac{2210}{50} = 44.2$$

حساب \bar{x} :

$$S^2 = \frac{\sum(f.xi^2)}{\sum f} - \bar{x}^2 = 231.36$$

حساب S^2 :

03 – الانحراف المعياري -La déviation standard:

يتم إستنتاجه إنطلاقا من حساب التباين فهو الجذر التربيع له و يرمز له بالرمز S

مثال : نفس المثال السابق التباين تساوي 231.36 و منه الإنحراف المعياري :

$$\text{حساب } S = \sqrt{s^2}: 15.25$$

الماضرة (03)

01 – مفاهيم عامة حول الإحصاء الاستدلالي التطبيقي :

هو فرع من فروع الإحصاء يشمل كل الأساليب الإحصائية والنظريات القائمة عليها

وتطبيقاتها العملية المستخدمة لتحليل البيانات (المعلومات) التي نحصل عليها من العينة وذلك

للإستنتاج أو الإستدلال عن معالم و خواص المجتمع التي سحبت منه العينة وتكون هذه

الاستنتاجات على شكل تقديرات أو اختبارات للفروض و اتخاذ قرارات .

1-1/ المجتمع الإحصائي :

هو المجال العام لكل الملاحظات الممكن التعرف عليها وفق شروط محددة، كما يمكن تعريف المجتمع

العام على أنه كل وحدة تتوفر فيها الخصائص المدروسة مهما كان عددها كبير .

>> يعرف المجتمع الإحصائي بأنه مجموعة كل البيانات (القيم) الخاصة بالظاهرة محل الدراسة

والمجموعة من كل المفردات المقصودة بهذه الدراسة <<

والمفردات في أي دراسة إحصائية، قد تكون أشخاصا أو حيوانات أو أشياء جامدة، أو سنوات، أو أشياء اعتبارية كمنشآت أو جمعيات، وقد يكون المجتمع محدودا، أي نستطيع تحديد العدد الكلي لقيمه، أي العدد الكلي لمفرداته (عدد القيم هو نفسه عدد المفردات)، وقد يكون غير محدود (لا نهائي)، أي أن العدد الكلي لمفرداته كبير جدا لا يمكن تحديده أو حصره ويرمز له بـ N

ملاحظة : يمكن أن يكون المجتمع الإحصائي معدد أو غير محدد ، حقيقي أو نظري - مثال :

- عدد طلبة جامعة الجلفة

- عدد المؤسسات الاقتصادية في الوطن .

- عدد التلاميذ المراهقين في الثانوية .

1-2/ المعلمة :

>> هي أي مقياس إحصائي تحسب قيمته من بيانات المجتمع ككل بدون استثناء، ونستخدمه

لوصف المجتمع محل الدراسة وتحديد معالمه، وبالتالي يطلق عليه معلمة <<

(نجاه رشيد الكيخيا، 2007، صفحة 24).

ومن المعالم أي المقاييس التي تصف لنا المجتمع، هي مقاييس النزعة المركزية (المتوسط الحسابي،

الوسيط، المنوال)، أو مقاييس التشتت (التباين، الانحراف المعياري)، أو مقاييس الالتواء والتقلطح وأكبر

قيمة أو أصغر قيمة أو أي مقاييس إحصائية أخرى تحسب من المجتمع، والمعلمة عبارة عن قيمة ثابتة

لا تتغير، لأنها تحسب من المجتمع محل الدراسة، والمجتمع ثابت لا يتغير إثناء اجراء الدراسة، ولذلك يطلق على المعالم أحيانا الثوابت الإحصائية، وعادة تستخدم الحروف اليونانية للتعبير عن المعالم، فيرمز للوسط الحسابي للمجتمع μ (ميو) و تباين المجتمع بالرمز σ^2 (سيجما تربيع) وللانحراف المعياري للمجتمع بالرمز σ وهكذا ...

1-3/ العينة :

في الكثير من الدراسات لا يمكن للباحث أن يتناول كل وحدات المجتمع الإحصائي ، لهذا يلجأ إلى إختيار بعض الوحدات الممثلة له، فالعينة هي مجموعة صغيرة نسبيا من المجتمع العام و يشترط في تكوينها ما يلي :

- أن تعكس كل صفات المجتمع العام (التجانس) .
- أن يعطي لكل فرد من أفراد المجتمع العام نفس الفرصة للانتماء إليها قصد القضاء على عامل التحيز .
- أن تكون العينة كبيرة نسبيا، حيث تعكس صفات المجتمع العام المأخوذة منه .

في أي دراسة إحصائية، يجب أن يكون الهدف هو دراسة المجتمع ككل وليس دراسة العينة، ولكن نستخدم العينة لأننا في اغلب الدراسات لا نستطيع أن نجمع بيانات كل مفردات المجتمع محل الدراسة، وذلك للأسباب التالية:

1/ إذا كان حجم المجتمع محل الدراسة كبير جدا، وكانت إمكانات البحث المادية محدودة ولا تسمح له بجمع البيانات عن كل مفردة من مفردات المجتمع.

2/ إذا كان حجم المجتمع لا نهائي أي من المستحيل دراسته ككل .

3/ إذا كان المجتمع محل الدراسة متجانسا، أي أن جميع مفرداته تتمتع بنفس الخواص، ففي هذه الحالة نجد أن دراسة المجتمع ككل، هي مضيعة للجهد والمال والوقت، فمثلا اختبار قطعة من قماش متجانس تكفي لاختبار القماش كله.

1 - 4 - الفرضيات الإحصائية :

تضع لنا الفرضيات الإحصائية توقع لقيم بعض الإحصائيات المتعلقة بالمجتمع العام ، و ترتبط الفرضيات الإحصائية مباشرة بفرضيات البحث، حيث يسمح لنا قبولها أو رفضها التأكد من مدى تحققها ، و هي على نوعان من حيث الصياغة :

1 - 4 - 1 - الفرضية الصفرية H_0 :

نتوقع من خلالها عدم وجود فروق بين مجموعتين أو أكثر، أو عدم وجود إرتباط أو تأثير و و تأثر بين المتغيرات المدروسة

مثال : لا توجد فروق ذات دلالة إحصائية بين الذكور و الإناث في مستويات قلق الإمتحان .

