922021

القسم	السنة	الفصل
ارشاد نفسي	الثانية	الثاني

الإحصاء التحليلي

مقرر

تحليل الارتباط والانحدار

Correlation Regression Analysis

– الغاية من تحليل الارتباط والانحدار.

– تحليل الارتباط .

الارتباط البسيط للمتغيرات الكمية .

الارتباط للمتغيرات النوعية .

– تحليل الانحدار

الانحدار الخطي البسيط .

إيجاد معادلة مستقيم الانحدار.

أهمية معادلة مستقيم الانحدار .

دراسة التباينات.

التقدير والتنبؤ.

8-1 الغاية من تحليل الارتباط والانحدار

تعرفنا في الفصول السابقة على بعض أساليب التحليل الإحصائي، التي يمكن استخدامها لتحليل متغير واحد مثل أطوال مجموعة من الأشخاص، أو أوزانهم أو دخولهم.... الخ، فرأينا كيف يمكن حساب النـزعة المركزية والتشتت واحتمال أن يأخذ المتغير قيمة ما أو تكون قيمته في مجال ما .

سوف نتعرف في هذا الفصل على أساليب تحليل جديدة نستطيع باستخدامها تحليل متغير ما، من خلال علاقته بمتغير آخر أو بعدة متغيرات، حيث نستطيع أن نذكر عدداً كبيراً من الأمثلة على الظواهر أو المتغيرات التي يوجد علاقة بينها فمثلاً: دخل الأسرة وإنفاقها، دخل الأسرة ومدخراتها، الغلة الزراعية وكمية السماد المستخدم، لون بشرة الإنسان وعدد شعر الرأس، .... الخ. ومن خلال دراسة وتحليل متغير ما بعلاقته مع متغير آخر، أو مجموعة من المتغيرات يمكننا التأكد من وجود هذه العلاقة وتحديد متانتها بين المتغيرات، وهذا الأسلوب يسمى بتحليل الارتباط (Correlation analysis)، كما نستطيع معرفة تأثير أحد المتغيرين على متغير آخر، أو تأثير عدة متغيرات على متغير واحد من خلال دراسة الانحدار (Regression).

إن تحليل الارتباط والانحدار يتم من خلال إجراء عدد من الحسابات، يسبقها التأكد من وجود علاقة بين المتغيرات عن طريق التحليل المنطقي بالاعتماد على معارفنا

العامة كما يمكن الاستعانة بأصحاب الخبرات للتأكد من منطقية العلاقة بين المتغيرات، وإلا سوف نحصل على مؤشرات تؤكد وجود العلاقة وهي غير موجودة منطقياً، وهذا ما يدعى بالارتباط الزائف، مثلاً تزايد المبيعات السنوية لمتجرين مختلفين من مدينتين مختلفتين ، والتأكيد المنطقي للعلاقة يتطلب بالإضافة إلى وجود العلاقة أن تكون الفترة الزمنية للعلاقة منطقية مثلاً: الغلة وكمية السماد الكيماوي للعام نفسه، أما الغلة وكمية السماد البلدي تكون الفترة الزمنية أكثر من عام، كما يجب ألا يكون هناك فاصل مكاني يلغي منطقية العلاقة حيث نعلم أن هناك علاقة بين دخل الأسرة ونفقاتها، لكن إذا أخذنا دخل عدة أسر في دمشق وإنفاق عدة أسر في عمان، فإن الفاصل المكاني ينفي هذه العلاقة. ونشير إلى أن أهمية تأكيد وجود العلاقة المنطقية لا يعني بالضرورة وجود علاقة سببية بين المتغيرين، أي أن أحدهما يؤثر بالآخر، فقد يكون أحدهما يؤثر بالآخر من خلال متغير آخر مثلاً: علاقة الزمن بتزايد عدد السكان في بلد ما، فالزمن كمتغير يعكس تأثير متغيرات أخرى كثيرة، ومن الممكن أن تكون كلتا الظاهرتين تتحركان بسبب متغيرات أخرى تؤثر عليها مثلاً: تطور عدد الحيوانات الداجنة في بلد ما، وإنتاجها النباتي فإنه يؤثر على كليهما عدد كبير من المتغيرات الأخرى.

بعد التأكد من وجود العلاقة بين الظواهر (المتغيرات) يجب تحديد أي المتغيرات يؤثر بالآخر، وهو ما سوف ندعوه بالمتغير المستقل (Independent variable) ونرمز له بالرمز x ، أما المتغير الذي يتأثر بالآخر ندعوه بالمتغير التابع (Dependent variable) ونرمز له بالرمز y، وإن تحديد طبيعة المتغيرات هو امتداد للخطوة السابقة، التأكد من وجود العلاقة، حيث يتم تحديد طبيعة المتغيرات في أكثر الأحيان من خلال التحليل المنطقي، فالدخل يؤثر باستهلاك الأسرة، فهو المستقل، والاستهلاك تابع والسماد يؤثر في الغلة فهو المستقل والغلة تابع... أما إذا تعذر تحديد طبيعة المتغيرات بالطريقة السابقة فعندها من المنطق القول إن من يحدث أولاً هو الذي يؤثر بالآخر، وبالتالي هو المستقل، وإذا تعذر حسب الطريقتين السابقتين تحديد طبيعة المتغيرات يكون للباحث الخيار في اعتبار أحدهما مستقلاً والآخر تابعاً، وأن الخطأ المرتكب سيكون ضئيلاً كما سنرى فيما بعد.

3

بعد القيام بالإجراءات السابقة يتم حساب الارتباط والانحدار للظواهر، مع العلم بأنه يمكن دراسة علاقة متغير مستقل واحد مع متغير تابع واحد، وهنا يدعى الارتباط البسيط والانحدار البسيط، أما إذا درسنا علاقة عدة متغيرات مستقلة بمتغير تابع واحد، فالارتباط يدعى متعدداً والانحدار يدعى متعدداً مثل علاقة إنفاق الأسرة كمتغير تابع بدخلها، وعدد أفرادها ومكان إقامتها ودرجة تعلم أفرادها... كمتغيرات مستقلة .

8-2 تحليل الارتباط (Correlation Analysis)

يمكن تقسيم العلاقات بين الظواهر إلى قسمين: علاقات تابعية عندما يكون لكل قيمة للمتغير المستقل قيمة بالذات للمتغير التابع، مثل علاقة مساحة الدائرة بنصف القطر، ومساحة المربع بالضلع، واستهلاك معمل ما في أحد المواد الأولية بحجم الإنتاج، وعلاقات ارتباطية عندما يكون مقابل كل قيمة للمتغير المستقل قيمة تقريبية أو احتمالية للمتغير التابع، وإننا هنا سوف ندرس النوع الثاني فقط، حيث تمت دراسة النوع الأول من العلاقات في المرحلة الثانوية.

والعلاقة بين الظواهر إما أن تكون خطية (Linear) يمكن تمثيلها بخط مستقيم، أو منحنية (Non-Linear) تمثل بمنحنى غير المستقيم .

وسوف ندرس هنا الارتباط، ولأسباب درسية كما يأتي :

8-2-1 الارتباط البسيط للمتغيرات الكمية

8-2-2 الارتباط للمتغيرات النوعية

8-2-1 الارتباط البسيط للمتغيرات الكمية

بعد أن يتأكد الباحث من وجود علاقة بين المتغيرات قيد الدراسة يحدد المتغير المستقل والمتغير التابع ويقوم بجمع البيانات عن الظاهرتين، حيث من خلال هذه البيانات يستطيع أن يحدد متانة هذه العلاقة وطبيعتها بالطرائق الآتية :

8-2-1-1 التحليل المباشر :

إذا تبين من خلال قراءة بيانات المتغيرين أن القيم الدنيا للمتغير المستقل تتقابل مع القيم الدنيا للمتغير التابع والعليا مع العليا، فالعلاقة بين المتغيرين هي علاقة طردية، أما إذا كانت القيم الدنيا للمتغير المستقل تتقابل مع القيم العليا للمتغير التابع وبالعكس، فالعلاقة عكسية أي الزيادة في المستقل يقابلها تناقص بالمتغير التابع، ولمعرفة نوع العلاقة فيما إذا كانت خطية أو منحنية، نقارن الفروق بين كل قيمة للمتغير المستقل والقيمة السابقة لها مع مثيلاتها للمتغير التابع، فإذا كانت متقاربة فالعلاقة خطية، أما إذا كانت متباعدة فالعلاقة غير خطية .

8-2-1-2 شكل الانتشار (Scatter diagram)

يقصد بشكل الانتشار تمثيل البيانات عن الظاهرتين على المحاور الإحداثية، حيث يمثل المتغير المستقل X على محور السينات والمتغير التابع Y على محور العينات، وتمثل كل ثنائية (X, Y) بنقطة على المستوى الإحداثي، والنتيجة تكون شكلاً مؤلفاً من عدد كبير من النقاط يساوي عدد المشاهدات المأخوذة عن متغيرين، حيث يدعى هذا الشكل بشكل الانتشار .

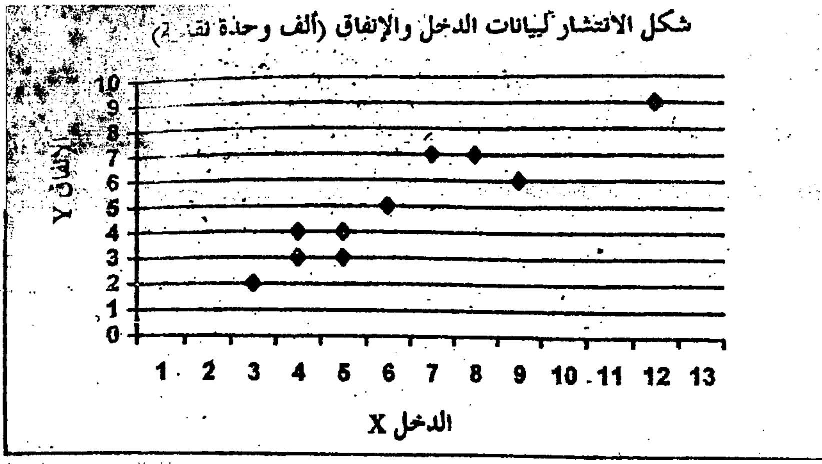
مثال (8-1) :

الجدول الآتي يمثل بيانات الدخل الشهري، والإنفاق الشهري لعينة مؤلفة من 10 أسر .

جدول رقم (8-1)

الأسرة	10	9	8	7	6	5	4	3	2	1
الدخل (بآلاف الوحدات)	12	5	4	9	8	7	6	3	5	4
الإنفاق (بآلاف الوحدات)	9	3	4	6	7	7	5	2	4	3

إن رسم شكل الانتشار للبيانات السابقة يعطي الشكل الآتي رقم (8-1).

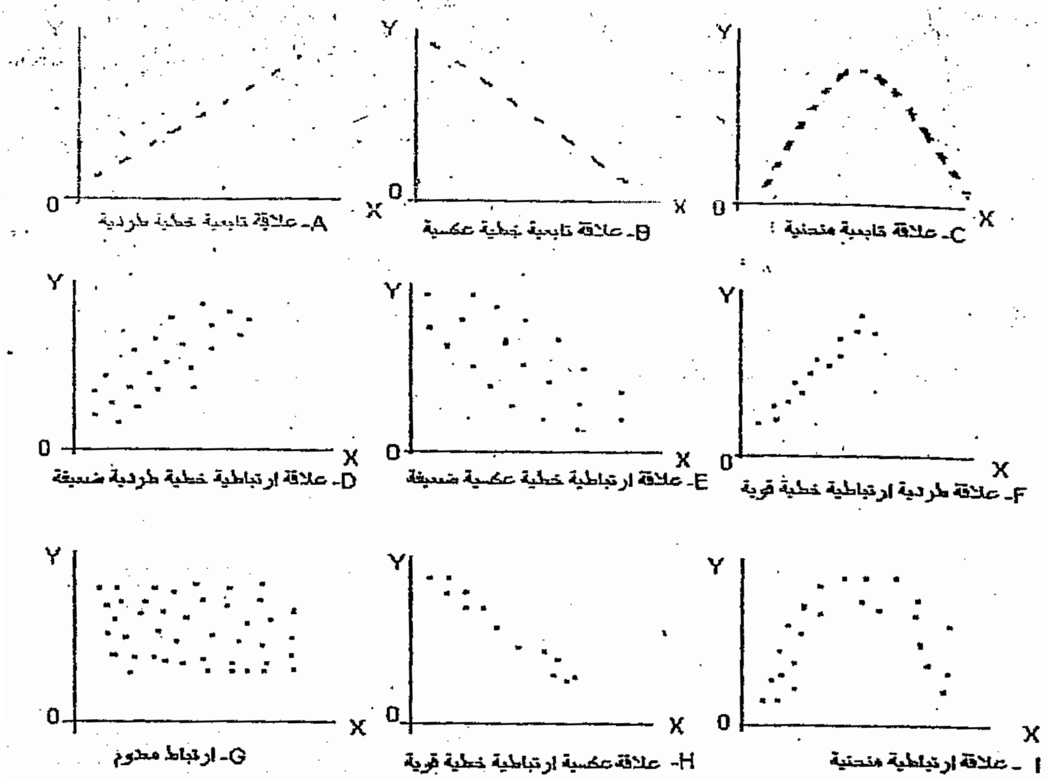


الشكل رقم (8-1)

يمكننا من خلال هذا الشكل ملاحظة طبيعة العلاقة، حيث القيم الدنيا للدخل تترافق مع القيم الدنيا للإنفاق والعليا مع العليا مما يبين أن العلاقة طردية. كما يمكن من الشكل معرفة متانة العلاقة، حيث تكون العلاقة تابعية عندما تقع النقاط على منحني واحد، وتكون ارتباطية وقوية كلما اقتربت النقاط من المنحني، ففي الشكل رقم (8-1) النقاط تنتشر حول مستقيم يمكن تمريره من بين النقاط مما يسمح لنا بالقول إن العلاقة خطية وقوية. وبشكل عام يمكن أن يأخذ شكل الانتشار أحد الأشكال الآتية الواردة في الشكل رقم (8-2)، ومن خلال قراءة هذه الأشكال يمكن تحديد طبيعة ونوع ومتانة العلاقة بين المتغيرات، فإذا كانت نقاط الانتشار تقع على مستقيم واحد تكون العلاقة قوية جداً، وبالتالي يكون الارتباط تاماً. ويمكن أن تكون العلاقة عكسية أو طردية حسب موقع المستقيم الأشكال: A, B (الشكل (8-2))، وإذا كانت تقع على منحني واحد كانت العلاقة تابعية منحنية الشكل ((8-2) - C). وعندما تكون نقاط الانتشار تشكل حزمة من النقاط لا تقع على منحني واحد تكون العلاقة طردية، وتكون العلاقة أقوى كلما قل عرض الحزمة، لاحظ الأشكال D, E, F, H, I في الشكل رقم(8-2) وعندما تكون النقاط مبعثرة دون أن يكون لها اتجاه معين يكون الارتباط معدوماً بين المتغيرين الشكل ((8-2) G)[1].

---

(1) لمزيد من المعلومات انظر كتاب : مبادئ الإحصاء، للأستاذ الدكتور ناظم خيذر، الصادر في دمشق عن مطبعة ابن حيان للعام 1982-1983 .



الشكل (8-2)

8-2-1-3 معامل الارتباط (Coefficient of correlation)

يساعد معامل ارتباط بيرسون (Pearson) على تحديد متانة وطبيعة العلاقة بين المتغيرين x و y وتعطى قيمة هذا المعامل بالعلاقة :

$$r_{xy} = \frac{\Sigma (x - \bar{x})\ (y - \bar{y})}{n\ S_x\ S_y} \qquad (1)$$

حيث : $\bar{x}$ و $\bar{y}$ الأوساط الحسابية للمتغير المستقل والمتغير التابع على التوالي .

S_x و S_y الانحرافات المعيارية للمتغير المستقل والمتغير التابع.

n عدد المشاهدات .