1 - 4 - 2 - الفرضية البديلة H_1 :

إجابة وحل للفرضية الصفرية H_0 حيث يتوقع الباحث وجود فروق بين مجموعتين في حالة الاختبار بمخرجين ولصالح مجموعة معينة في حالة الاختبار بمخرج واحد.

* كل فرضية صفرية تقابلها فرضية بديلة واحدة والفرضيات البديلة الممكنة ثلاث:

$H_1 : \mu_1 \neq \mu_2$: فرضية بديلة بمخرجين او حدين (one-tailed test)

$H_1 : \mu_1 > \mu_2$: فرضية بديلة بمخرج واحد لصالح المجموعة الأولى (two-tailed test)

$H_1 : \mu_1 < \mu_2$:فرضية بديلة بمخرج واحد لصالح المجموعة الثانية (two-tailed test)

1 - 5 - القيمة الحرجة :

وهي القيمة التي تفصل بين منطقة الرفض ومنطقة القبول، و يتم إستخراجها إنطلاقا من القيمة
المجدولة التي هي نقطة تقاطع مستوى الدلالة مع درجة الحرية .

1 - 6 - دلالة الاختبار :

هي دلالة إحصائية تساعد الباحث على الخروج بنتائج واتخاذ قرار بقبول H_0 و رفض H_1 و رفض H_0
و قبول H_1 بمستوى خطأ مقبول هو عادة 5 أخطاء في المائة 0.05 او خطأ في المائة 0.01 او خطأ في
الالف 0.001 و هو المستوى الأكثر دقة في القياس.

1 - 7 - القرار الإحصائي :

يقسم مجال متغير دلالة الاختبار إلى مجالين (منطقتين) تسمى إحداهما بمنطقة الرفض والمنطقة الثانية
منطقة القبول. وبناءً على ذلك يكون القرار الإحصائي برفض الفرض الصفري إذا وقعت قيمة دلالة الاختبار
في منطقة الرفض ويكون عدم رفض الفرض الصفري إذا وقعت في منطقة القبول .

ملاحظة هامة : أولا - يتم قبول أو رفض الفرضية المصاغة من خلال المقارنة بين القيم

المحسوبة حسب كل إختبار و قيم مجدولة معدة مسبقا حسب كل أسلوب إحصائي حيث :

***** إذا كانت القيمة المحسوبة أكبر أو تساوي من القيمة المجدولة فإننا نرفض الفرضية الصفرية

و نقبل البديلة و نقول بأن النتيجة دالة أي رفض H_0 و قبول H_1 ***** .

ثانياً - يتم تحديد نوع الأسلوب الإحصائي الموافق للدراسة بناء على طبيعة صياغة الفرضية الإحصائية

و مستوى القياس الخاص بالبيانات المدروسة و طبيعة العينة المدروسة .

1 - 8 - مستويات القياس :

أ - المستوى الاسمي : يعبر فيه عن المتغير بصفات فهو نوعي ويساعد على التمييز فقط كالجنس ولون العينين...الخ. في هذا المستوى يمكن أن تعطي للصفات أرقاماً غير أن هذه الأرقام لا تسمح بإجراء عمليات حسابية عليها مثل أرقام الولايات (الجلفة 17) او قاعات التدريس ...الخ .

ب - المستوى الرتبي : يعبر فيه عن المتغير برتب بحيث ترتب القياسات تصاعدياً أو تنازلياً، في هذا المستوى تؤدي الأرقام دور التمييز لكنها أكثر دقة من المستوى الإسمي فهي تعطي فكرة عن موقع الفرد بالنسبة لباقي الأفراد، كترتيب الطلبة المتفوقين و الراسبين، المرتبة الأولى المرتبة الثانية...المرتبة الأخيرة

ج - مستوى المسافات المتساوية (الفترى) : يعبر فيه عن المتغير بقيمة عددية ويفترض أن المسافة بين القيمة والقيمة التي تليها متساوية، و نجد أن أغلب المتغيرات تقاس عند هذا المستوى، كما أن الصفر فيه قيمة غير حقيقية بل هي افتراضية فقط، أي انه لا يعبر عن غياب الظاهرة فمثلاً الطالب الذي يتحصل على درجة الصفر في مقياس الإحصاء لا يعني أن هذا الطالب ليست له معلومات عن وحدة الإحصاء المدرسة.

د - مستوى النسبة : هو أدق مستويات القياس ينطلق هذا المستوى من الصفر الحقيقي الذي يشير إلى إنعدام الظاهرة المدروسة، كغياب الكلور من الماء المقطر، حيث تستخدم في هذا المستوى والمستوى الذي سبقه كل العميلات الحسابية .

المحاضرة (04)

إختبار كاف تربيع χ^2

ترجع النشأة الأولى لإختبار كا2 إلى البحث الذي نشره كارل بيرسون في أوائل القرن العشرين وهي تعد من أهم اختبارات الدلالة الإحصائية وأكثرها شيوعاً لأنها لا تعتمد على شكل التوزيع ولذا فهي تعد من المقاييس اللابارامترية أي مقاييس التوزيعات الحرة ولأنها تحسب لكل خلية من خلايا أي جدول تكرارى ثم تجميع القيم الجزئية للحصول على القيمة الكارلية لكا2 .

وتستخدم كا2 لحساب دلالة فروق التكرار أو البيانات العددية التي يمكن تحويلها إلى تكرار مثل النسب والاحتمال .

* يستخدم إختبار كاف تربيع لإختبار الفروض والمعنوية للبيانات الاسمية , وهي أنواع منها:

1/ إختبار المعنوية للعينة الواحدة (كاي تربيع - لحسن المطابقة)

2/ إختبار المعنوية لأكثر من عينة (كاي تربيع - للاستقلالية)

أولاً: إختبار المعنوية للعينة الواحدة (كاف تربيع لحسن المطابقة)

يستخدم اختبار كاف تربيع لحسن المطابقة لإختبار هل النتائج المشاهدة تختلف عن النتائج المتوقعة .

لحسن المطابقة: χ^2 - شروط إجراء اختبار كاف تربيع

1- عدد مشاهدات العينة أكبر من 50 ($n > 50$)

2- التكرار المتوقع المناظر لكل فئة لا يقل عن 5 ($f_e < 5$)

ب - خطوات اختبار كاي لحسن المطابقة :

1- صياغة فرض العدم والفرض البديل:

* لا يوجد اختلاف بين النتائج المشاهدة والنتائج المتوقعة: H_0

* يوجد اختلاف بين النتائج المشاهدة والنتائج المتوقعة: H_1

2- قيمة إحصاء الاختبار كاف تربيع بعد تكوين جدول يساعدنا في حسابه على النحو التالي

الفئات	التكرارات المشاهدة	التكرارات المتوقعة	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
	f_o	f_e			
المجموع					$\sum \frac{(f_o - f_e)^2}{f_e}$

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

χ^2 : إحصاء الاختبار

ج- القيمة الجدولية لكاف تربيع :

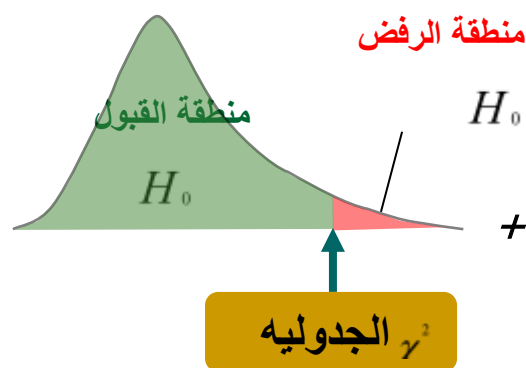
ودرجة الحرية من (عدد الفئات - 1) α نحدد مستوى المعنوية

نستخرج قيمة كاف تربيع الجدولية من خلال نقطة تقاطع مستوى الدلالة مع درجة الحرية $\chi^2(n-1, \alpha)$

د- اتخاذ القرار: نتخذ القرار بناءً على قيمة إحصاء إختبار كاف تربيع المحسوبة و مقارنتها بالمجدولة فإذا

كانت المحسوبة أكبر نرفض الصفرية ونقبل البديلة و العكس صحيح (نحدد منطقة الرفض و القبول على

الرسم التالي):



ونقبل الفرض البديل H_0 إذا وقعت قيمة إحصاء الاختبار في منطقة الرفض فإننا نرفض العدم H_1

أما إذا وقعت قيمة إحصاء الاختبار في منطقة القبول فإننا نقبل فرض العدم H_0

مثال : في دراسات سابقة عن المرضى النفسيين تم سؤالهم عن مستواهم الدراسي فكانت النتائج

كالتالي : 5% في المرحلة الجامعية / 15% في المرحلة الثانوية

30% في المرحلة المتوسطة / 50% في المرحلة الابتدائية

ولكن حالياً كانت النتائج ل 60 شخص كالتالي :

المرحلة الثانوية	عدد المرضى
جامعي	6
ثانوي	20
متوسط	10
ابتدائي	24
المجموع	60

هل يمكن أن نقرر إن نتائج برنامج هذا العام الفعلية تختلف عن البرامج السابقة؟ $\alpha = 0.05$

الحل:

لا يوجد اختلاف بين النتائج المشاهدة والنتائج المتوقعة: H_0

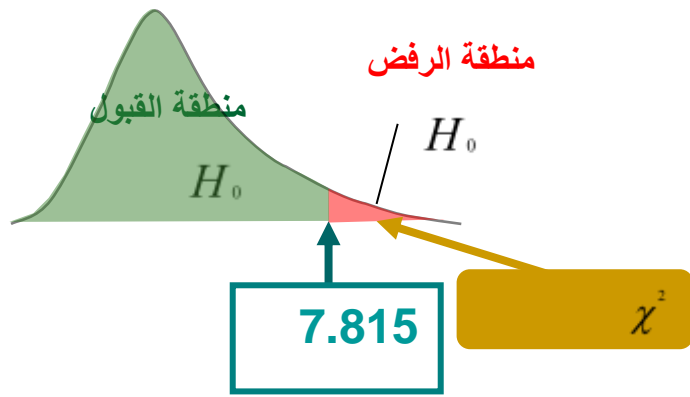
يوجد اختلاف بين النتائج المشاهدة والنتائج المتوقعة: H_1

نمط التغير	التكرارات f_o المشاهدة	النسبة	التكرارات f_e المتوقعة	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
جامعي	6	5%	$0.05 * 60 = 3$	3	9	3
ثانوي	20	15%	$0.15 * 60 = 11$	9	81	7.36
متوسط	10	30%	$0.30 * 60 = 18$	-8	64	3.55

ابتدائي	24	50%	$0.50 * 60 = 30$	-6	36	1.2
المجموع	60					10.72

قيمة إحصاء الاختبار $\chi^2 = 10.72$

3- الجدولية = قيمة $\chi^2 = 7.815$



4- وقع إحصاء الاختبار في منطقة الرفض

فإننا نرفض فرض العدم ونقبل الفرض البديل أي أن هناك اختلافا بين النتائج المشاهدة والنتائج المتوقعة

مثال 2: قامت وحدة محو الأمية بوزارة التعليم بتصميم برنامج دعائي يستهدف تحفيز ودفع غير

المتعلمين إلى تغيير اتجاهاتهم بحيث يصبحون أكثر إيمانا بفائدة التعليم و كانت نتائج البرامج السابقة في

هذا المجال كالآتي : * 23% يصبحون أكثر إيمانا بأهمية التعليم (تغيير إيجابي).