ويمكن إصلاح العلاقة السابقة وكتابتها بأشكال أخرى أكثر سهولة لحساب معامل الارتباط :

$$r_{xy} = \frac{n(\Sigma x y) - (\Sigma x)(\Sigma y)}{\sqrt{\left[n \Sigma x^2 - (\Sigma x)^2\right]\left[n \Sigma y^2 - (\Sigma y)^2\right]}} \qquad (2)$$

$$r_{xy} = \frac{\Sigma x y - n \bar{x} \bar{y}}{\sqrt{\left[\cancel{n} \Sigma x^2 - n \bar{x}^2\right]\left[\Sigma y^2 - n \bar{y}^2\right]}} \qquad (3)$$

وسوف نوضح فيما يأتي كيفية حساب معامل الارتباط بالعلاقات الثلاث السابقة من خلال العودة إلى المثال رقم (8-1) .

جدول رقم (8-2)

الأسرة	الإنفاق (ألف) X	الإنفاق (ألف) Y	xy	x*x	y*y	(x-x)	(x-y)	(x-x) * (y-y)
1	4	3	12	16	9	-2.3	-2	4.6
2	5	4	20	25	16	-1.3	-1	1.3
3	3	2	6	9	4	-3.3	-3	9.9
4	6	5	30	36	25	-0.3	0	0
5	7	7	49	49	49	0.7	2	1.4
6	8	7	56	64	49	1.7	2	3.4
7	9	6	54	81	36	2.7	1	2.7
8	4	4	16	16	16	-2.3	-1	2.3
9	5	3	15	25	9	-1.3	-2	2.6
10	12	9	108	144	81	5.7	4	22.8
المجموع	63	50	366	465	294	0	0	51

الحل حسب الصيغة رقم (1) .

نوجد الانحراف المعياري لـــ x وكذلك لـــ y :

$$S_x = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$$

$$= \sqrt{\frac{465}{10} - (6.3)^2} = 2.6096$$

$$S_y = \sqrt{\frac{\Sigma y^2}{n} - \bar{y}^2}$$

$$= \sqrt{\frac{294}{10} - (5)^2} = 2.0976$$

ومنـــه :

$$r_{xy} = \frac{\Sigma (x - \bar{x})\ (y - \bar{y})}{n\ S_x\ S_y}$$

$$= \frac{51}{(10)\,(2.6096)\,(2.0976)} = +\ 0.9317$$

الحل حسب الصيغة رقم (2) :

$$r_{xy} = \frac{n\,\Sigma xy - \Sigma x\,\Sigma y}{\sqrt{\left[n\,\Sigma x^2 - (\Sigma x)^2\right]\left[n\,\Sigma y^2 - (\Sigma y)^2\right]}}$$

$$= \frac{10\,(366) - (63)\,(50)}{\sqrt{\left[10\,(465) - (63)^2\right]\left[10\,(294) - (50)^2\right]}} = +\ 0.9317$$

الحل حسب الصيغة رقم (3) :

$$r_{xy} = \frac{\Sigma xy - n\,\bar{x}\,\bar{y}}{\sqrt{\left[\Sigma x^2 - n\,\bar{x}^2\right]\left[\Sigma y^2 - n\,\bar{y}^2\right]}}$$

$$= \frac{366 - 10\ (6.3)\,(5)}{\sqrt{\left[465 - 10\,(6.3)^2\right]\left[294 - 10\,(5)^2\right]}} = +\ 0.9317$$

8-2-1-3-1 مزايا ومساوئ معامل ارتباط بيرسون :

1- يأخذ معامل ارتباط r قيمة تتراوح بين −1 إلى +1 ويمكن إثبات ذلك رياضياً.

2- تكون قيمة r تساوي الواحد عندما تكون العلاقة تامة، وتبلغ قيمته الصفر عندما يكون الارتباط معدوماً، وتقترب من الواحد عندما تكون العلاقة قوية جداً.

3- يأخذ معامل الارتباط قيماً موجبة عندما تكون العلاقة طردية، وسالبة عندما تكون عكسية.

4- قيمة معامل الارتباط مجردة من وحدات القياس، وهذه من أهم مزاياه حيث يتم التخلص من وحدات القياس عند التقسيم على الانحرافات المعيارية (لاحظ الصيغة 1) في حين أن أهم مساوئ معامل ارتباط بيرسون أنه لا يعبر عن متانة العلاقة بشكل صحيح إلا إذا كانت العلاقة خطية، أما إذا كانت العلاقة منحنية، فإن قيمته لا تعبر عن مكانة العلاقة، وفي هذه الحالة يجب استخدام مقاييس أخرى، كما أنه لا يمكن استخدام معامل بيرسون في حالة المتغيرات النوعية.

$$r_{xy} = \frac{\Sigma\Sigma f_{ij} x_i y_j - n\bar{x}\bar{y}}{n S_x S_y}$$

$$= \frac{12480 - 100(7.56)(15.88)}{100(2.38)(2.25)}$$

$$= +0.88$$

$$r_{xy} = \frac{n\Sigma\Sigma f_{ij} x_i y_j - (\Sigma f_i x_i)(\Sigma f_j y_j)}{\sqrt{[n\Sigma f_i x_i^2 - (\Sigma f_i x_i)^2][n\Sigma f_j y_j^2 - (\Sigma f_j y_j)^2]}}$$

$$r_{xy} = \frac{100(12480) - (756)(1588)}{\sqrt{[100(6282) - (756)^2][100(25724) - (1588)^2]}} = +0.88$$

وتجب الإشارة هنا إلى أنه يمكن استخدام أسلوب الانحرافات المختصرة في الصيغة الأولى أو الثانية لحساب معامل الارتباط .

8-2-1-5 الدلالة الإحصائية لمعامل الارتباط

عند دراسة الارتباط بين متغيرين فإننا في أغلب الأحيان نستخدم عدداً من المشاهدات (n) وهي عبارة عن عينة عشوائية، حيث نحصل من هذه العينة على قيمة لمعامل الارتباط ولكن السؤال الذي يطرح هو: هل هذه القيمة لمعامل الارتباط هي فعلاً موجودة في المجتمع الإحصائي أم لا ؟ وللتأكد من ذلك نقوم عادةً باختبار دلالة أو معنوية معامل الارتباط، وجوهر هذا الاختبار يقوم على أساس اختبار الفرضية القائلة بأن قيمة معامل الارتباط في المجتمع الإحصائي تساوي الصفر، فإذا ما رفضنا هذه الفرضية نكون عندها قد تأكدنا من وجود ارتباط في المجتمع، والمعامل الناتج يعبر عنه أما إذا قبلنا الفرضية فهذا يعني أن قيمة معامل الارتباط لا تعبر عن الارتباط في المجتمع الإحصائي، وإن قيمته يمكن أن تساوي الصفر، وبالتالي ليس لمعامل الارتباط أية دلالة أو قيمة غير معنوية .

ولإجراء الاختبار نستخدم التحويل t حيث :

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

وندعو هذه القيمة بالقيمة المحسوبة لـ t ، نقارن هذه القيمة مع القيمة النظرية لـ t المستخرجة من جداول ستودنت عند مستوى دلالة α محدد وعدد درجات حرية $n-2$ حيث n عدد المشاهدات .

لنتأكد الآن من الدلالة الإحصائية لمعامل الارتباط الذي أوجدناه سابقاً من بيانات الجدول رقم (8-2) حيث وجدنا أن : $r = 0.93$ وإن $n = 10$:

$$t = 0.93\sqrt{\frac{10-2}{1-(0.93)^2}} = 7.15$$

أما قيمة t النظرية عند مستوى معنوية 5% و 8 درجات حرية :

$$t_{(0.05)(8)} = 2.3$$

نلاحظ أن قيمة t المحسوبة أكبر من t النظرية، وبالتالي نرفض الفرضية من أن قيمة معامل الارتباط بالمجتمع الإحصائي تساوي الصفر، وبالتالي قيمة r التي حصلنا عليها هي قيمة معنوية .

أو بطريقة أخرى نحسب الخطأ المعياري لمعامل الارتباط :

$$S_r = \sqrt{\frac{1-r^2}{n-2}}$$

فإذا كانت قيمة معامل الارتباط r أكبر من ثلاثة أمثال خطئه المعياري، يكون لمعامل الارتباط أهمية إحصائية أو دلالة إحصائية، وذلك باحتمال 99.7% ، وبتطبيق هذه الطريقة على بيانات المثال رقم (1) نجد :

$$S_r = \sqrt{\frac{1-(0.93)^2}{10-2}} = 0.129$$

وهنا نلاحظ أن $r > 3\ S_r$ وبالتالي نقول إن لمعامل الارتباط دلالة إحصائية، وإن قيمته في المجتمع تختلف عن الصفر.

8-2-2 الارتباط للمتغيرات النوعية

شاهدنا سابقاً بعض الأساليب التي نستطيع بمساعدتها الكشف عن مكانة العلاقة الارتباطية للمتغيرات الكمية، ولكنّ أياً من هذه الأساليب لا يصلح لحساب مكانة العلاقة الارتباطية للمتغيرات النوعية، حيث تستخدم في هذا المجال عدة أساليب مختلفة تبعاً للحالة المدروسة، وسوف نقوم فيما يأتي بدراسة هذه الأساليب بالتفصيل.

8-2-2-1 معامل الاقتران

يستخدم معامل الاقتران عندما يكون لدينا متغيران نوعيان، ولكل متغير قيمتان أو صفتان كأن ندرس العلاقة بين الإصابة بالسرطان لمجموعة من الأشخاص مدخنين وغير مدخنين، فالمتغيران هنا المرضى (مصاب بالسرطان وغير مصاب)، والتدخين (مدخن وغير مدخن)، أو دراسة العلاقة بين لون الزهرة ورائحتها..

ولحساب معامل الاقتران نقسم المفردات حسب المتغيرات والصفات كما في الجدول (8-4).

جدول رقم (8-4)

المتغير A / المتغير B	الصفة 1	الصفة 2	المجموع
الصفة 1	n_{11}	n_{12}	$'n_2$
الصفة 2	n_{21}	n_{22}	$'n_2$
المجموع	n_1	n_2	N

حيث : $n_{11} + n_{12} + n_{21} + n_{22} = N$

ويحسب معامل الاقتران كما يأتي :

$$r_c = \frac{n_{11} * n_{22} - n_{12} * n_{21}}{n_{11} * n_{22} + n_{12} * n_{21}}$$

وتكون العلاقة بين المتغيرات قوية جداً كلما اقتربت قيمة معامل الاقتران من الواحد وبشكل عام نقول إن هناك علاقة عندما تكون قيمة r_c أكبر من 0.5.

مثال (3-8) :

بهدف دراسة العلاقة بين التدخين والإصابة بالتهاب اللثة قام أحد الباحثين بجمع بيانات عن 200 شخص كان توزيعهم حسب المتغيرات والصفات كما في الجدول (5-8).

جدول رقم (5-8)

المتغير A / المتغير B	مدخن	غير مدخن	المجموع
مصـــاب	70	10	80
غير مصاب	20	100	120
المجموع	90	110	200

$$r_c = \frac{70*100 - 20*10}{70*100 + 20*10} = \frac{6800}{7200} = 0.944$$

وهذا يعني أن هناك علاقة قوية بين الإصابة بالتهاب اللثة والتدخين. وتجب الإشارة هنا إلى أنه ليس لإشارة r_c أي مدلول، لأن إشارته تختلف حسب ترتيب الصفات في الجدول، لذلك يفضل حسابه بالقيمة المطلقة للدلالة على وجود علاقة أو لا. (لاحظ لو بدلنا السطر الأول بالثاني في الجدول السابق لأصبحت قيمة معامل الاقتران 0.944 -) .

3-2-2-2 معامل التوافق

رأينا أن معامل الاقتران يستخدم لقياس متانة العلاقة بين متغيرين نوعيين، لكل منهما صفتان (أي الجدول يحوي أربع خلايا)، في حين لا يمكن استخدامه إذا كان لأحد المتغيرين على الأقل أكثر من صفتين حيث يصبح الجدول في هذه الحالة مركباً في أكثر من أربع خلايا، والمعامل الذي يمكن استخدامه هو معامل التوافق وصيغته كما يأتي:

$$r_A = \sqrt{\frac{G-1}{G}}$$

حيث :

$$G = \sum \left(\frac{f_{ij}^2}{f_i * f_j} \right)$$

f_{ij} : عدد التكرارات في الخلية الواقعة في السطر i والعمود j .

f_i : مجموع السطر i .

f_j : مجموع العمود j .

مثال (8-4) :

بهدف دراسة تأثير الحالة التعليمية للأم على مستوى الابن في المدرسة الابتدائية تم جمع البيانات الآتية عن 200 تلميذ في إحدى المدارس .

جدول رقم (8-6)

الحالة التعليمية للأم / مستوى الابن في المدرسة	أمية وابتدائية	ثانوية وإعدادية	فوق الثانوية	المجموع
متفوق	5	15	30	50
وسط	5	40	15	60
ضعيف	70	15	5	90
المجموع	80	70	50	200

$$G = \frac{5^2}{80 \times 50} + \frac{5^2}{80 \times 60} + \frac{70^2}{80 \times 90} + \frac{15^2}{70 \times 50} + \frac{40^2}{70 \times 60}$$

$$+ \frac{15^2}{70 \times 90} + \frac{30^2}{50 \times 50} + \frac{15^2}{60 \times 50} + \frac{5^2}{50 \times 90} = 1.613$$

$$r_A = \sqrt{\frac{G-1}{G}} = \sqrt{\frac{1.613-1}{1.613}} = 0.616$$

وهذا يعني أن هناك علاقة قوية بين حالة الأم التعليمية ومستوى الابن في الصف. وبشكل عام نقول إن هناك علاقة قوية أو جيدة بين المتغيرين، كلما كانت قيمة معامل التوافق أكبر من 0.5 .

8-2-2-3 معامل ارتباط الرتب (Coefficient of Rank Correlation)

يستخدم معامل ارتباط الرتب أو معامل ارتباط سبيرمان (Sperman) نسبة إلى واضعه، لقياس متانة العلاقة بين متغيرين نوعيين، لكل منهما عدد كبير من الصفات، أو إذا كان أحد المتغيرين نوعياً والآخر كمياً. وكذلك الأمر يمكن استخدامه في حالة المتغيرات الكمية أحياناً لاختصار الحسابات .

يتم حساب ارتباط الرتب باستبدال الصفات النوعية بقيم عددية تعبر عنها وتسمى الرتب Rank، حيث يتم وضعها عن طريق إيجاد العلاقة الداخلية للصفات ولكل متغير على حده، فنبدأ بأضعف الصفات مثلاً ونعطيها الرتبة 1 والأقوى منها رقم 2 وهكذا، بحيث يتولد لدينا متوالية عددية لكل متغير يكون مجموعها متساوياً، لأن عدد الصفات لكل متغير متساوية، وبعد ذلك نحسب معامل ارتباط الرتب بالعلاقة:

$$r_s = 1 - \frac{6\Sigma D^2}{n(n^2-1)}$$

حيث D : فرق رتبة X من رتبة Y لكل ثنائية .

n : عدد المشاهدات .

مثال (8-5) :

ليكن لدينا الجدول الآتي الذي يمثل البيانات الخاصة بشكل علبة الحلوى والطلب على هذا النوع من الحلوى .

جدول رقم (8-7)

شكل علبة الحلوى (x)	نوع الطلب (y)	رتب x	رتب y	D	D^2
علبة بسيطة دون رسوم	ضعيف	1	1	0	0
علبة بسيطة برسوم	وسط	2	2	0	0
علبة بسيطة برسوم وغلاف	جيد جداً	3	4	1	1
علبة مزخرفة	جيد	4	3	1	1
علبة حديدية تقليدية	عالي جداً	5	5	0	0
المجموع					2

$$r_s = 1 - \frac{6\Sigma D^2}{n(n^2-1)}$$

$$= 1 - \frac{6\times 2}{5(5^2-1)} = 1 - \frac{12}{120} = 0.9$$

أي أن هناك علاقة قوية جداً بين شكل العلبة ونوع الطلب عليها. ويجب الإشارة هنا إلى أن معامل ارتباط الرتب يتمتع بالمزايا نفسها التي يتمتع بها معامل ارتباط بيرسون، باعتبار أنه مستنتج من-معامل ارتباط بيرسون بعد تعويض ΣX و Σy و ΣX^2 و ΣY^2 بما يقابلها باعتبارها متواليات عددية بعد إعطاء الرتب للمتغيرين .

8-2-2-3-1 معامل ارتباط الرتب في حالة تكرار عدد من الصفات:

مثال (8-6) :

تبين البيانات الآتية غزارة الأمطار ولون الغيم في إحدى المدن في فصل الشتاء .

17

جدول رقم (8-8)

لون الغيم (x)	كمية المطر ميليمتر/الدقيقة (y)	رتب x	رتب y	D	D^2
أبيض متفرق	0	1	1	0	0
أبيض كثيف	2	2	2	0	0
رمادي	3	3	3.5	-0.5	0.25
رمادي غامق	3	4	3.5	0.5	0.25
أسـود	4	5	5	0	0
المجموع		15	15		0.5

نلاحظ من هذه البيانات أن أحد المتغيرين نوعي وهو لون الغيوم والآخر كمي.

ولقد تم وضع رتب x مبتدئين بالأضعف حسب معرفتنا بحالات هطول المطر، والأضعف في هذه الحالة إذا كان لون الغيم أبيض متفرقاً، وأعطيناه الرتبة 1، والأفضل هو أبيض كثيف 2 وهكذا... أما المتغير الآخر هو كمية المطر وبياناته كمية، وهنا لن تكون هناك صعوبة في وضع الرتب، حيث أقل هذه البيانات سوف يأخذ الرتبة 1 وهكذا، ولكن هناك مشكلة تكرار إحدى صفات الرقم 3، في هذه الحالة تعطى الرتب كما يأتي : حسب الترتيب أول صفة يجب أن تعطى الرتبة 3 في البيانات، والصفة المكررة الأخرى يجب أن تعطي الرتبة 4 ، ولكن بما أن الصفات متساوية فيجب أن تساوي بين الرتبتين، ويتم ذلك من خلال إيجاد الوسط الحسابي للرتبتين 3 و 4 وهو 3.5، فيعطى كل منها الرتبة 3.5 بهذه الطريقة نكون قد حافظنا على مجموع الرتب لكل من المتغيرين بحيث يكون متساوياً، انظر الجدول رقم (8-8) من المثال (6-8) .

جدول رقم (13-8)

مجموع الرتب D	f	Df	fD^2
2	–	–	–
3	–	–	–
4	10	40	160
5	28	140	700
6	52	312	1872
7	10	70	490
8	–	–	–
9	–	–	–
المجموع	10	562	3222

نحسب S_D^2 من العلاقة :

$$S_D^2 = \frac{\Sigma f D^2}{\Sigma f} - \left(\frac{\Sigma f D}{\Sigma f}\right)^2$$

$$= \frac{3222}{100} - \left(\frac{562}{100}\right)^2 = 0.6356$$

نحسب معامل الارتباط من العلاقة :

$$r_s = \frac{S_D^2 - S_x^2 - S_y^2}{2 * S_x * S_y}$$

$$= \frac{0.6356 - 1 - 2.09}{2 * 1 * 1.447} = - 0.849$$

وهذا يعني أن العلاقة قوية جداً، وعكسية بين المستوى التعليمي للأم وعدد الولادات.

8-3 تحليل الانحدار Regression Analysis :

رأينا سابقاً كيف يمكننا تحديد وجود/أو عدم وجود علاقة بين المتغيرات، وكيف نستطيع تحديد طبيعة العلاقة ومكانتها، ولكن هل يمكننا بالإضافة إلى ذلك تحديد مدى تأثير المتغير المستقل X على المتغير التابع Y؟ وحساب مقدار التغير في المتغير التابع عندما يزداد المتغير المستقل وحدة واحدة؟ وتقدير قيمة المتغير التابع عند قيمة محددة للمتغير المستقل؟ إن الإجابة على هذه الأسئلة يكون من خلال دراسة الانحدار Regression.

والانحدار هو تمثيل العلاقة بين المتغيرات بمعادلة رياضية، حيث يمكن تقسيم الانحدار حسب نوع المعادلة وعدد المتغيرات إلى :

1- الانحدار البسيط: وهو دراسة العلاقة بين متغيرين أحدهما مستقل والآخر تابع، وهو بدوره يقسم إلى قسمين :

- الانحدار الخطي البسيط Simple linear Regression :

وذلك عندما يمكن تمثيل العلاقة بخط مستقيم .

- الانحدار غير الخطي البسيط Simple non-linear Regression :

وذلك عندما تكون العلاقة بين المتغيرين تُمثل بعلاقة غير معادلة المستقيم: درجة ثانية – ثالثة – أسية – لوغارتيمية .

2- الانحدار المتعدد Multiple Regression : وهو دراسة الانحدار بين عدة متغيرات مستقلة ومتغير واحد تابع، ويقسم بدوره إلى قسمين:

- الانحدار المتعدد الخطي Multiple non-Linear Regression :

وتكون بتمثيل العلاقة بين المتغيرات المستقلة والمتغير التابع بعلاقة غير خطية .

وسوف نقوم بدراسة الانحدار الخطي البسيط بشكل مفصل، وذلك لاستعمالاته المتعددة ، ولكي يكون مدخلاً لفهم علاقات الانحدار الأخرى ، ونترك

الانحدار المتعدد الخطي أو اللاخطي بالإضافة إلى الارتباط والانحدار المتعدد للسنوات القادمة .

8-3-1 الانحدار الخطي البسيط :

إن دراسة الانحدار بين متغيرين x و y تبدأ بتمثيل العلاقة بينهما بخط مستقيم يدعى مستقيم الانحدار، والمعادلة الممثلة لهذا المستقيم تسمى بمعادلة الانحدار، وهي من الشكل :

$$\hat{y}_i = a + b\, x_i$$

حيث : $\hat{y}_i$ – القيمة المقدرة أو المحسوبة للمتغير التابع.

a – ثابت الانحدار .

b – معامل الانحدار .

ومن الممكن إيجاد انحدار y على x ، أو انحدار x على y، غير أنه يفضل التركيز على انحدار y على x نظراً لأنه أكثر أهمية من انحدار x على y من جهة، وباعتبار أننا حددنا مسبقاً المتغير المستقل والمتغير التابع من جهة أخرى.

بعد إيجاد معادلة الانحدار الممثلة للعلاقة يمكن الاستمرار بدراسة الانحدار باستخدام أساليب تحليل أخرى سوف نتعرض لها لاحقاً ولكن من الضروري أولاً بيان كيفية إيجاد معادلة الانحدار .

8-3-1-1 إيجاد معادلة مستقيم الانحدار :

إن التدقيق في الشكل رقم (8-1) أو في الشكل رقم (8-3) الذي يمثل شكل الانتشار لبيانات الدخل والإنفاق لــ 10 أسر كما هو وارد في الجدول رقم (8-14).

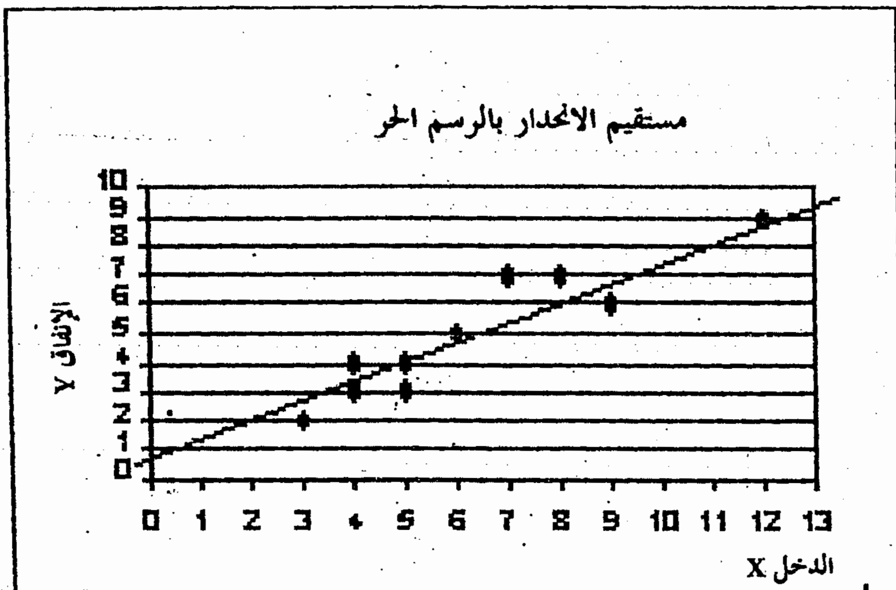
جدول رقم (8-14)

رقم الأسرة	1	2	3	4	5	6	7	8	9	10
دخل الأسرة (1000) وحدة X	4	5	3	6	7	8	9	4	5	12
إنفاق الأسرة (1000) وحدة Y	3	4	2	5	7	7	6	4	3	9

نرى أنه يمكن من خلال نقاط الانتشار رسم عدد لا نهائي من المستقيمات،
وذلك لتعذر الحصول على مستقيم يمر من كل هذه النقاط، وبالتالي يمكن اختيار
مستقيم يمر خلال الشكل ممثلاً للعلاقة، ويكون هذا المستقيم ممثلاً بشكل أفضل
بالاعتماد على طريقة علمية ما في إيجاده، حيث توجد عدة طرائق تستخدم لإيجاد
أفضل مستقيم، وسوف نناقش بعضها أدناه .

– **طريقة الرسم الحر في إيجاد مستقيم الانحدار Freehand Method :**

تعتمد هذه الطريقة في إيجاد مستقيم الانحدار على رسم مستقيم يمر من جميع نقاط الانتشار أو معظمها، وإن عملية الحصول على المستقيم الأفضل يعتمد على الخبرة في عملية رسم . انظر الشكل (3-8) .



الشكل رقم (3-8)

لإيجاد معادلة الانحدار $\hat{y}_i = a + b X_i$ لا بد من إيجاد الثوابت a, b، حيث يتم ذلك كما يأتي : إن الثابت b هو عبارة عن ميل المستقيم، ويعطى بالعلاقة :

$$b = \frac{\Delta Y}{\Delta X}$$

نفتش عن نقطتين يمر منها المستقيم ونحسب ΔY و ΔX حيث نرى في الشكل (8-3) أن المستقيم يمر من النقطتين التاليتين (5, 4) و (12,9)، وبالتالي يكون :

$$\Delta Y = 9 - 4 = 5$$

$$\Delta X = 12 - 5 = 7$$

وبالتالي :

$$b = \frac{5}{7} = 0.714$$

أما قيمة الثابت a فهي عبارة عن نقطة التقاطع مع محور العينات وهي في الشكل (8-3) تساوي تقريباً 0.7. وعليه تكون معادلة الانحدار من الشكل :

$$\hat{y}_i = 0.7 + 0.714$$

تتميز هذه الطريقة بالسهولة والسرعة، ولكن من سيئاتها أنها قد تعطي نتائج غير دقيقة ومختلفة تبعاً للخبرة الشخصية، واختلاف أسلوب الرسم، حيث يمكن أن نحصل على عدد من الخطوط المستقيمة تساوي عدد القائمين على عملية الرسم .

– طريقة المربعات الصغرى The Method of Least Squares :

إن جوهر هذه الطريقة يتلخص بجعل الانحرافات بين القيم الفعلية y_i للمتغير التابع عن مستقيم الانحدار y_i أصغر ما يمكن أي :

$$\Sigma \left(y_i - \hat{y}_i\right) \rightarrow \min$$

ولكن $\Sigma\,(y_i - \hat{y}_i) = 0$ حيث $\hat{y}$ يمثل القيمة المتوسطة لـــ y_i عند قيمة معينة لـــ x كما سنرى فيما بعد، ولكي نتخلص من هذه النتيجة نأخذ مربعات الفروق وعليه نشكل التابع f :

$$f = \Sigma\,(y_i - \hat{y}_i)^2 \rightarrow \min$$

أو :

$$f = \Sigma\,(y_i - a - b\,x_i)^2 \rightarrow \min$$

نلاحظ أن قيمة هذا التابع ترتبط بالمتغيرين a و b، حيث تزداد بازديادهما وتنقص بتنقصانهما، وإذا ما نظرنا إلى شكل الانتشار لوجدنا أن عدد الخطوط المستقيمة التي يمكن أن تمر عبر نقاط الشكل هو لا نهائي، وتختلف عن بعضها باختلاف أحد المتغيرين على الأقل، فإذا ما بلغ التابع F النهاية الصغرى تكون الانحرافات أقل ما يمكن، وبالتالي نشتق التابع اشتقاقاً جزئياً بالنسبة لـــ a أولاً، ومن ثم لـــ b ونعدم بعد ذلك المشتق :

$$\frac{\partial F}{\partial a} = 2\,\Sigma\,(y_i - a - b\,x_i)\,(-1) = 0$$

وبإصلاح العلاقة نحصل على المعادلة الآتية التي تدعى المعادلة الطبيعية الأولى:

$$\Sigma\,Y_i = na + b\,\Sigma\,X_i$$

$$\frac{\partial F}{\partial b} = 2\,\Sigma\,(y_i - a - b\,x_i)\,(-x_i) = 0$$

ومنه :

$$\Sigma\,x_i\,y_i = a\,\Sigma\,x_i + b\,\Sigma\,x_i^2$$

وهي المعادلة الطبيعية الثانية .

وبالحل المشترك لهاتين المعادلتين الطبيعيتين :

$$\Sigma\,y_i = na + b\,\Sigma x_i$$

$$\Sigma\,x_i\,y_i = a\,\Sigma\,x_i + b\,\Sigma\,x_i^2$$

عن طريق استخدام إحدى الطرائق المعروفة لحل معادلتين بمجهولين نحصل على الثوابت a و b المجهولين، ومن هذه الطرائق نذكر :

- الحذف بالتعويض .
- الحذف بالطرح .
- الحل المصفوفي .
- المعينات .

ونلاحظ من المعادلتين السابقتين أن جميع القيم الأخرى يمكن الحصول عليها من نات الفعلية .

سوف نقوم فيما يلي بإيجاد معادلة مستقيم الانحدار للبيانات الواردة في الجدول م (8-14)، والتي أعدنا كتابتها بالجدول رقم (8-15) لإيجاد القيم المطلوبة لحل ادلتين :

جدول رقم (8-15)

رقم الأسرة	x	y	x*y	x^2	y^2
1	4	3	12	16	9
2	5	4	20	25	16
3	3	2	6	9	4
4	6	5	30	36	25
5	7	7	49	49	49
6	8	7	56	64	49
7	9	6	54	81	36
8	4	4	16	16	16
9	5	3	15	25	9
10	12	9	108	144	81
المجموع	63	50	366	465	294

– الحل باستخدام طريقة الحذف بالطرح :

نعوض في المعادلتين بالقيم التي حصلنا عليها من الجدول رقم (8-15) :

$$50 = 10 * a + 63\,b \qquad \text{I}$$

$$366 = 63 * a + 465\,b \qquad \text{II}$$

نلاحظ أنه لحذف أحد المتغيرين من المعادلتين بالطرح يجب ضرب المعادلة الأولى — 6.3، وبالتالي نحصل على المعادلة الآتية :

$$315 = 63 * a + 396.9\,b \qquad \text{III}$$

وبطرح المعادلة III من II نحصل على المعادلة IV .

$$15 = 68.1\,b \qquad \text{IV}$$

ومنه :

$$b = \frac{51}{68.1} = 0.7489$$

وبتعويض قيمة b في المعادلة I نحصل على قيمة $a = 0.2819$.