* 65% لا تتغير اتجاهاتهم (لا تغيير).

* 12% تتغير اتجاهاتهم بحيث يصبحون أكثر نفورا من التعليم (تغيير سلبي)

بالنسبة لهذا العام كانت نتائج البرنامج الذي اجري على 90 شخصا غير متعلم على النحو التالي:

نمط التغيير	عدد الأفراد
تغيير ايجابي	52
لا تغيير	34
تغيير سلبي 22	4
المجموع =	90

هل يمكن إن نقرر إن نتائج برنامج هذا العام الفعلية تختلف عن البرامج السابقة؟ $\alpha = 0.05$

الحل:

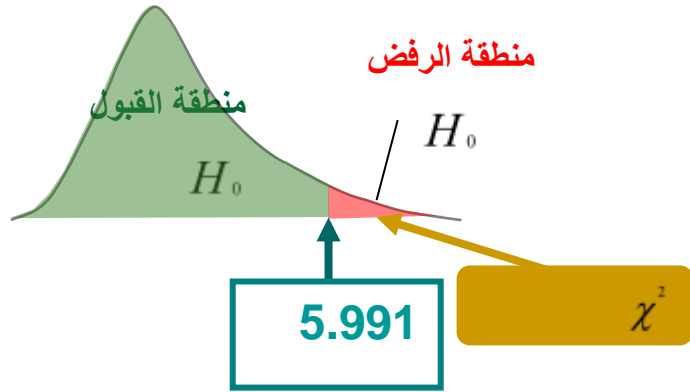
لا يوجد اختلاف بين النتائج المشاهدة والنتائج المتوقعة: H_0 :

يوجد اختلاف بين النتائج المشاهدة والنتائج المتوقعة: H_1 :

2 - قيمة إحصاء الاختبار $\chi^2 = 61.86$

نمط التغير	التكرارات f_o المشاهدة	النسبة النسبة	التكرارات f_e المتوقعة	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
تغير ايجابي	52	23 %	$0.23 \times 90 = 20.7$	31.3	979.69	47.32
لاتغير	34	65 %	$0.65 \times 90 = 58.5$	-24.5	600.25	10.26
تغير سلبي	4	12 %	$0.12 \times 90 = 10.8$	-6.8	46.24	4.28
المجموع	90					61.86

3- الجدوليه $\chi^2 = 5.991 = \chi^2(2, 0.05)$ قيمة



4- وقع إحصاء الاختبار في منطقة الرفض

فإننا نرفض فرض العدم ونقبل الفرض البديل أي أن هناك اختلافا بين النتائج المشاهدة والنتائج المتوقعة

المحاضرة (05)

ثانيا: اختبار كاف تربيع للإستقلالية

نحتاج في حالات كثيرة إلى التعرف عما إذا كانت هناك علاقة بين صفتين من صفات مجتمع ما فمثلاً قد نحتاج لمعرفة هل توجد علاقة بين مستوى الدخل والمستوى التعليمي؟ أو هل توجد علاقة بين لون العينين ولون الشعر في مجتمع ما؟ أو هل توجد علاقة بين المستوى التحصيلي ودخل الأسرة؟

يستخدم اختبار كاف تربيع للإستقلالية للإجابة على مثل هذه الأسئلة (هل توجد علاقة بين متغيرين إسميين أو متغير إسمي والآخر ترتيبى) ويعتمد على مقارنة القيم المشاهدة مع القيم المتوقعة، لذلك يجب أن نختار عينة عشوائية من المجتمع محل الدراسة ثم تصنف مشاهدات هذه العينة حسب مستويات كل صفة من الصفتين ووضعها في جدول يسمى جدول التوافق .

أولاً - خطوات اختبار كاف تربيع للإستقلالية :

1- صياغة فرض العدم والفرض البديل :

H_0 : لا يوجد علاقة بين الصفتين أو لا يوجد ارتباط بين الصفتين

H_1 : يوجد علاقة بين الصفتين أو لا يوجد ارتباط بين الصفتين

2- قيمة إحصاء اختبار كاف تربيع للإستقلالية المحسوبة :

إذا كان لكل من الصفتين A,B مستويان إثتان فقط , وكانت التكرارات المشاهدة هي a,b,c,d وذلك كما

يلي :

	B1	B2
A1	a	B
A2	c	D

ففي هذه الحالة تكون العلاقة الإحصائية للإختبار

$$\chi^2 = \frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

3- القيمة الجدولية لكاي تربيع :

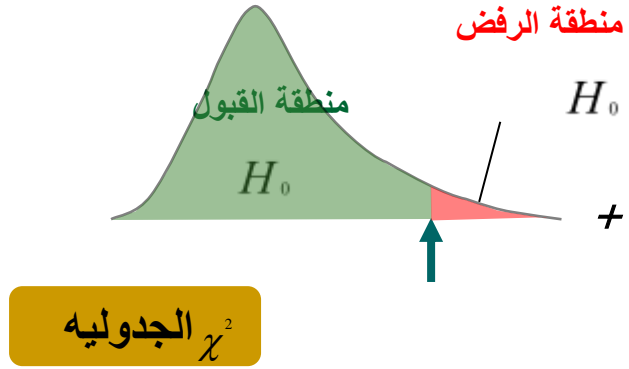
• نقوم أولاً بحساب درجة الحرية و هي (عدد الأعمدة - 1) (عدد الأسطر - 1)

و بنفس الطريقة أي تقاطع درجة الحرية مع مستوى الدلالة نستخرج القيمة المجدولة أي له تقريباً توزيع كاف

تربيع بدرجة حرية واحدة. $\chi^2(1, \alpha)$

4- اتخاذ القرار :

- نتخذ القرار بناءً على المقارنة بين القيمة المحسوبة و الجدولة وفق القاعدة العامة السابقة الذكر



ونقبل الفرض البديل H_0 إذا وقعت قيمة إحصاء الاختبار في منطقة الرفض فإننا نرفض فرض H_1

, أما إذا وقعت قيمة إحصاء الاختبار في منطقة القبول فإننا نقبل فرض العدم H_0

مثال : في بحث لدراسة العلاقة بين شرب الشاي و الجنس تم اختيار عينة حجمها 88 من المقيمين في

إحدى المدن وتم تصنيفهم في الجدول الآتي . هل تدل هذه البيانات على وجود علاقة بين شرب الشاي نوع

الجنس؟ ، استخدم مستوى معنوية $\alpha=0.05$

	ذكور	إناث	المجموع
يشربون الشاي	40	33	73
لا يشربون الشاي	3	12	15
المجموع	43	45	88

الحل :

H_0 : لا توجد علاقة بين شرب الشاي ونوع الجنس.