– الحل باستخدام طريقة المعينات (الطريقة المباشرة) :

المعين الأساسي Δ :

$$\Delta = \begin{vmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{vmatrix} = n\,\Sigma x_i^2 - (\Sigma x_i)(\Sigma x_i) = n\,\Sigma x_i^2 - (\Sigma x_i)^2$$

المعين المساعد Δ_a :

$$\Delta_a = \begin{vmatrix} \Sigma y_i & \Sigma x_i \\ \Sigma x_i\, y_i & \Sigma x_i^2 \end{vmatrix} = (\Sigma y_i)\,(\Sigma x_i^2) - (\Sigma x_i\, y_i)\,(\Sigma x_i)$$

المعين المساعد Δ_b :

$$\Delta_b = \begin{vmatrix} n & \Sigma y_i \\ \Sigma x_i & \Sigma x_i\, y_i \end{vmatrix} = n\,\Sigma\, x_i\, y_i - \Sigma x_i\, \Sigma y_i$$

$$b = \frac{\Delta b}{\Delta} = \frac{n \Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n \Sigma x_i^2 - (\Sigma x_i)(\Sigma x_i)}$$

وبتقسيم البسط والمقام على n نحصل على :

$$= \frac{\Sigma x_i y_i - \overline{x} \Sigma y_i}{\Sigma x_i^2 - \overline{x} \Sigma x_i}$$

وهذا يساوي أيضاً :

$$= \frac{\Sigma x_i y_i - n \overline{x}\, \overline{y}}{\Sigma x_i^2 - n \overline{x}^2}$$

أما الثابت a :

$$a = \frac{\Delta a}{\Delta} = \frac{(\Sigma y_i)(\Sigma x_i^2) - (\Sigma x_i y_i)(\Sigma x_i)}{n \Sigma x_i^2 - (\Sigma x_i)(\Sigma x_i)}$$

ويمكننا بطريقة أخرى أسهل إيجاد قيمة الثابت a بعد الحصول على قيمة الثابت b من المعادلة الطبيعية الأولى إذا قسمنا هذه المعادلة على n فإننا نحصل على :

$$\overline{y} = a + b\, \overline{x}$$

ومنه :

$$a = \overline{y} - b\, \overline{x}$$

وبتطبيق هذه العلاقات بشكل مباشر على المثال السابق نستطيع تحديد قيمة ثابتي المعادلة a و b :

$$b = \frac{366 - 6.3 * 50}{465 - 6.3 * 63} = \frac{51}{68.1} = 0.7489$$

$$= \frac{366 - 10(6.3 * 5)}{465 - 10(6.3)^2} = \frac{51}{68.1} = 0.7489$$

$$a = \frac{50\,(465) - 366\,(63)}{10\,(465) - (63)\,(63)} = \frac{192}{681} = 0.2819$$

$$= 5 - 07489 * 603 = 0.2819$$

حيث :

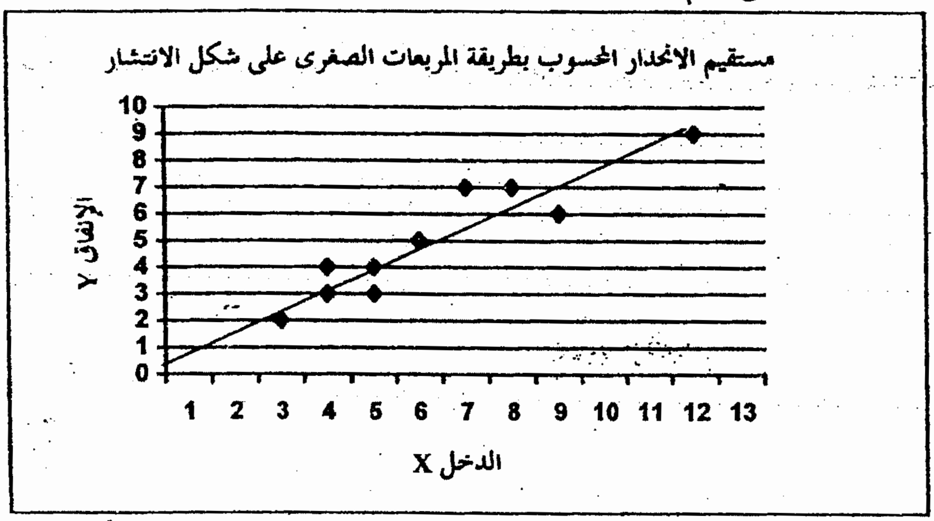
$$\bar{x} = \frac{\Sigma x}{n} = \frac{63}{10} = 6.3$$

$$\bar{y} = \frac{\Sigma Y}{n} = \frac{50}{10} = 5$$

وهكذا، فإن معادلة الانحدار تكون من الشكل الآتي

$$\hat{y}_i = 0.2819 + 0.7489\, x_i$$

انظر الشكل رقم (8-4) .



الشكل رقم (8-4)

8-3-1-2 أهمية معادلة مستقيم الانحدار :

تقدم معادلة مستقيم الانحدار التي يتم إيجادها بطريقة المربعات الصغرى بمجموعة من المعلومات الهامة عند تحليل البيانات وهي :

A – تعطي معادلة الانحدار القيمة التي يأخذها المتحول التابع بالمتوسط عندما يأخذ المتحول المستقل قيمة محددة. ولتوضيح ذلك نورد المثال الآتي :

مثال (8-8) :

يبين الجدول رقم (8-16) كمية السماد المستخدم وغلة القطعة من القمح لـ 12 قطعة أرض استخدمت في تجربة قام بها أحد الباحثين لمعرفة مدى تأثير السماد على الغلة فكانت النتائج :

جدول رقم (8-16)

x^2	$x\,y$	الغلة (10) كيلو غرام y	كمية السماد (1) كيلو غرام x	رقم قطعة الأرض
0	0	19	0	1
0	0	20	0	2
0	0	21	0	3
1	21	21	1	4
1	22	22	1	5
1	23	23	1	6
4	46	23	2	7
4	48	24	2	8
4	50	25	2	9
9	75	25	3	10
9	78	26	3	11
9	81	27	3	12
42	444	276	18	المجموع

لإيجاد معادلة مستقيم الانحدار الممثلة للعلاقة بين كمية السماد والغلة يجب تحديد قيمة الثوابت a و b :

$$b = \frac{\Sigma x_i y_i - \bar{x}\Sigma y_i}{\Sigma x_i^2 - \bar{x}\Sigma x_i}$$

$$= \frac{444 - 1.5 * 276}{42 - 1.5 * 18} = \frac{30}{15} = 2$$

$$a = \bar{y} - b\,\bar{x}$$

$$= 23 - 2 * 1.5 = 20$$

$$\hat{y}_i = 20 + 2x_i$$

إن القيمة التي تعطيها هذه المعادلة تتضح من خلال بيانات الجدول رقم (8-16)، فعندما تكون كمية السماد 0 كغ يكون متوسط الغلة 20 وحدة (200 كغ)، وكذلك يكون متوسط الغلة 22 وحدة عند استخدام 1 كغ سماد للغلة وهكذا... فإذا عوضنا في المعادلة المتحول المستقل X بقيمة محدد ولتكن 2 كغ، فإنها تأخذ الشكل الآتي :

$$\hat{y}_i = 20 + 2 * 2 = 24$$

أي عند استخدام 2 كغ سماد للقطعة، فإن الغلة سوف تكون 24 وحدة، وهي تعبر عن متوسط الغلة عند استخدام 2 كغ من السماد، انظر الجدول رقم (8-16). وهكذا فإن معادلة مستقيم الانحدار تعطي دوماً قيمة التابع بالمتوسط، حيث تبقى هذه النتيجة نفسها دون تغيير، حتى لو كانت البيانات موضع الدراسة لا تحتوي على قيمة الأوساط الحسابية للتابع كما هو الحال في مثالنا. وللتأكد من ذلك يمكننا حذف المشاهدات التي تحوي أوساطاً حسابية لـ y ، وأن المعادلة سوف تبقى نفسها والقيم التي تعطيها هي القيم المتوسطة نفسها. وأن مجموع القيم الفعلية للتابع يساوي مجموع القيم المحسوبة من المعادلة .

B – نستطيع من خلال معادلة مستقيم [illegible]

يأخذها المتغير التابع بالمتوسط، عندما تكون قيمة المتغير المستقل معدومة، وهذه القيمة هي قيمة ثابت معادلة الانحدار a، وللتأكد من ذلك نعوض قيمة X بالصفر في معادلة الانحدار للمثال السابق فنجد :

$$\hat{y}_i = 20 + 2 * 0 = 20$$

وهذا يعني أنه عندما تكون كمية السماد معدومة أي تساوي الصفر، فإن الغلة بالمتوسط تساوي 20 وحدة، وهي متوسط الغلة في حالة عدم استخدام السماد. انظر الجدول رقم (8-16) .

C – تساعد معادلة مستقيم الانحدار على معرفة مقدار التغير في المتغير التابع (زيادة أو نقصان)، عندما يزداد المتغير المستقل وحدة واحدة، مع العلم أن التغير سيكون بالمتوسط، وللتأكد من ذلك انظر المثال السابق: فإذا استخدمنا 1 كغ سماد فإن متوسط الغلة سوف يعادل 22 وحدة، وتصبح الغلة بالمتوسط 24 وحدة إذا استخدمنا 2 كغ سماد، أي عندما يزداد المتغير المستقل وحدة واحدة، فإن المتغير التابع سوف يزداد بالمتوسط بمقدار 2 وحدة وهي قيمة b معامل الانحدار.

D – تبلغ القيمة المقدرة للمتغير التابع y_i قيمة الوسط الحسابي للمتغير التابع عندما تكون قيمة المتغير المستقل تساوي وسطه الحسابي $\bar{x}$ أي بمعنى أن مستقيم الانحدار يمر بالنقطة ($\bar{y}$ $\bar{x}$)، حيث يمكننا التأكد من ذلك من خلال المعادلة الآتية :

$$\bar{y} = a + b\ \bar{x}$$

E – تستخدم معادلة مستقيم الانحدار للتنبؤ بقيم المتغير التابع عندما يأخذ المتغير المستقل قيماً محددة، ولكن هنا يجب الانتباه إلى أي مدى يمكن أن تحافظ هذه العلاقة على نوعها وطبيعتها خارج حدود هذه البيانات، وبشكل عام يمكننا القول إنه يمكن استخدام المعادلة في حدود معينة من الثقة على طرفي مجال قيم المتغير المستقل .

اختبار الفرضيات

إن اختبار الفرضيات مسألة هامة جداً في التحليل الاحصائي ، لأنها تساعد على تفسير المعطيات و النتائج التجريبية و اتخاذ القرار المناسب لتلك المعطيات و نستند في اتخاذ القرار على الفرضية الصفرية (العدم) H_0 .

و هذه الفرضية لا تعترف بوجود الفروقات الحقيقية بين العينات أو بين متوسطات العينات المدروسة .

و للتحقق من صحة أي فرضية نقوم بسحب عينة عشوائية من المجتمع المدروس و نقوم بإجراء الحسابات اللازمة ثم نقارن القيمة التي حصلنا عليها من العينة مع القيمة النظرية التي نستخرجها من الجداول الخاصة باختبارات المعنوية ، و هذه الاختبارات متعددة و سنذكر منها فقط الاختبارات التالية :

1- اختبار t ستودنت .

2- اختبار كاي مربع X^2 .

3- اختبار فيشر F .

وهذه الاختبارات تسمى اختبارات المعنوية لأنها تستخدم لاختبار صحة الفرضية الصفرية (العدم) أو عدم صحتها و ذلك بمقارنة القيم الفعلية (المحسوبة) لهذه الاختبارات مع القيم النظرية التي نستخرجها من الجداول الخاصة بهذه الاختبارات بعد معرفة درجات الحرية و مستوى المعنوية المطلوب .

وقبل الحديث عن اختبارات المعنوية لا بد من توضيح بعض المفاهيم و العناوين و هي :

أولاً- الفرضية الصفرية (فرضية العدم) : و نرمز لها بالرمز H_0 .

عندما نريد اتخاذ قرار في أي موضوع من المواضيع قيد البحث من المفيد أن وضع فروض أو تخمينات عن المجتمع المدروس و هذه الفروض أو التخمينات قد تكون صحيحة أو غير صحيحة ، فمثلاً عندما نريد مقارنة عدة أسماء تجارية لمعالجة حالة صحية محددة بهدف اختبار الأفضل فالفرض الإحصائي يقول أن كل الأسماء التجارية المدروسة متساوية في العلاج أو لا يوجد فروقات بين تأثيرها أي الفرق بينها يساوي الصفر .

لنفترض أن هذه الأسماء التجارية عددها 3 فتكون الفرضية الصفرية كالتالي :

$$H_0 : \overline{X}_1 = \overline{X}_2 = \overline{X}_3$$

و كما هو واضح من العلاقة السابقة أو من الفرضية السابقة بأنه لا يوجد فروقات معنوية (حقيقية) بين تلك الأسماء التجارية و إن وجدت الفروقات فهي ليست حقيقية و إنما تعود للصدفة و العشوائية .

و بذلك يمكننا أن نقدم النص التالي للفرضية الصفرية التي تستخدم في جميع مجالات حياتنا مهما تكن .

بمعنى أن كل الباحثين في مختلف المجالات يضعون الفرضية الصفرية كأساس نظري لمقارنة نتائج أبحاثهم .

<u>الفرضية الصفرية تنص</u> : على عدم وجود فروقات معنوية (جوهرية أو حقيقية) بين متوسطات العينات المدروسة و إن وجدت بعض هذه الفروقات الظاهرية فهي ليست حقيقية و إنما تعود للصدفة و العشوائية أي لطريقة أخذ العينة و تسجيل القراءات أو بسبب حجم العينة أو لأي أسباب أخرى .

ثانياً- الفرضية البديلة (العكسية) : و نرمز لها بالرمز H_1 .

هذه الفرضية كما هو العنوان فرضية عكسية أي أننا نلجأ إليها عندما لا تتحقق الفرضية الصفرية ، أي عندما يكون المتوسط أكبر أو أصغر من الفرض و للتوضيح نأخذ المثال التالي : الفرضية الصفرية المتعلقة بمتوسط طول مجموعة من الطلاب تقول : $H_0: \overline{X} = 167cm$.

و الفرضية البديلة تقول أن متوسط طول هذه المجموعة هو أقل أو أكثر من 167 سم أو لا يساوي 167 سم

أي : $H_1: \overline{X} \neq 167cm$.

ثالثاً- مستوى المعنوية :

مستوى المعنوية في العرف الإحصائي يعبر عن درجة الاحتمال التي تقبل أو ترفض عنده الفرضية الصفرية و بتعبير آخر مستوى المعنوية يعبر عن درجة الدقة المطلوبة للنتائج أو درجة الخطأ المسموح فمثلاً عندما نقول إننا نريد دقة النتائج عند المستوى 5% فهذا يعني أن الخطأ المسموح هو 5% و دقة النتائج هي 95% و نفس الشيء يقال بالنسبة للمستوى 1% فالدقة المطلوبة هنا هي 99% و هكذا بالنسبة لبقية المستويات فكل باحث يختار مستوى المعنوية المناسب للظاهرة التي يدرسها فيمكن أن يختار مثلا واحد بالألف لأن النتائج المطلوبة تتطلب دقة عالية جداً و هذا يحدث مع الظواهر التي تتطلب مثل هذه الدقة لا سيما تلك الأشياء التي تتعلق بحياة الإنسان .

اختبارات المعنوية :

أولاً: اختبار t ستودنت:

هذا الاختبار يستخدم بكفاءة عالية عند مقارنة متوسطي عينتين أو متوسطي مجتمعين أو أي متوسطين ، و تحدد مجالات استخدامه في التالي :

أ- تقدير مجال الثقة - تقدير حدود المجتمع ...

ب- مقارنة متوسطي عينتين مستقلتين .