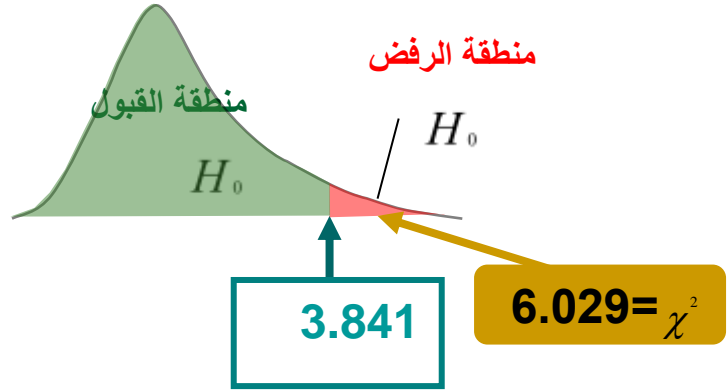
H_1 : توجد علاقة بين شرب الشاي ونوع الجنس.

وتكون قيمة إحصاء الاختبار هي :

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} = \frac{88(480-99)^2}{73 \times 15 \times 43 \times 45} = 6.029$$

ونحصل على القيمة الحرجة من جدول توزيع كاف تربيع فنجدها :

$$\chi^2(1,0.05) = 3.841$$



وقيمة إحصاء الاختبار أكبر من القيمة الجدوليه , أي أنها تقع في منطقة الرفض وبالتالي فإننا نرفض H_0

ونقبل H_1 وهو أن هناك علاقة بين شرب الشاي والنوع.

مثال: أجري بحث اجتماعي لدراسة العلاقة بين الجنس والاتجاه للزواج من الاقارب أخذت عينة من 57

فردا وكانت النتائج على النحو التالي :

الجنس	ذكر	أنثى	المجموع
الاتجاه للزواج من الاقارب			
مؤيد	10	15	25
غير مؤيد	20	12	32

المجموع	30	27	57
---------	----	----	----

• هل هناك ارتباط أو علاقة بين الجنس والاتجاه للزواج من الأقارب أم أن الصفتين مستقلتان عن بعضهما البعض أي لا علاقة بين الجنس والاتجاه للزواج من الأقارب بمستوى معنوية 0.05 ؟

الحل :

H_0 : لا توجد علاقة بين الاتجاه للزواج من الأقارب ونوع الجنس.

H_1 : توجد علاقة بين الاتجاه للزواج من الأقارب ونوع الجنس.

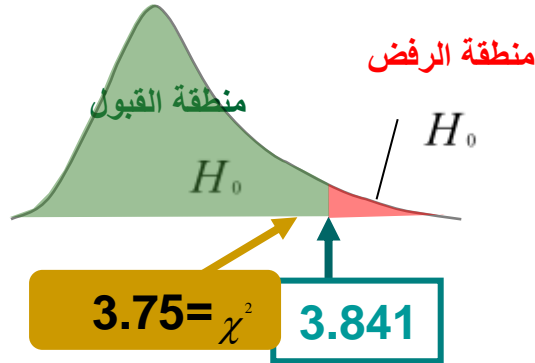
وتكون قيمة إحصاء الاختبار هي :

$$\chi^2 = \frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)} = \frac{75(120 - 300)^2}{25 \times 32 \times 30 \times 27} = \frac{75 \times (-180)^2}{648000}$$

$$= \frac{75 \times 32400}{2430000} = 3.75$$

ونحصل على القيمة الحرجة من جدول توزيع كاف تربيع فنجدها :

$$\chi^2(1,0.05) = 3.841$$



وقيمة إحصاء الاختبار أصغر من القيمة الجدولية ، أي أنها تقع في منطقة القبول وبالتالي فإننا نقبل H_0

وهو أنه ليس هناك علاقة بين الاتجاه للزواج من الأقارب و الجنس .

المحاضرة (06)

الارتباط R

1- تعريف الارتباط :

هو أسلوب إحصائي لقياس مدى قوة العلاقة بين متغيرين أو أكثر و هو محصور بين $1+$ و $1-$ بحيث يشير الرقم 1 إلى وجود علاقة إيجابية (correlation Positive) قوية بين المتغيرين، وذلك يعني أن حدوث أي تغير إيجابي في أحد المتغيرات، سيتبعه تغير إيجابي في المتغير الآخر. ويشير الرقم $1-$ إلى وجود علاقة سلبية (Negative correlation) قوية بين المتغيرين، مما يعني أن حدوث أي تغير إيجابي في أحد المتغيرات، سيتبعه تغير سلبي في المتغير الآخر، بينما يشير الصفر إلى عدم وجود علاقة على الإطلاق بين المتغيرين، أو عدم وجود علاقة واضحة بين التغيرات أو التقلبات

2- خصائص معاملات الارتباط :

- يستخدم معامل الارتباط للتعرف على طبيعة العلاقة بين متغيرين أو أكثر (طردي/عكسي)، فعندما يلاحظ تغير في المتغير X يتبعه تغير في المتغير Y
- يستخدم في اختبار صحة الفروض الارتباطية (العلائقية) سواء كانت فرضيات صفرية أو فرضية بديلة موجهة أو غير موجهة.
- قيمة معامل الارتباط إلى درجة العلاقة بين المتغيرات المستقلة والمتغيرات التابعة وليس تفسير هذه الظاهرة.

- عندما يكون معامل الارتباط مرتفع بين متغيرين $(X.Y)$ لا يعني أن المتغير (X) سبب وجود (Y) أو العكس.

ملاحظة: الخطأ الشائع الذي يقع فيه الباحثون هو تفسير معاملات الارتباط على علاقات سببية (علاقة العلة بالمعلول).

* يجب أن تكون العلاقة منطقية فمن الممكن أن تكون هناك علاقة بين الطول والوزن، المسافة والزمن، القوة والسرعة، المراجعة والتحصيل الدراسي، القلق والثبات الانفعالي، ولكن هل من المعقول أن تكون هناك علاقة بين طول أصابع القدمين والذكاء؟ أو هل هناك علاقة بين الطول والتحصيل النظري في التربية البدنية؟ ان مثل هذه العلاقة غير منطقية لأنه لا يمكن تفسير مثل هذا النوع من العلاقات واغلبها منعدم الارتباط.

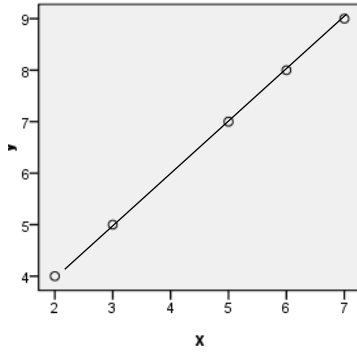
* يشترط في تطبيق معامل الارتباط البسيط بين متغيرين أن تكون العلاقة بين $(X.Y)$ خطية أي أن كل زيادة في (X) تصحبها زيادة في (Y) أو أن كل تناقص في (X) يصاحبه تناقص في (Y) أو الزيادة في (X) تتبعها نقص في (Y) أو العكس.

3 - لوحة الانتشار:

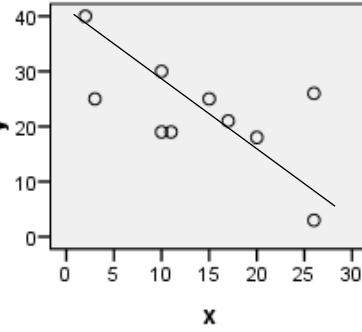
للتأكد من العلاقة بين متغيرين خطية يمكن رسم لوحة الانتشار Scatter Diagramme، تمثل هذه اللوحة المسافة الموجودة بين المحورين الممثلين لدرجة المتغيرين سحابة من النقاط فإذا حصلنا على سحابة تشكل خطا مستقيما ذا اتجاه واحد نقول ان العلاقة بين المتغيرين علاقة خطية .

ملاحظة : تقدير القوة بين متغيرين بفضل لوحة الانتشار:

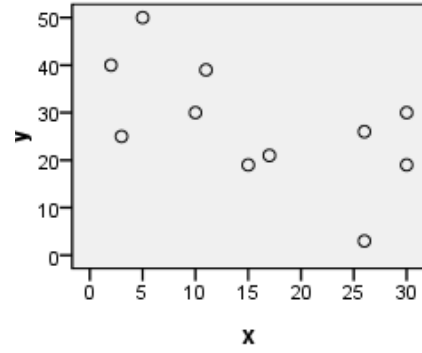
تتراوح قيمة معامل الارتباط بين 1 و -1 مرورا بالصفر.



موجبة قوية (يوجد ارتباط)



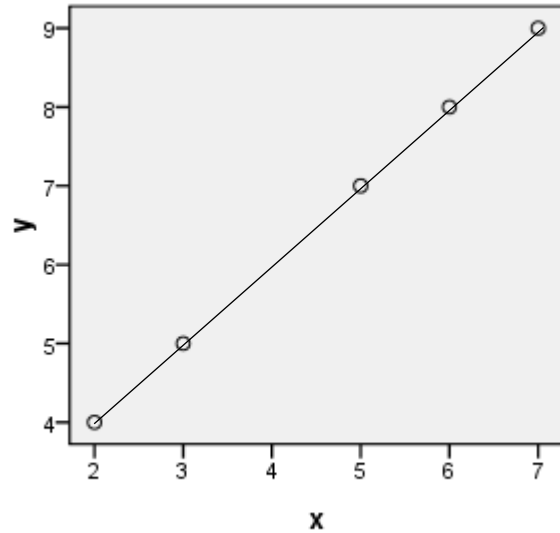
سالبة قوية (يوجد ارتباط خطي)



عدم وجود ارتباط

مثال: تحديد طبيعة وقوة الارتباط بين القلق والتحصيل الدراسي بإفترض أن لدينا البيانات التالية لـ 5 أفراد:

التحصيل y	القلق x	n
5	3	1
4	2	2
9	7	3
7	5	4
8	6	5



ملاحظة هامة : سنحاول في هذا السداسي التركيز على معامل الارتباط الخاص بقياس العلاقة بين الدرجات **بيرسون** و معامل الارتباط الرتبي **سبيرمان** .