جـ- مقارنة متوسطي عينتين غير مستقلتين (مرتبطتين) .

أولاً- تقدير مجال الثقة :

إن العلاقة الرياضية المناسبة لإنشاء مجال الثقة يحوي القيمة الحقيقية لمتوسط المجتمع M تكون :

$$\overline{X} - \frac{\delta}{\sqrt{n}}(t_\alpha) < M < \overline{X} + \frac{\delta}{\sqrt{n}}(t_\alpha)$$

حيث أن الاحتمال $1-\alpha$ يساعدنا في حساب قيمة t_α النظرية (الجدولية) حيث α تمثل مستوى المعنوية الذي نختاره لدقة النتائج .

δ - الانحراف المعياري

$\overline{X}$ - المتوسط الحسابي للعينة

M - متوسط المجتمع المدروس .

t_α - هي القيمة النظرية التي نستخرجها من الجدول الخاص بتوزيع t عند مستوى المعنوية المحدد و أمام درجات الحرية n-1 .

و العلاقة الرياضية السابقة تفيدنا بمعرفة حدود العينة المدروسة أي حدها الأعلى و الأدنى أو حدود المجتمع المدروس بمعرفة حديه الأدنى و الأعلى .

فالحد الأدنى يساوي

$$\overline{X} - \frac{\delta}{\sqrt{n}}(t_{\alpha})$$

هذين الحدين كما يلاحظ من العلاقة السابقة

و الحد الأعلى يساوي

$$\overline{X} + \frac{\delta}{\sqrt{n}}(t_{\alpha})$$

و الحدين الأعلى و الأدنى تم حسابهما من العلاقة الخاصة بتوزيع ستودنت أي :

$$t = \frac{\overline{X} - M}{\frac{\delta}{\sqrt{n}}}$$

$\overline{X}$ متوسط العينة .

M متوسط المجتمع

δ الانحراف المعياري

و لتوضيح ذلك نأخذ المثال التالي :

مثال 1 :

سحبنا عينة مكونة من 25 طالب من طلاب كلية التربية و قمنا بحساب متوسط الوزن لهؤلاء الأشخاص فوجدناه 69 كغ بانحراف معياري قدره 3 كغ .

و **المطلوب** : حدد مجال الثقة لهذه العينة بدرجة ثقة تساوي 95% إذا علمت أن قيمة t النظرية عند المستوى 5% تساوي 2,064 .

الحل :

الحد الأدنى لوزن العينة =

$$\overline{X} - \frac{\delta}{\sqrt{n}}(t_{0,05}) = 69 - \frac{3}{\sqrt{25}}(2,064) = 67,76$$

الحد الأعلى لوزن العينة =

$$\overline{X} + \frac{\delta}{\sqrt{n}}(t_{0,05}) = 69 + \frac{3}{\sqrt{25}}(2,064) = 70,24$$

كما نلاحظ أن الحد الأدنى لأوزان هذه العينة هي 67,76 كغ و الحد الاعلى 70,24 كغ .

إذن $(67,76 < M < 70,24)$

أي أقل من 70,24 و أكثر من 67,76 .

مثال 2 :

إذا كان متوسط سكر الدم لجميع العاملين في إحدى كليات الجامعة يساوي 89 ، سحبنا عينة من 50 شخصاً فوجدنا أن متوسط سكر الدم لهؤلاء الأشخاص هو 94 بانحراف معياري قدره 2,5 .

و المطلوب :

هل هذه العينة تنتمي لمجتمع العاملين المذكورين أعلاه عند مستوى المعنوية 1% (درجة الثقة 99%) إذا علمت أن قيمة t عند المستوى 1% تساوي 2,58 .

الحل : الحد الأدنى للعينة السابقة :

$$\overline{X} - \frac{\delta}{\sqrt{n}}(t_{0,01}) = 94 - \frac{2,5}{\sqrt{50}}(2,58) = 93,08$$

الحد الأعلى لنفس العينة :

$$\overline{X} + \frac{\delta}{\sqrt{n}}(t_{0,01}) = 94 + \frac{2,5}{\sqrt{50}}(2,58) = 94,91$$

من هنا يتضح أن العينة ذات المتوسط 94 لا تتبع المتوسط 89 لأن حدي الثقة لمتوسط المجتمع التابعة له هذه العينة تقع ما بين 93,08 و 94,91 و لا تشمل القيمة 89 .

و يمكن أن نثبت هذا الكلام بطريقة أخرى و هي باستخدام اختبار t أي باستخدام العلاقة الرياضية الخاصة بهذا الاختبار في حال وجود عينة و مجتمع كما هو وارد في مثالنا السابق .

حيث : $\delta_{\overline{X}}$ تمثل الخطأ القياسي للعينة $t = \frac{\overline{X} - M}{\delta_{\overline{X}}}$ و تساوي $\frac{\delta}{\sqrt{n}}$

إذن :

$$t = \frac{94-89}{\frac{2,5}{\sqrt{50}}} = 14,28$$

و هي قيمة t الفعلية (المحسوبة) .

و هذه القيمة نقارنها مع القيمة النظرية (الجدولية) لهذا الاختبار و التي تساوي 2,58 .

نلاحظ أن القيمة الفعلية (المحسوبة) لـــ t أكبر من القيمة النظرية لذلك نرفض نص الفرضية الصفرية ($H_0 : M = \overline{X}$) الذي يقول أن متوسط العينة يساوي متوسط مجموعة العاملين في إحدى كليات الجامعة أو ينتمي له و بالتالي نقبل نص الفرضية البديلة ($H_1 : M \neq \overline{X}$) الذي يقول أن متوسط العينة لا يساوي متوسط

مجموعة العاملين إذن العينة لا تنتمي للمجتمع المذكور أعلاه أي لمجتمع مجموعة العاملين في إحدى كليات الجامعة .

ثانياً- مقارنة متوسطي عينتين مستقلتين :

نفترض أنه لدينا عينتين عشوائيتين غير مرتبطتين أي أن أفراد العينة الأولى يختلفون تماماً عن أفراد العينة الثانية

و عدد أفراد العينيتين n_1، n_2 تم اختيارهما من مجتمعين طبيعيين فالفرضية الصفرية لهاتين العينتين هي :

$H_0: \overline{X}_1 = \overline{X}_2$ و الفرضية البديلة $H_1: \overline{X}_1 \neq \overline{X}_2$ و العلاقة الرياضية لاختبار t في هذه الحالة تساوي :

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S_D}$$

S_D

حيث تمثل δ_D الخطأ القياسي و يتوقف حسابها على حجم العينة .

فإذا كان حجم العينة كبيراً (أكبر من 30) فتحسب δ_D من العلاقة :

$$S_D = \sqrt{\frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2}}$$

حيث : S_1^2 $\delta_1^2 = \frac{\sum\left(X_1 - \overline{X}_1\right)^2}{n_1}$ التباين للعينة الأولى .

S_2^2 $\delta_2^2 = \frac{\sum\left(X_2 - \overline{X}_2\right)^2}{n_2}$ التباين للعينة الثانية .

أما إذا كان حجم العينة صغيراً (30 و أقل) فإن قيمة الخطأ القياسي تحسب من العلاقة التالية :

$$S_D = \sqrt{\frac{\sum\left(X_1-\overline{X}_1\right)^2+\sum\left(X_2-\overline{X}_2\right)^2}{n_1+n_2-2}} \times \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

و يمكن أن تحسب $\sum\left(X_1-\overline{X}_1\right)^2$ و $\sum\left(X_2-\overline{X}_2\right)^2$ بطريقة تربيع القيم التالية :

$$\sum\left(X_1-\overline{X}_1\right)^2=\sum X_1^2-\frac{\left(\sum X_1\right)^2}{n_1}$$

$$\sum\left(X_2-\overline{X}_2\right)^2=\sum X_2^2-\frac{\left(\sum X_2\right)^2}{n_2}$$

فهنا نختار الطريقة التي نراها أسهل .

<u>مثال للعينات الصغيرة :</u>

إحدى مستشفيات التوليد الخاصة

قمنا بزيارة قسم التوليد في [redacted] و سجلنا أوزان ستة أطفال حديثي الولادة و كانت كما يلي :

X_1 : 3,8 3,7 2,9 3,5 2,6 3,3

و ذهبنا إلى المشفى الوطني و بقسم التوليد أيضاً سجلنا أوزان ثمانية أطفال :

X_2 : 3,7 4,6 5,4 6,2 4,2 3,5 5,3 5,5

و المطلوب :

هل يوجد اختلاف معنوي (حقيقي) بين متوسطي العينتين إذا علمت أن قيمة t عند المستوى 1% تساوي 3,055 .

الحل :

$$t = \frac{\overline{X_1} - \overline{X_2}}{S_D}$$

بما أن العينة صغيرة فإن قيمة الخطأ القياسي S_D و تحسب من العلاقة :

$$S_D = \sqrt{\frac{\sum\left(X_1 - \overline{X}_1\right)^2 + \sum\left(X_2 - \overline{X}_2\right)^2}{n_1 + n_2 - 2}} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

و لحساب هذه الانحرافات ننشئ الجدول التالي :

X_1	$(X_1 - \overline{X}_1)$	$(X_1 - \overline{X}_1)^2$	X_1^2	X_2	$(X_2 - \overline{X}_2)$	$(X_2 - \overline{X}_2)^2$	X_2^2
3,8	0,5	0,25	14,44	3,7	−1,1	1,21	13,69
3,7	0,4	0,16	13,69	4,6	0,2	0,04	21,16
2,9	−0,4	0,16	8,41	5,4	0,6	0,36	29,16
3,5	0,2	0,04	12,25	6,2	1,4	1,96	38,44
2,6	−0,7	0,49	6,76	4,2	−0,6	0,36	17,64
3,3	0	0	10,89	3,5	1,3	1,69	12,25
				5,3	0,5	0,25	28,09
				5,5	0,7	0,49	30,25
19,8	0	1,1	66,44	38,4		6,36	190,68

إذن يمكن أن نطبق إحدى الطريقتين الانحرافات أو تربيع القيم بعد أن قمنا بحساب متوسط العينة الأولى .

$$\overline{X}_1 = \frac{\sum X_1}{n_1} = \frac{19,8}{6} = 3,3$$

$$\overline{X}_2 = \frac{\sum X_2}{n_2} = \frac{38,4}{8} = 4,8$$

$$\sum\left(X_1 - \overline{X}_1\right)^2 = \sum X_1^2 - \frac{\left(\sum X_1\right)^2}{n_1} = 66,44 - \frac{(19,8)^2}{6} = 1,1$$

$$\sum\left(X_2 - \overline{X}_2\right)^2 = \sum X_2^2 - \frac{\left(\sum X_2\right)^2}{n_2} = 190,68 - \frac{(38,4)^2}{8} = 6,36$$

$$\times\sqrt{\frac{1}{6}+\frac{1}{8}}$$

$$S_D = \sqrt{\frac{1,1+6,36}{6+8-2}} = 0,426$$

$$t = \frac{3,3-4,8}{0,426} = -3,\text{[redacted]}52$$

و الآن نقارن قيمة t الفعلية $|-3,52|$ بالقيمة المطلقة مع قيمة t النظرية 3,055 نلاحظ أن قيمة t المحسوبة أكبر من قيمة t النظرية لذلك نرفض نص الفرضية الصفرية $(H_0: \overline{X}_1 = \overline{X}_2)$ الذي يقول بعدم وجود فروقات معنوية (حقيقية) بين متوسطي العينتين و بالتالي نقبل نص الفرضية البديلة $(H_1: \overline{X}_1 \neq \overline{X}_2)$

(إذن يوجد فرق معنوي (حقيقي) أو جوهري بين متوسط أوزان الأطفال في

المشفى العسكري

~~مستشفى الجامعي~~ و أوزان الأطفال في المشفى الوطني .

مثال آخر للعينات الكبيرة :

لدينا العينتين التاليتين : الأولى و هي الشاهد و عدد أفرادها 50 و متوسطها الحسابي 10,51 و تباينها 7,99 و عدد أفراد العينة الثانية 64 و متوسطها الحسابي يساوي 9,23 و التباين 7,23

السؤال :

هل يوجد اختلاف معنوي بين متوسطي العينتين إذا علمت أن قيمة t النظرية (الجدولية) تساوي 1,96 عند المستوى 5% .

الحل :

$$S_D = \sqrt{\frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2}}$$

بما أن n_1+n_2 كبيراً لذلك S_D تحسب من العلاقة التالية :

و في نص المسألة :

$$\left[\delta_1^2 = 7,99\right] و \left[\overline{X}_1 = 10,51\right] و \left[n_1 = 50\right]$$

$$\left[\delta_2^2 = 7,23\right] و \left[\overline{X}_2 = 9,23\right] و \left[n_2 = 64\right]$$

$$S_D = \sqrt{\frac{7,99}{50} + \frac{7,23}{64}} = 0,52$$

$$t = \frac{10,51 - 9,23}{0,52} = 2,46$$

و هي قيمة t الفعلية

و للإجابة عن السؤال السابق نقارن قيمة t المحسوبة 2,46 مع قيمة t النظرية (الجدولية) 1,96 نلاحظ أن قيمة t المحسوبة أكبر من قيمة t النظرية لذلك

نرفض نص الفرضية الصفرية $(H_0: \overline{X}_1 = \overline{X}_2)$ الذي يقول بعدم وجود فروقات معنوية بين متوسطي العينتين أو بتساوي متوسطي العينتين و بالتالي نقبل نص الفرضية البديلة $(H_1: \overline{X}_1 \neq \overline{X}_2)$

إذن يوجد فرق جوهري (معنوي) بين متوسطي العينتين المدروستين أي يوجد اختلاف بين المتوسطين .

مثال 1 :

أجريت معايرة كمية الخضاب في الدم لعينة مؤلفة من 48 طفلاً فكان متوسط هذه العينة 11,3 مختارة من مجتمع انحرافه المعياري 2,5 غ و المطلوب : أوجد مجال الثقة لمجتمع الأطفال الذي أخذت منه هذه العينة إذا علمت أن قيمة t النظرية بدرجة ثقة 95% تساوي 1,96 أو يمكن أن نطرح السؤال بشكل آخر وهو أوجد حدود المجتمع الذي أخذت منه العينة أو حدود العينة بنفس الشروط السابقة .

الحل :

مجال الثقة المطلوب هو :

$$L_1 = \overline{X} - \frac{\delta}{\sqrt{n}}(t_{0,05})$$

$$L_2 = \overline{X} + \frac{\delta}{\sqrt{n}}(t_{0,05})$$

و منه نجد أن الحد الأدنى يساوي :

$$L_1 = 11,3 - \frac{2,5}{\sqrt{49}}(1,96) = 10,6$$

$$L_2 = 11,3 + \frac{2.5}{\sqrt{49}}(1,96) = 12$$

إذن مجال الثقة في مجتمع الأطفال هو [10,6 و 12] أو بمعنى آخر أن حدود مجتمع العينة يقع بين 10,6 و 12 .

مثال 2 :

لدى قياس معدل البولة في الدم لـ 25 شخصا مختارين من مجتمع إحصائي له التوزيع الطبيعي و وجدنا أن المتوسط الحسابي لمعدل البولة في دم العينة يساوي $\overline{X}=0,29g/l$ و الانحراف المعياري $\delta=0,05g/l$ و المطلوب : إيجاد مجال الثقة لمتوسط مجتمع الأشخاص الذي أخذت منه هذه العينة إذا علمت أن قيمة t النظرية بدرجة ثقة 90% تساوي 1,711 .

الحل :

مجال الثقة هو :

$$L_1=\overline{X}-\frac{\delta}{\sqrt{n}}(t_{0.01})$$

$$L_2=\overline{X}+\frac{\delta}{\sqrt{n}}(t_{0.01})$$

و الآن نجد الحد الأدنى لمجتمع العينة :

$$L_1=0,29-\frac{0,05}{\sqrt{25}}(1,711)=0,288$$

$$L_2=0,29+\frac{0,05}{\sqrt{25}}(1,711)=0,291$$

إذن مجال الثقة في مجتمع العينة هو [0,288 و 0,291]

أو بمعنى آخر أن حدود مجتمع العينة يقع بين 0,288 و 0,291 .