04 - معامل الارتباط بيرسون Pearson :

هو معامل خاص بقياس الارتباط بين متغيرين و تكون البيانات المقاسة على شكل درجات و يرمز له بالرمز **R_p** و يتميز بالخصائص التالية :

- يستخدم في البيانات الكمية.
- هو معامل يوجد ضمن مستوى المسافات المتساوية والنسبية
- هو اختبار بارامتري(معلمي) يعطى وفق المعادلة التالية:
-

$$R_p = \frac{n\sum(x.y) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}}$$

حيث: * **R_p** معامل الارتباط

- **n** حجم العينة
- **x.y** متغيران

مثال: لمعرفة العلاقة بين الذكاء الوجداني و تقدير الذات عند مستوى الدلالة 0.05 طبق المقياسين على

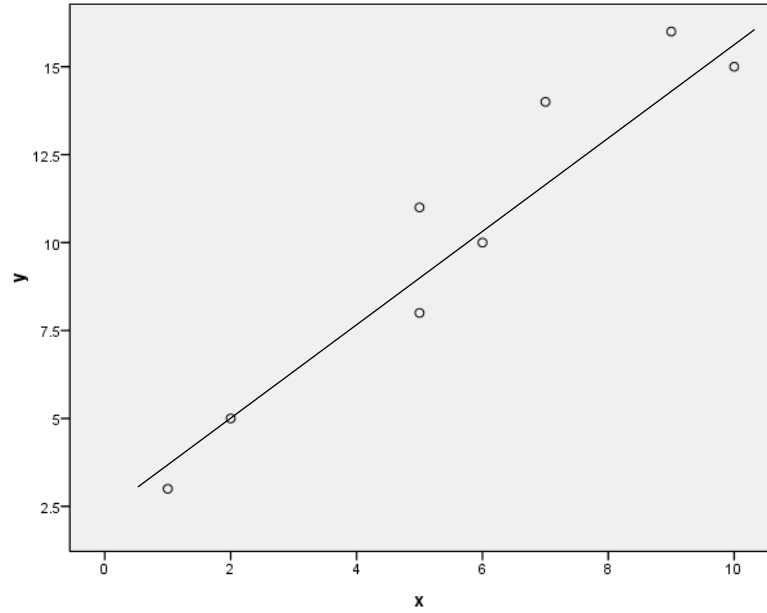
عينة قوامها 08 طلبة و كانت النتائج كما هي موضحة في الجدول التالي :

الذكاء الوجداني (x)	تقدير الذات (y)
10	15
2	5
5	11
6	10

7	14
1	3
9	16
5	8
\sum 45	82

الحل:

الدكاء الوجداني (x)	تقدير الذات (y)	x.y	x^2	y^2
10	15	150	100	225
2	5	10	4	25
5	11	55	25	121
6	10	60	36	100
7	14	98	49	196
1	3	3	1	9
9	16	144	81	256
5	8	40	25	64
\sum 45	82	560	321	996



1/ صياغة الفرضيات : لا يوجد
 H_0 : ارتباط بين X و Y

يوجد

H_1 : ارتباط بين X و Y

2/ نوع البيانات كمية

3/ تحديد نوع الإختبار المناسب : بما أننا بصدد دراسة العلاقة بين متغيرين و البيانات جاءت على شكل درجات فإن إختبار معامل الارتباط بيرسون هو الأسلوب الإحصائي المناسب :

4/ الإجراء الحسابي : $R = 0.96$ و هذا ما يدل على أن العلاقة طردية موجبة قوية جدا

$$R_p = \frac{n\sum(x.y) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}} = \frac{8.560 - 45.82}{\sqrt{[8.321 - 2025][8.996 - 6724]}} = 0.96$$

5/ درجة الحرية $df : 8 - 2 = 6$

6 / مستوى الدلالة $\alpha : 0.05$

7/ القيمة الجدولة هي $R_t = 0.70$

8/ القرار الإحصائي : بما أن القيمة المحسوبة أكبر من الجدولة أي $R_t 0.7 > R_c 0.96$ نرفض الفرض الصفري و نقبل البديلة و نقول بأنه توجد علاقة ذات دلالة إحصائية بين الذكاء الوجداني و تقدير الذات لدى الطلبة

خطوات اختبار الفرضيات الارتباطية:

- 1/ صياغة الفرض الصفري مقابل الفرض البديل.
- 2/ معرفة نوعية البيانات (كمية. كيفية)
- 3/ تحديد نوع الاختبار (بيرسون، سبيرمان).
- 4/ رسم لوحة الانتشار بناء على البيانات المعطاة.
- 5/ حساب معامل الارتباط (بيرسون، سبيرمان) بناء على علاقة الارتباط.
- 6/ حساب درجة الحرية **df**:

- اختبار بيرسون $df = n-2$

- اختبار سبيرمان $df = n-1$

7/ تحديد مستوى الدلالة.

8/ تحديد قيمة معامل الارتباط (R) مع الجدولية (R_t) و فق جدول الارتباط المختار وهذا بتحديد نقطة تقاطع df مع α

9/ اتخاذ القرار بقبول او رفض H_0 :

- إذا كانت $R_c > R_t$ نرفض الفرض الصفري التي تقول بعدم وجود ارتباط.

- إذا كانت $R_c < R_t$ نقبل الفرض الصفري التي قول بوجود ارتباط.

المحاضرة (07)

05 - معامل الارتباط الرتبي سبيرمان Spearman :

أحيانا تكون بيانات الظاهرتين أو إحداهما بيانات غير كمية لكنها ذات طبيعة ترتيبية مثل تقديرات الطلاب في اختبار معين (A.B.C...) أو تكون البيانات كمية لا تتوفر فيها بعض الخصائص المطلوبة فنلجأ حينئذ لإستبدال قيم البيانات بترتيبها ونستخدم ما يسمى بمعامل ارتباط الرتب سبيرمان Spearman حيث يرمز له R_s ويكون حسابه من خلال الخطوات التالية :

1/ نرتب بيانات الظاهرتين في موقعيهما حسب الترتيب التصاعدي ونسمي هذه رتب القيم

$$EX : A^8.A^7.B^6.B^5.B^4.C^3.C^2.F^1$$

2/ نحسب فروق الرتب ومجموع مربعاتها فيكون معامل ارتباط الرتب: تعطى علاقة معامل الإرتباط سبيرمان كما يلي :

$$R_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

حيث: R_s : معامل ارتباط الرتب

d^2 : مربع الفروق بين رتب نفس الفرد على المتغيرين X.Y

n : عدد افراد العينة

1 و 6 : ثابتان لا يتغيران

ملاحظة :

- تعطى الرتبة 1 إلى أضعف قيمة .
- إذا وجدت مفردتان أو أكثر لهما نفس القيمة فإننا رتبتهن تكون متوسط الرتب التي سيأخذونها لو لم تكن لهم نفس القيمة.
- لمعامل سبيرمان نفس الخواص السابقة لمعامل بيرسون.