مثال :

إذا كان متوسط عمر صمام في جهاز طبي من نوع ألفا هو 6,5 سنة بانحراف معياري قدره 0,9 سنة بينما متوسط عمر صمام آخر في جهاز طبي من نوع بيتا هو 6 سنة بانحراف معياري 0,8 سنة .

و المطلوب :

اختبار معنوية الفروق بين متوسط عمر عينة الصمامات من نوع ألفا التي حجمها 36 و متوسط عمر عينة الصمامات من نوع بيتا و حجمها 49 بدرجة ثقة 95%.

الحل :

لدينا المعطيات التالية :

$n_1 = 36 \quad , \quad \overline{X_1} = 6.5 \quad , \quad \delta_1 = 0.9$

$n_2 = 49 \quad , \quad \overline{X_2} = 6 \quad , \quad \delta_2 = 0,8$

نلاحظ أن عدد أفراد العينة الأولى و الثاني كبير لذلك سنقوم بحساب الخطأ القياسي بما يتناسب و العينات الكبيرة إذن :

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S_D} , S_D = \sqrt{\frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2}} = \sqrt{\frac{(0,9)^2}{36} + \frac{(0,8)^2}{49}} = 0,189$$

ملاحظة : نلاحظ من العلاقة السابقة أننا قمنا بتربيع قيمة الانحراف المعياري لكلا العينتين و هذا سببه أننا نريد التباين و ليس الانحراف المعياري كما هو واضح في العلاقة السابقة حيث يرمز للتباين بـــ δ^2

$$t = \frac{6,5 - 6}{0,189} = 2,64$$

ونعود لحساب t الفعلية من العلاقة السابقة :

و نستخرج قيمة t النظرية من الجدول بدرجة ثقة 95% و أمام درجات الحرية :

$$n_1 + n_2 - 2 = 36 + 49 - 2 = 83$$

فنجد أن قيمة t الفعلية (المحسوبة) 2,64 مع قيمة t النظرية 1,96 نلاحظ أن قيمة t الفعلية أكبر من قيمة t النظرية لذلك نرفض نص الفرضية الصفرية $(H_0: \overline{X_1} = \overline{X_2})$ الذي ينص بعدم وجود فروقات معنوية (جوهرية) بين متوسط عمر الصمام من نوع ألفا و عمر الصمام من نوع بيتا و بالتالي نقبل نص الفرضية البديلة $(H_1: \overline{X_1} \neq \overline{X_2})$ إذن يوجد فرق معنوي (حقيقي و جوهري) بين متوسط عمر الصمامين .

مثال :

أجريت دراسة على عينتين من الأطفال المولودين حديثاً إحداهما من الذكور و الأخرى من الإناث حجمها على التوالي $n_1=n_2=10$ و وجد أن متوسط وزن الطفل في العينة الأولى (ذكور) = 3,5 و متوسط وزن العينة الثانية (إناث) = 3 و تباين العينة الأولى 25 و الثانية 22 .

و المطلوب :اختبار معنوية الفروق بين متوسطي العينتين إذا علمت أن t تحسب عند المستوى 0,05% (بدرجة ثقة 95%) .

الحل :

المعطيات المتوفرة بنص المسألة : $n_1=10$ ، $\overline{X_1} = 3,5$ ، $\delta_1^2 = 25$

$n_2 = 10$ ، $\overline{X_2} = 3$ ، $\delta_2^2 = 22$

و كما نلاحظ أن عدد أفراد العينتين صغير لذلك سنقوم بحساب الخطأ القياسي S_D بما يتوافق و العينات الصغيرة :

$$t=\frac{\overline{X_1}-\overline{X_2}}{S_D},S_D=\sqrt{\frac{\sum\left(X_1-\overline{X_1}\right)^2+\sum\left(X_2-\overline{X_2}\right)^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}=\sqrt{\frac{\delta_1^2}{n_1}+\frac{\delta_2^2}{n_2}}=2,168$$

لعدم توفر البيانات المطلوبة طبقنا قانون العينات الأسرية

إذن :

$$t=\frac{3,5-3}{2,168}=0,23$$

و هي قيمة t الفعلية .

نقارن قيمة t الفعلية 0,23 مع قيمة t النظرية بدرجة ثقة 95% و درجات الحرية 18 فنجدها تساوي 1,734 نلاحظ أن قيمة t الفعلية أصغر من القيمة النظرية لذلك نقبل نص الفرضية الصفرية الذي يقول بعدم وجود فروقات بين متوسطي الوزن في كلتا العينتين $\left(H_0:\overline{X_1}=\overline{X_2}\right)$ و بالتالي نرفض نص الفرضية البديلة .

إذن لا يوجد فرق معنوي (جوهري) حقيقي بين متوسط وزن عينة الذكور و متوسط وزن عينة الإناث و الفرق الموجود (الظاهري) بين العينتين غير حقيقي و إنما يعود للصدفة و العشوائية .

تطبيقات غير محلولة :

مثال 1:

قام أحد الباحثين بأخذ عينة حجمها n=33 من مياه بحيرة سد 16 تشرين و قام بقياس كمية الصوديوم في هذه العينة فوجد أن متوسط وزن الصوديوم في هذه العينة 24,11 ملغ/ليتر و الانحراف المعياري 2,44 ملغ/ليتر

و المطلوب :

إيجاد مجال الثقة لهذه العينة (حدي مجتمع العينة) بدرجة ثقة 95% (قيمة t النظرية عند المستوى 5% = 1,96) .

مثال 2:

أجري اختبار في دار التوليد بدمشق لمعايرة التيروكسين لدى 49 وليداً فوجدنا أن المتوسط الحسابي لهذه العينة $\overline{X} = 9,8$ بانحراف معياري قدره 3,1

والمطلوب :

تقدير حدود الثقة لمجتمع العينة (مجال الثقة) إذا علمت أن قيمة t تحسب عند المستوى 0,1% (درجة الثقة 90%) علماً أن قيمة t النظرية بدرجة ثقة 90 % = 1,296

مثال 3:

مجموعتين من المرضى ، تتكون الأولى من 50 شخصاً و الثانية من 100 شخص ، المجموعة الأولى أعطيت نوعاً جديداًمن الحبوب المنومة فكان وسط ساعات النوم 7,82 بانحراف 0,25 ساعة و أعطيت المجموعة الثانية نوعا

معروفاً من الحبوب فكان متوسط ساعات النوم 6,75 ساعة بانحراف معياري 0,3 ساعة .

و المطلوب :

اختبار معنوية الفرق بين نوعي الحبوب المستخدمة بمستوى 0,05 (قيمة t عند المستوى 0,05 = 1,96) .

مثال 4:

لدينا العينتين التاليتين اللتين تمثلان مستوى التحصيل لدى الطلبة :

X_1 : 15 14 10 9 11 16

X_2 : 20 19 14 15 16 17

و المطلوب :

اختبار معنوية الفرق بين مستوى التحصيل لكلا المجموعتين إذا علمت أن :

قيمة t تحسب بدرجة ثقة 95% و تساوي 2,179 .

ثالثاً- مقارنة متوسطي عينتين غير مستقلتين (مرتبطتين) :

في كثير من الأحيان نأخذ عينة واحدة و نسجل المشاهدات لمفرداتها ، ثم تخضع هذه العينة تحت تأثير معين و نعود لتسجيل الجديدة لمفردات العينة نفسها و بعملية المقارنة لهاتين القراءتين لمفردات العينة يمكن أن نستنتج مدى تأثير هذا المؤثر الذي وضعت العينة تحته .

و هذا الاختبار يستخدم كثيراً في المجالات المختلفة لحياة الإنسان ، فمثلاً يستخدم في مجال تجارب العقاقير الطبية لبيان تأثير عقار معين قبل استخدامه و بعده و

كذلك تأثير برنامج غذائي معين على زيادة الوزن و أشياء أخرى كثيرة و يتم حساب قيمة الاختبار الفعلية كما يلي :

$$t = \frac{\bar{d}}{\delta_{\bar{d}}}$$

حيث : $\bar{d}$ تمثل المتوسط الحسابي للفرق بين قيم أفراد العينة المدروسة قبل و بعد الاختبار (التأثير)

$\delta_{\bar{d}}$ الخطأ القياسي للفرق و يتم حساب المتوسط السابق و الخطأ القياسي للفرق كما يلي :

1- نحسب الفرق بين كل زوج مقابل من أزواج القيم (أفراد العينة قبل و بعد التأثير)

$$\sum (X_1 - X_2) = \sum d$$

2- نحسب متوسط الفرق :

$$\bar{d} = \frac{\sum d}{n}$$

3- نحسب تباين الفرق :

$$\delta_d^2 = \frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2 \text{ أو } \delta_d^2 = \frac{\sum \left(d - \bar{d}\right)^2}{n} = \frac{\sum d^2 - \frac{\left(\sum d\right)^2}{n}}{n}$$

4- الخطأ القياسي للفرق :

$$\delta_{\bar{d}} = \sqrt{\frac{\delta_d^2}{n}}$$

5- نقارن قيمة t المحسوبة مع t النظرية (الجدولية) عند مستوى المعنوية المطلوب و أمام درجات الحرية $n-1$ ثم نأخذ القرار الإحصائي بقبول الفرضية الصفرية أو برفضها .

مثال :

أخذت عينة من عشرة أطفال و سجلت أوزانها الأولية ثم أعطيت وجبة غذائية محددة و لمدة خمسة عشر يوما ، ثم سجلت أوزانها من جديد و كانت النتائج على الشكل التالي :

الوزن قبل التغذية X_1	17	17,5	18	18,5	19	19,5	20	21	21	21,5	المجموع
الوزن بعد التغذية X_2	17,7	18,5	18,9	19,3	20,1	19,9	20,9	21,7	21,7	22,3	-
$d=X_1-X_2$	-0,7	-1	-0,9	-0,8	-1,1	-0,4	-0,9	-0,7	-0,7	-0,8	-8
d^2	0,49	1	0,81	0,64	1,21	0,16	0,81	0,49	0,49	0,64	6,74

و المطلوب : اختبار معنوية الفروق بين الوزنين بدرجة ثقة 95% .

الحل : الفروض

$$H_0 : \bar{d} = 0$$
$$H_1 : \bar{d} \neq 0$$

$$\bar{d} = \frac{\sum d}{n} = \frac{X_1 - X_2}{n} = \frac{-8}{10} = -0,8$$

نحسب تباين الفرق :

$$\delta_d^2 = \frac{\sum d^2 - \frac{\left(\sum d\right)^2}{n}}{n} = \frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2 = \frac{6,74 - \frac{(-8)^2}{10}}{10} = \frac{6,74}{10} - \left(\frac{-8}{10}\right)^2 = 0,034$$

إذن الخطأ يساوي :

$$\delta_{\bar{d}} = \sqrt{\frac{\delta_d^2}{n}} = \sqrt{\frac{0,034}{10}} = 0,058$$

نحسب قيمة t الفعلية :

$$t = \frac{\bar{d}}{\delta_{\bar{d}}} = \frac{-0,8}{0,058} = -13,79 = |13,79|$$

نضع الفروض

$$H_0 : \bar{d} = 0$$

$$H_1 : \bar{d} \neq 0$$

القرار الإحصائي : نقارن قيمة t المحسوبة (الفعلية) 13,79- بالقيمة المطلقة مع قيمة t الجدولية (النظرية) عند المستوى 0,05% و درجات الحرية n-1 أي : 9=1-10 فتساوي 2,26 فنلاحظ أن t المحسوبة أكبر من t الجدولية لذلك نص الفرضية الصفرية الذي يقول بعدم وجود فروقات معنوية أي عدم وجود تأثير للبرنامج الغذائي على زيادة الوزن و بالتالي نقبل الفرضية البديلة إذن يوجد فرق معنوي اي يوجد تأثير للبرنامج الغذائي (الوجبة الغذائية) على زيادة أوزان الأطفال .

توزيع كاي مربع X^2

كثيراً ما يحتاج الباحث لمعرفة مدى مطابقة نتائج حصل عليها بطريقة الملاحظة أو التجربة و نتائج أخرى يمكن الحصول عليها بالطرق النظرية لنفس الظاهرة .

فاختبار كا مربع أو اختبار المطابقة يجري بين مؤشر تجريبي و الآخر نظري و عادة النتائج التجريبية ليس من الضروري أن تتفق مع النتائج المتوقعة وفقاً لقواعد الاحتمال أو وفقا للتوزيع النظري ، فعلى سبيل المثال إن الاحتمالات النظرية تؤدي بنا إلى توقع 50 أنثى و 50 ذكر في 100 ولادة و الحقيقة أنه من النادر الحصول على هذه النتيجة بالضبط .

و هذا الاختبار يستعمل بشكل واسع في التحليل الوراثي و الطبي و مجالات كثيرة و متعددة و نرمز للتكرارات الفعلية (التجريبية) بالرمز [O]و التكرارات النظرية أو المتوقعة [E] و بذلك تكون معادلة الاختبار :

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

حيث : χ^2 – كاي مربع

O – المشاهدات الفعلية (الحقيقية)

E – المشاهدات النظرية (المتوقعة)

- القيمة الناتجة لـــ χ^2 تقارنها مع القيمة النظرية التي نستخرجها من الجدول الخاص بهذا الاختبار (توزيع كا مربع) و الموجود في كتب الاحصاء ، على أساس درجات الحرية k-1 (حيث k تمثل عدد المجموعات المدروسة) ومستوى المعنوية المطلوب أي درجة الثقة المطلوبة (درجة الدقة التي نختارها لنتائج البحث) .

ثم نجري المقارنة بين قيمة χ^2 الفعلية (المحسوبة) من العلاقة الخاصة بهذا الاختبار و الموجودة أعلاه مع القيمة النظرية (الجدولية) لهذا الاختبار .

فإذا كانت قيمة كاي مربع χ^2 الفعلية أكبر أو تساوي قيمة كاي مربع النظرية نرفض الفرضية الصفرية $(H_0:O=E)$ و التي تقول بعدم وجود فروقات معنوية (حقيقية) بين التكرارات التجريبية (الفعلية) و التكرارات النظرية و بالتالي نكون قد قبلنا الفرضية البديلة $(H_1:O\neq E)$.

أما إذا كانت قيمة كاي مربع الفعلية (المحسوبة) أصغر من قيمة كاي مربع النظرية (الجدولية) نقبل الفرضية الصفرية $(H_0:O=E)$ القائلة بعدم وجود فروقات معنوية (جوهرية) و نرفض الفرضية البديلة (العكسية) $(H_1:O\neq E)$ إذن لا يوجد فرق معنوي بين التكرارات الفعلية و التكرارات النظرية .

مثال 1 :

بالعودة إلى سجلات أحد مشافي التوليد وجدنا توزيع 1000 طفل مولود تبعاً للجنس كالآتي : 560 إناث و 440 ذكور .

و السؤال :

هل هده النتيجة تؤيد أو تؤكد الفرضية القائلة بتساوي الذكور و الإناث عند الولادة بدرجة ثقة 95% .

الحل :

الفرضية الصفرية و التي تقول بتساوي ولادة الذكور و الإناث $H_0:O=E$

O – تمثل المشاهدات الفعلية كما هو وارد في نص المسألة .

E – المشاهدات النظرية حسب احتمال ولادة الذكور و الإناث .

و الفرضية البديلة (العكسية) و التي تقول بعدم تساوي ولادة الذكور و الإناث $H_1:O\neq E$

إذن التوزيع الفعلي هو كالتالي : O = 560 إناث + 440 ذكور

و التوزيع النظري هو : E = 500 إناث + 500 ذكور (حسب احتمال ولادة الذكور و الإناث) .

و الآن نطبق العلاقة كاي مربع :

$$\chi^2=\sum\frac{(O-E)^2}{E}=\frac{(560-500)^2}{500}+\frac{(440-500)^2}{500}=14,4$$

و هي القيمة الفعلية .

و الآن نستخرج من جدول كاي مربع القيمة النظرية عند المستوى 5% و درجات الحرية k-1=2-1=1 فنجدها 3,841 بينما قيمة كاي مربع الفعلية 14,4 و هي أكبر من القيمة النظرية (الجدولية) و هذا يؤدي إلى رفض الفرضية الصفرية السابقة و القائلة بعدم وجود فروقات معنوية (حقيقية) بين ولادة الذكور و الإناث و بالتالي نقبل نص الفرضية القائلة بعدم تساوي ولادة الذكور و الإناث أي وجود اختلاف جوهري (معنوي) بين ولادة الذكور و الإناث و الاختلاف الموجود هو اختلاف حقيقي و لا يعود للصدفة و العشوائية كما تنص الفرضية الصفرية .

مثال 2 :

أثبتت الدراسات في العام الماضي أن 65% من السكان يفضلون الملابس التي تصنع محلياً و بعد ذلك أدخلت الكثير من التعديلات لتحسين جودة الإنتاج و صار هناك شعور بأن هذه النسبة قد زادت و للتأكد من صحة الشعور أو عدمه أخذت

عينة حجمها 300 شخص و اتضح أن هناك 188 شخص يستعملون الملابس المنتجة محلياً .

السؤال :

هل هناك ما يؤكد تصور المنتجين المحليين أو هل يتطابق ما يتصوره المنتجين المحليين مع نتائج العينة باحتمال قدره 5% .

الحل :

الفرضية الصفرية $H_0:O=E$ و الفرضية البديلة $H_1:O\neq E$

إذن التوزيع الفعلي هو كالتالي : O = 188 يستعملون الملابس المحلية + 122 يستعملون الملابس المستوردة و التوزيع النظري هو :

E = الملابس المحلية ($195=0,65\times300$) + الملابس المستوردة (105) .

و الآن نطبق العلاقة كاي مربع :

$$\chi^2=\sum\frac{(O-E)^2}{E}=\frac{(188-195)^2}{195}+\frac{(122-105)^2}{105}=0,718$$

و هي القيمة الفعلية

و الآن نستخرج من جدول كاي مربع القيمة النظرية $\chi^2=3,841$

و الآن نقارن قيمة χ^2 الفعلية 0,718 مع القيمة النظرية 3,841 فنجد أن القيمة الفعلية (المحسوبة) أصغر من القيمة النظرية لذلك نقبل نص الفرضية الصفرية الذي يقول بأن 65% من السكان يفضلون الملابس المحلية و نرفض الفرضية البديلة التي تقول بأن النسبة أكثر من ذلك .

إذن عند إجراء التطابق بين الدراسات التي أجريت في الماضي و التي أثبتت أن 65% من السكان يفضلون الملابس المصنوعة محلياً وبين الشعور بزيادة النسبة نبين أن النسبة 65% ما زالت قائمة و لم تتغير نسبة استعمال الناس للملابس المحلية .

اختبار الاستقلالية أو اختبار χ^2 لبيانات جدول التوافق :

هو نفس اختبار التطابق أو اختبار كاي مربع إلا أنه يرتب النتائج (البيانات الإحصائية) في صفوف و أسطر أي ضمن جداول فهنا ندرس مدى استقلالية صفتين أو خاصتين فالخاصة الأولى نضعها ضمن الأسطر و الخاصية الثانية ضمن الأعمدة و نطبق العلاقة المعروفة لاختبار كاي مربع :

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

مثال :

قام فريق من الأطباء بدراسة لمعرفة آثار التدخين على الصحة العامة و تم اختبار عينتين تضم الأولى 425 من المدخنين و الثانية 425 من غير المدخنين و بعد إجراء الفحوصات اللازمة توصل فريق الأطباء إلى النتائج التالية :

الحالة الصحية / العينة	جيدة	غير جيدة	المجموع
يدخن	311	114	425
لا يدخن	375	50	425
المجموع	686	164	850

و المطلوب :

اختبار مدى تأثير التدخين على الصحة العامة بدرجة ثقة 99%

الحل :

لدينا توزيعين التوزيع الأول يتعلق بالنتائج التجريبية أي النتائج التي توصل إليها فريق الأطباء وهي موجودة في الجدول أعلاه .

أما النتائج النظرية فنحن من سيقوم بحسابها ثم نضعها في جدول كالجدول السابق و بذلك سيصبح لدينا جدولين الأول يحتوي النتائج الفعلية و الجدول الثاني النتائج النظرية .

- الأشخاص اللذين يدخنون و حالتهم الصحية جيدة = $\frac{686 \times 425}{850} = 343$ شخص

كما لاحظنا قمنا بضرب مجموع السطور الذي يحتوي على المدخنين و هو 425 بمجموع الأعمدة الذي يحوي المدخنين أيضاً و هو 686 و قسمنا على المجموع العام .

- الأشخاص اللذين يدخنون و حالتهم الصحية غير جيدة = $\frac{164 \times 425}{850} = 82$ شخصاً

قمنا بضرب مجموع السطور الذي يحوي على غير المدخنين 425 و مجموع العمود الذي يحوي غير المدخنين أيضا 164 و قسمنا على المجموع العام .

- الأشخاص اللذين لا يدخنون و حالتهم الصحية جيدة = $\frac{686 \times 425}{850} = 343$

- الأشخاص اللذين يدخنون و حالتهم الصحية غير جيدة = $\frac{164 \times 425}{850} = 82$

و بعدها ننشئ الجدول التالي :

العينة \ الحالة الصحية	جيدة	غير جيدة	المجموع
يدخن	343	82	425
لا يدخن	343	82	425
المجموع	686	164	850

نتائج الفحوصات نظرياً

ثم نطبق العلاقة :

$$\chi^2=\sum\frac{(O-E)^2}{E}=\frac{(311-343)^2}{343}+\frac{(114-82)^2}{82}+\frac{(375-343)^2}{343}+\frac{(50-82)^2}{82}=30,96$$

و هي القيمة الفعلية

و الآن نستخرج القيمة النظرية لكاي مربع من الجدول الخاص بهذا الاختبار بعد معرفة عدد درجات الحرية من العلاقة

(عدد الأسطر - 1) × (عدد الأعمدة - 1) = (2-1) (2-1) = 1

و مستوى المعنوية 0,01% تكون قيمة كاي مربع الجدولية

الفروض الإحصائية : الصفرية H_0:O=E $\chi^2=6,635$

البديلة H_1:O≠E

القرار الإحصائي : إن قيمة χ^2 الفعلية 30,96 أكبر من قيمة χ^2 الجدولية (النظرية) 6,635 لذلك نرفض الفرضية الصفرية القائلة بعدم وجود تأثير

للتدخين على الصحة العامة و بالتالي نقبل بالفرضية البديلة القائلة بوجود تأثير للتدخين على الصحة العامة ، إذن يوجد تأثير معنوي (جوهري) حقيقي للتدخين على الصحة العامة .

مثال : في دراسة لمعرفة مدى الارتباط بين الحالة الزوجية و القدرة على الادخار فقام فريق من الباحثين بأخذ عينة مؤلفة من 880 أعزب و عينة أخرى مؤلفة من 2447 متزوج و سجلت النتائج التالية :

الحالة الزوجية	عدد المدخرين	غير المدخرين	المجموع
أعزب	490	390	880
متزوج	1552	895	2447
المجموع	2042	1285	3327

و المطلوب : اختبار الفرض القائل بأن سلوك المتزوجين في الادخار لا يختلف اختلافاً جوهرياً عن سلوك غير المتزوجين بدرجة ثقة 95% .

الحل : ننشئ جدول مماثل للجدول السابق يتعلق بالقيم النظرية (النتائج النظرية) كما يلي :

عدد المدخرين من غير المتزوجين =

$$\frac{\text{(مجموع السطر الذي يحوي غير المتزوجين} \times \text{مجموع العمود الذي يحوي المدخرين)}}{\text{المجموع العام}}$$

عدد المدخرين غير المتزوجين = $\frac{2042 \times 880}{3327} = 540,1$

عدد غير المدخرين من غير المتزوجين = $\frac{1285 \times 880}{3327} = 339,3$

عدد المدخرين من المتزوجين = $\frac{2042 \times 2447}{3327} = 1501,9$

عدد غير المدخرين من المتزوجين = $\frac{1285 \times 2447}{3327} = 945,1$

و الآن نضع النتائج السابقة في جدول مشابه للجدول السابق :

الحالة الزوجية	عدد المدخرين	غير المدخرين	المجموع
أعزب	540,1	339,9	880
متزوج	1501,9	945,1	2447
المجموع	2042	1285	3327

و نطبق العلاقة كاي مربع :

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(490-540,1)^2}{540,1} + \frac{(390-339,9)^2}{339,9} + \frac{(1552-1501,9)^2}{1501,9} + \frac{(895-945,1)^2}{945,1} = 16,36$$

و هي قيمة χ^2 الفعلية

نستخرج من الجدول قيمة χ^2 النظرية أمام درجات الحرية

(عدد الأسطر −1) × (عدد الأعمدة −1) = (2−1) × (2−1) = 1

و مستوى المعنوية 0,05% فنجدها تساوي 3,841 .

الفروض الاحصائية : الفرضية الصفرية $H_0:O=E$

الفرضية البديلـة $H_1:O\neq E$

القرار الاحصائي : نلاحظ أن قيمة χ^2 الفعلية 16,34 أكبر من قيمة χ^2 النظرية لذلك نرفض نص النظرية الصفرية القائلة بأن سلوك غير المتزوجين في الادخار لا يختلف اختلافا معنويا (جوهريا) عن سلوك المتزوجين و بالتالي نقبل نص الفرضية البديلة (العكسية) القائلة بأن سلوك غير المتزوجين في الادخار يختلف اختلافاً معنوياً (جوهرياً) عن سلوك المتزوجين .

مثال محلول :

أردنا اختبار فعالية أحد الأدوية الجديدة في شفاء الأمراض فأخذنا 200 مريض و قسمناهم إلى مجموعتين الأولى 100 مريض و أعطيناهم الدواء و الثانية 100 مريض و لم نعطهم الدواء (مجموعة المراقبة) و بعد فترة محددة من الزمن سجلنا في جدول خاص عدد الأشخاص اللذين تعافوا من المرض و اللذين لم يتعافوا من المرض من كلتا المجموعتين

العينة	شفوا	لم يشفوا	المجموع
المجموعة الأولى	75	25	100
المجموعة الثانية	65	35	100
المجموع	140	60	200

و **المطلوب** : اختبار الفرض القائل بعجم فعالية الدواء بدرجة ثقة 99% .

الحل :

يجب أن ننشئ جدول للنتائج النظرية للاختبار :

- الأشخاص اللذين شفوا من المجموعة الأولى = $\frac{140\times100}{200}=70$ شخص

- الأشخاص اللذين لم يشفوا من المجموعة الأولى = $\frac{60\times100}{200}=30$ شخص

- الأشخاص اللذين شفوا من المجموعة الثانية = $\frac{140\times100}{200}=70$ شخص

- الأشخاص اللذين لم يشفوا من المجموعة الثانية = $\frac{60\times100}{200}=30$ شخص

ننشئ الجدول :

العينة	شفوا	لم يشفوا	المجموع
المجموعة الأولى	70	30	100
المجموعة الثانية	70	30	100
المجموع	140	60	200

نطبق العلاقة :

$$\chi^2=\sum\frac{(O-E)^2}{E}=\frac{(75-70)^2}{70}+\frac{(25-30)^2}{30}+\frac{(65-70)^2}{70}+\frac{(35-30)^2}{30}=2,38$$

و هي القيمة الفعلية لـــ χ^2 .

نستخرج من الجدول الخاص بتوزيع كاي مربع χ^2 القيمة النظرية عند درجات الحرية المساوية 1 و مستوى

المعنوية 1% فنجدها تساوي 6,635

الفروض الإحصائية : الفرضية الصفرية $H_0{:}O=E$

الفرضية البديلة $H_1{:}O\neq E$

القرار الإحصائي : إن قيمة χ^2 الفعلية أصغر من قيمة χ^2 النظرية لذلك نقبل نص الفرضية الصفرية بأنه لا يوجد تأثير للدواء الجديد على المرض (غير فعال) و بالتالي نرفض الفرضية البديلة القائلة بوجود تأثير للدواء الجديد بفعاليته ، إذن : الدواء الجديد غير فعال في معالجة المرض لأن الأشخاص اللذين شفوا لوحدهم يساوي عدد الأشخاص اللذين شفوا باستخدامه .

بعض الأمثلة غير المحلولة الخاصة باختبار كاي مربع χ^2 :

مثال 1 :

في دراسة قام بها فريق من الباحثين لمعرفة نسبة ولادة الذكور للإناث و سجلت نتائج 400 حالة ولادة في أخذ مشافي التوليد فكانت النتيجة على الشكل التالي :

185 ذكور + 215 إناث

و المطلوب :

اختبار الفرض القائل بتساوي ولادة الذكور و الإناث بدرجة ثقة 95% (χ^2 النظرية 3,841) .

مثال 2 :

في دراسة لمعرفة العلاقة بين لون البشرة للإنسان ولون الشعر أخذت عينة مؤلفة من 927 شخص و بعد المعاينة سجلت النتائج التالية :

لون الشعر / لون الجلد	أسود	أشقر	خرنوبي	المجموع
أبيض	63	45	140	248
أسود	32	12	60	104
حنطي	210	170	195	575
المجموع	305	227	395	927

و المطلوب :

اختبار الفرض القائل بعدم وجود ارتباط بين لون البشرة و لون الشعر بدرجة ثقة %95

(قيمة χ^2 النظرية 9,488) .

مثال 3 :

في دراسة لمعرفة الارتباط بين نوعية المقررات التي يدرسها الطالب و النتيجة في الامتحان ، سجلت نتائج الطلاب اللذين نجحوا و اللذين رسبوا في ثلاثة مقررات فحصلنا على النتائج الفعلية التالية :

اسم المقرر / النتيجة	احصاء حيوي	فسيولوجيا	تدريب عملي	المجموع
نجاح	240	220	345	805
رسوب	60	70	56	186
المجموع	600	290	401	991

و المطلوب :

اختبار الفرض القائل بعدم وجود علاقة بين النتيجة و نوعية المقررات العلمية بدرجة ثقة 99% (χ^2 النظرية = 9,21) .

العمود الأول = الأرقام المسلسلة.

العمود الثاني = التطبيق الأول.

العمود الثالث = التطبيق الثاني.

العمود الرابع = الفرق بين التطبيق الأول والثاني.

العمود الخامس = انحراف الفرق عن المتوسط في العمود الرابع.

العمود السادس = مربع الانحراف للفرق.

الحل:

١-إيجاد الفرق بين التطبيقين.

٢-إيجاد المتوسط الحسابي للفرق وهو $\frac{\text{١٠}}{\text{١٠}} = \text{١}$

٣-إيجاد الانحراف للفرق عن المتوسط.

٤-إيجاد مربع الانحراف للفرق.

٥-تطبيق المعادلة على النحو التالي .

$$\text{ت} = \frac{|\text{م ف}|}{\sqrt{\dfrac{\text{ع}^2\text{ف}}{\text{ن} - \text{١}}}}$$

ولإيجاد $\text{ع}^2\text{ف} = \frac{\text{مج}(\text{ح}^2\text{ف})}{\text{ن}}$

صورة أخرى :

$$\text{ت} = \frac{|\text{م ف}|}{\sqrt{\dfrac{\text{مج}(\text{ح}^2\text{ف})}{\text{ن}(\text{ن} - \text{١})}}}$$

وبالرجوع إلى (ت) الجدولية عند درجة حرية (١٥٠) لا نجد قيمتها في الجدول ، ولذا تكشف عن قيمتها مقابل درجة الحرية الأدنى = (١٠٠) ومستوى دلالة (٠٥,٠) ، فنجد أنّ قيمة (ت) =١,٦٦(دلالة الطرف الواحد)، و ١,٩٨(دلالة الطرفين).