- درجة الحرية : $df = n-1$

مثال :

البيانات التالية توضح تقدير عينة من 08 طلبة تخصص علم النفس و المراد هو معرفة هل توجد علاقة ذات دلالة إحصائية بين رتب الطلبة في المواد العلمية (X) و المواد الأدبية (Y) و كانت النتائج كما هي موضحة في الجدول التالي :

x	y
A	80
F	90
B	60
B	60
C	80
C	70
A	90
B	60
8	

الحل:

x	y	رتبة x	رتبة y	d	d ²
A	80	7.5	5.5	2	4

F	90	1	7.5	-6.5	42.25
B	60	5	2	3	9
B	60	5	2	3	9
C	80	2.5	5.5	-3	9
C	70	2.5	4	1.5	2.25
A	90	7.5	7.5	0	0
B	60	5	2	3	9
8				0	84.5

$$R_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6.84.5}{8(8^2-1)} = -0.006$$

ومنه الارتباط عكسي ضعيف جدا

1/ تحويل البيانات الكمية الى رتب: سبق وذكرنا أن من شروط استخدام معامل ارتباط سبيرمان أنه لا توجد

تكرارات كثيرة في الرتب، من هذه الملاحظة تعترضنا حالتين:

1-1/ ان لا توجد تكرارات في الرتب: في تحويل البيانات الكمية الى رتب تعطى الرتبة 01 الى أضعف

القيم الكمية ومنتصاعد في ترتيب القيم الكمية حتى نصل الى اعلى درجة كمية في الترتيب.

البيانات الكمية للمتغير x : 16 — 14 — 12 — 8 — 10 — 19

ترتيب قيم المتغير x : 5 — 4 — 3 — 1 — 2 — 6

1-2/ حالة البيانات المتكررة: في حالة تكرار مجموعة من القيم فإننا نحسب المتوسط الحسابي لرتب هذه

القيم

البيانات الكمية x : 20 — 19 — 12 — 12 — 8 — 8 — 4

رتب : x 1—3—3—3—5.5 —5.5— 7—8

$$\frac{2+3+4}{3} = 3\frac{5+6}{2} = 5.5$$

6- الانحدار (التنبؤ) :

1-6/ تعريف: يتمثل التنبؤ في تقدير قيمة متغير (x) اعتمادا على نتائج متغير ثاني (y) له علاقة بالمتغير الأول.

- يتم التنبؤ على أساس وجود ارتباط بين المتغير المتنبأ به والمتغير الآخر.
- ترتفع قيمة التنبؤ كلما زادت قوة الارتباط بين المتغيرين.
- يتم التنبؤ من خلال معادلة رياضية تربط بين متغيرين تعرف باسم معادلة الانحدار معطلة بالخطأ المعياري للتنبؤ.

2-6/ معادلة خط الانحدار:

$$y = A + B(x)$$

مثل التنبؤ للمتغير y

حيث A, B تسمى ثوابت التنبؤ

A : هي نقطة تقاطع الخط الموجود في سحابة الانتشار و الذي يمر بجميع النقاط مع محور الترتيب (Y)

$$A = \bar{y} - B(\bar{x})$$

الجزء المقطوع من محور (Y) ويحسب من المعادلة

B مدى ارتفاع الخط البياني (ميل خط الانحدار) في كل مرة تزيد فيها وحدة للمتغيرين X,y ويحسب

$$B = \frac{sy}{sx} \cdot r \text{ بالمعادلة}$$

لحل معادلة الانحدار نحسب أولاً الثابت B ولحساب هذا المعامل نحتاج الى حساب الانحراف المعياري للمتغير X أو SX وكذلك الانحراف المعياري Y أي SY و R معامل الارتباط ، ثم نحسب معادلة التنبؤ بالخطأ المعياري للتنبؤ لان المعادلة مبنية على انحراف النقاط على الخط الموجود في وسط سحابة الانتشار .

$$\text{يعطى الخطأ المعياري للتنبؤ بالمعادلة: } sxy = sy - \sqrt{1 - r^2}$$

حيث sxy : الخطأ المعياري للتنبؤ

sy : الانحراف المعياري لـ y

r : معامل الارتباط

مثال: تحصل محمد على العلامة 14 في مقياس الإحصاء الاستدلالي ولم يجز امتحان مادة spss كم تكون علامته في هذا المقياس علماً أن : $r = 0.90$ / $\bar{y} = 12$ / $\bar{x} = 10$ / $sy = 2$ / $sx = 2$

الحل:

$$y^* = y \pm syx \quad \text{أولاً: نكتب معادلة التنبؤ بالخطأ المعياري}$$

$$r = \frac{2}{2} \cdot 0.90 = 0.90 \quad \text{ثانياً: نحسب الثابت } B = \frac{sy}{sx} \cdot r$$

$$A = \bar{y} - B(\bar{x}) = 12 - 0.90(10) = 3 \quad \text{ثالثاً: نحسب الثابت}$$

$$Y = A + B(x) = 3 + 0.90(14) = 15.6 \quad \text{و بالتعويض في معادلة الانحدار نجد:}$$

ومنه يأخذ الطالب في مقياس spss علامة 15.6

رابعاً: حساب الخطأ المعياري للتنبؤ حيث يمثل الخطأ المحتمل الذي يمكن ارتكابه في تقدير علامة الطالب ويحسب $sxy = sy - \sqrt{1 - r^2}$

$$y^* = 15.6 \pm 0.86 \quad \text{ومن معادلة التنبؤ مصححة بالخطأ المعياري :}$$

الخطأ المعياري SYX معادلة التنبؤ y