أمّا عند مستوى دلالة (٠١,٠) فإنّ قيمة (ت) الجدولية = ٢,٣٦(دلالة الطرف الواحد) و ٢,٦٣ (دلالة الطرفين).

وبما أن (ت) المحسوبة = ٨,٤٦ ، وهي أكبر من قيمة (ت) الجدولية في جميع الحالات ، فإنّ ثمّة فروقاً دالة إحصائياً بين المجموعتين لصالح المتوسط الأفضل.

٣/٢-دلالة فرق متوسطين مرتبطين لعينة واحدة:

يطبّق هذا القانون في حال جرى اختبار على مجموعة من الأفراد ، ثم أعيد الاختبار على المجموعة نفسها في وقت آخر ، أي أنّ الاختبارين أجريا على العينة نفسها بفاصل زمني .

مثال: أوجد قيمة (ت) لتطبيق اختبارين في مادة (الرياضيات) على عينة مؤلّفة من عشرة طلاّب ، كما هو مبيّن في بيانات الجدول التالي:

م	التطبيق١	التطبيق٢	ف	ح ف	ح² ف
١	٨	٦	٢	١	١
٢	٦	٢	٤	٣	٩
٣	٤	٥	-١	-٢	٤
٤	٧	٦	١	٠	٠
٥	٩	٧	٢	١	١
٦	٥	٤	١	٠	٠
٧	٧	٦	١	٠	٠
٨	٧	٧	٠	-١	١
٩	٥	٤	١	٠	٠
١٠	٦	٧	-١	-٢	٤
مج	٦٤	٥٤	١٠		٢٠

$$ت = \frac{١}{\frac{٢}{\sqrt{٩}}} = \frac{١}{٠,٤٧} = ٢,١٣$$

وبالرجوع إلى (ت) الجدولية عند درجة حرية ١٠-١=٩، وعند مستوى دلالة (٥.ر.) نجد أن قيمة ت = ١,٨٣، دلالة الاتجاه الواحد، و٢,٢٦ دلالة الاتجاهين .وعند مستوى دلالة (١.ر.) قيمة ت =٢,٨٢، دلالة الطرف الواحد ، و٣,٢٥ دلالة الطرفين.

وبما أنّ (ت) المحسوبة = ٢,١٣، فهي دالة فقط عند مستوى (٠,٠٥) ، لدلالة الطرف الواحد فقط.

ملاحظة: يحدّد الاتجاه الواحد أو الاتجاهين عند مستوى دلالة (٠,٠٥) أو (٠,٠١) تبعاً لتحديد فرضيات البحث من البداية.

٤/٢- دلالة فرق متوسطين لعينتين غير متجانستين:

عندما يختلف حجم العينة في المجموعتين ، يختلف أيضاً الانحراف المعياري لكل عينة عن الأخرى اختلافاً كبيراً ، وفي هذه الحال لا يمكن استخدام (ت) كما سبق أن أوضحنا ولكن يستخدم بدلاً منها صور أخرى.

مثال:

أوجد دلالة فرق متوسطين لعينتين غير متجانسين ، بالنظر لاختلاف حجم العينة ، من خلال البيانات في الجدول التالي :

البيانات	المجموعة الأولى	المجموعة الثانية
الوسط الحسابي	١١,٣٥	١٨,٠٧
الوسيط	١١,٠٠	٢٠,٢٣
التباين	١٩,٢٢	٥,٣٢
عدد الأفراد	١٤	٢٥

الحل:

١.حساب التجانس عن طريق النسبة الفائية (ف):

ف = $\frac{١٩,٢٢}{٥,٣٢}$ = ٣,٦١ ، وبالرجوع لقيمة(ف)الجدولية عند درجة حرية ٢٥,١٤

ومستوى دلالة (٠٥ ,.) نجد أنها =٢,١١، وعند مستوى دلالة (٠١ ,.)=٢,٨٩

وبما أنّ قيمة (ف) المحسوبة أكبر من الجدولية، فالعينتان غير متجانستين لأن الفرق بين $ع^2$١،$ع^2$٢ ، هو فرق معنوي سواء عند مستوى (٠٥ ,.) أوعند مستوى (٠١ ,.)

٢-تطبيق صورة المعادلة التالية .

$$ت = \frac{\text{م}١ - \text{م}٢}{\sqrt{\frac{\text{ع}^2\,١}{\text{ن}\,١} + \frac{\text{ع}^2\,٢}{\text{ن}\,٢}}}$$

$$ت = \frac{١٨,٠٧ - ١١,٣٥}{\sqrt{\frac{١٩,٢٢}{١٤} + \frac{٥,٣٢}{٢٥}}} = \frac{٦,٧٢}{\sqrt{١,٣٧ + ,٢١}} = \frac{٦,٧٢٢}{\sqrt{١,٥}} = \frac{٦,٧٢}{١,٢٦}$$

=٥,٣٣

٣-استخراج قيمة (ت) الجدولية لكل من العينة الأولى ، العينة الثانية.

قيمة(ت) الجدولية = ٢,١٦ العينة الأولى ، عند مستوى دلالة (٠٥ ,.) ودرجة حرية(١٣)

وقيمة(ت) الجدولية = ٢٫٠٦ للعينة الثانية عند مستوى دلالة (٠,٠٥) ودرجة حرية(٢٤)

ملحوظة : تطبق الخطوات السابقة نفسيها بدلالة٠١, إذا أراد الباحث ذلك.

٢/٥ــ مثال توضيحي:

((فاعلية طريقة المناقشة في تدريس مادة الجغرافية))

دراسة تجريبية لطلاب الصف الثاني الثانوي الأدبي في مدارس مدينة دمشق.

أدوات البحث:

استبانة لتعرف آراء الطلبة (ذكور +إناث) نحو طريقة المناقشة في التدريس.

فرضية البحث:

" لا يوجد فرق ذو دلالة إحصائية بين متوسط درجات آراء الذكور ومتوسط آراء الإناث نحو استخدام طريقة المناقشة في التدريس."

التحقق من الفرضية:

لحساب دلالة الفروق بين المتوسطات الحسابية بين المجموعات ولمقارنة فيما بينها، بحسب الجنس، استخدام اختبار ستيودنت (ت) عند مستوى دلالة (٠٫٠٥)

والهدف من ذلك هو تحديد ما إذا كان هناك فروق دلالة إحصائية بين المتوسطات الحسابية، فإذا كانت قيمة (ت) المحسوبة أصغر من قيمة (ت) الجدولية، فالفروق ظاهرية يمكن إرجاعها إلى عاملي الحظ والصدفة، وإذا كانت قيمة (ت) المحسوبة أكبر أو تساوي قيمة (ت) الجدولية فالفروق دالة إحصائية "فروق جوهرية حقيقية".

بعد تفرغ إجابات الاستبانة بحسب الجنس (الذكور والإناث) تبيّن ما يلي :

مجموع تكرارات إجابات الذكور=٢٩١٩ ، مجموع ن = ٤٠ ، متوسط تكرار الذكور=٧٢٫٩٧٥

مجموع تكرارات إجابات الإناث = ٢٧٢٨، مجموع ن = ٣٥ ، متوسّط تكرار الإناث =٩٤ر٧٧

وحتى نستخدم الصورة المناسبة لاختبار ((ت)) لابد لنا من معرفة تجانس العينتين، ويقاس مدى التجانس باستخدام اختبار ((ف)) بقسمة التباين الأكبر على التباين الأصغر ،أي بالنسبة الفائية ، حيث أنّ :

$$\text{ف} = \frac{\text{التباين الأكبر}}{\text{التباين الأصغر}}$$

فإذا كانت قيمة (ف) الجدولية أكبر من قيمة(ف) المحسوبة ،عند درجة حرية ما ، ومستوى دلالة ما ، فعندها نقول إنّ العينتين متجانستين، ونستخدم صورة اختبار (ت) المناسبة لذلك، وهي:

صورة اختبار (ت) لــ:(متوسطين غير مرتبطين لعينتين غير متساويتين و متجانستين).

$$\text{ت} = \frac{\text{م}_1 - \text{م}_2}{\sqrt{\frac{\text{ع}_1^2 \times \text{ن}_1 + \text{ع}_2^2 \times \text{ن}_2}{\text{ن}_1 + \text{ن}_2 - 2}\left[\frac{1}{\text{ن}_1} + \frac{1}{\text{ن}_2}\right]}}$$

حيث أن : ت = قيمة (ت) المحسوبة - (م١) أو (سَ١) = المتوسط الحسابي للمجموعة الأولى.

- (م٢) أو (سَ٢) = المتوسط الحسابي للمجموعة الثانية.

-ع^{2}١= تباين المجموعة الأولى. -ع^{2}٢= تباين المجموعة الثانية.

-ن١= عدد أفراد المجموعة الثانية . -ن٢= عدد أفراد المجموعة الثانية.

وإذا كانت قيمة (ت) المحسوبة أكبر من (ت) الجدولية ، فعندها نقول إنّ العينتين غير متجانستين ،ونستخدم صورة اختبار(ت) المناسبة لذلك، وهي:

صورة اختبار(ت) لـ:((متوسطين غير مرتبطين لعينتين غير متساويتين وغير متجانستين))

-تطبيق صورة المعادلة التالية:

$$ت = \frac{م_{1} - م_{2}}{\sqrt{\frac{ع_{1}^{2}}{ن_{1}} + \frac{ع_{2}^{2}}{ن_{2}}}}$$

-ثم تطبيق صورة المعادلة التالية:

$$ت = \frac{(ت_{1})\text{الجدولية للعينة الأولى} \times \frac{ع_{1}^{2}}{ن_{1}} + (ت_{2})\text{الجدولية للعينة الثانية} \times \frac{ع_{2}^{2}}{ن_{2}}}{\frac{ع_{1}^{2}}{ن_{1}} + \frac{ع_{2}^{2}}{ن_{2}}}$$

وحتى نتأكد من تجانس عينتي البحث نطبق اختبار (ف) السابق ذكره ، وهو:

ف = التباين الكبير= ١٤,٤٩ = ٢,٧

التباين الصغير= ٥,٥٩

وبالرجوع لقيمة((ف)) الجدولية عند درجة حرية(٣٩) كبير،ودرجة حرية صغير(٣٤) ، وعند مستوى دلالة(٠,٠٥) نجدها = ١,٨٠

وبما أنّ قيمة (ف) الجدولية أصغر من قيمة (ف) المحسوبة نقول إنّ العينتين غير متجانستين، لذلك نستخدم صورة اختبار((ت)) لـ:

((متوسطين غير مرتبطين لعينتين غير متساويتين وغير متجانستين))

١- تطبيق صورة المعادلة التالية:

$$\text{ت} = \frac{\left|\text{م}_{١} - \text{م}_{٢}\right|}{\sqrt{\dfrac{\text{ع}^{2}_{١}}{\text{ن}_{١}} + \dfrac{\text{ع}^{2}_{٢}}{\text{ن}_{٢}}}}$$

$$\text{ت} = \frac{\text{٧٥,٤٧} - \text{٧٧,٩٤}}{\sqrt{\dfrac{\text{١٤,٤٩}}{\text{٤٠}} + \dfrac{\text{٥,٠٩}}{\text{٣٥}}}}$$

$$= \frac{\text{٢,٤٧}}{\sqrt{\text{٠,٥٢}}} = \text{٣,٤٣}$$

٢- نستخرج قيمة (ت) الجدولية للعينة الأولى = ٢,٠٢١، عند درجة حرية (٣٩) ومستوى دلالة(٠,٠٥). وقيمة(ت) الجدولية للعينة الثانية=٢,٠٢١ عند مستوى دلالة (٠,٠٥) ودرجة حرية (٣٤)

٣- تطبيق صورة المعادلة التالية:

$$\text{ت} = \frac{\text{(ت١)الجدولية للعينة الأولى} \times \dfrac{\text{ع}^{2}_{١}}{\text{ن}_{١}} + \text{(ت٢)الجدولية للعينة الثانية} \times \dfrac{\text{ع}^{2}_{٢}}{\text{ن}_{٢}}}{\dfrac{\text{ع}^{2}_{١}}{\text{ن}_{١}} + \dfrac{\text{ع}^{2}_{٢}}{\text{ن}_{٢}}}$$

$$\text{فتكون ت} = \frac{\text{٢,٠٢١} \times \dfrac{\text{١٤,٤٩}}{\text{٤٠}} + \text{٢,٠٢١} \times \dfrac{\text{٥,٠٩}}{\text{٣٥}}}{\dfrac{\text{١٤,٤٩}}{\text{٤٠}} + \dfrac{\text{٥,٠٩}}{\text{٣٥}}}$$

$$= \frac{٠,١٦ \times ٢,٠٢١ + ٠,٣٦ \times ٢,٠٢١}{٠,١٦ + ٠,٣٦}$$

$$= \frac{١,٠٥}{٠,٥٢} = ٢,٠١٩$$

والجدول التالي يوضّح معطيات تلك المعادلة :

	حجم العينة	المتوسط الحسابي	التباين	(ت) المحسوبة	(ت) الجدولة
ذكور	ن١=٤٠	م١=٧٥,٤٧	$ع_1^2$=١٤,٤٩	٢,٠١٩	٣,٤٣
إناث	ن٢=٣٥	م٢=٧٧,٩٤	$ع_2^2$=٥,٥٩		

يبين الجدول السابق أنّ قيمة (ت) المحسوبة (٢,٠١٩) أصغر من قيمة (ت) الجدولة(٣,٤٣)،عند مستوى دلالة (٠,٠٥)، مما يشير إلى عدم وجود فروق دالة إحصائياً بين آراء الذكور وآراء الإناث بشأن طريقة المناقشة في التدريس ، وهذا يسمح لنا بقبول الفرضية.

ثالثاً– اختبار كاي مربّع (كا٢)

١ – استخدام كاي مربّع :

إنّ الفكرة الأساسية التي يقوم عليها هذا الأسلوب الإحصائي (كاي مربّع) مصوغة على أساس الفرض الصفري ، هي أن التكرار الملاحظ في الفئة أو الفئات موضع الدراسة ، يختلف عن التكرار المتوقع أو الفرضي اختلافاً يرجع إلى الصدفة.

ويتمّ حساب (كاي مربّع)بوساطة المعادلة التالية:

$$كا^2 = \text{مجموع} \frac{(\text{التكرار الملاحظ} - \text{التكرار المتوقع})^2}{\text{التكرار المتوقع}}$$

٢-نماذج تطبيقيّة لاستخدام كاي مربّع :

٢/١- التوزيع العادل بين القيم :

مثال : لنفرض أنّنا قمنا بتحليل مضمون قصّة للأطفال ، وحصلنا على تكرارت عدد من القيم ، ونريد معرفة مدى عدالة توزيع هذه القيم ، على النحو التالي :

القيمة	الصدق	الأمانة	العدالة	التعاون	محبّةالوطن	الصداقة	المجموع
التكرار	٩	٧	٦	٥	٥	٤	٣٦

لدينا التكرار الملاحظ ، ونريد التكرار المتوقّع ، ونفرضه هنا متوسط التكرارات (٦) فتكون قيمة كا٢ المحسوبة =

$$\frac{(٩-٦)^٢}{٦} + \frac{(٧-٦)^٢}{٦} + \frac{(٦-٦)^٢}{٦} + \frac{(٥-٦)^٢}{٦} + \frac{(٥-٦)^٢}{٦} + \frac{(٤-٦)^٢}{٦} = ٣٫٦١$$

وبالعودة إلى جداول (كا٢) نجد أنّ قيمة $كا^2$ الجدولية =١١٫٠٧عند درجة حرية (٥) ومستوى دلالة (٠٫٠٥) وهي أكبر من قيمة $كا^2$ المحسوبة ، وبناء" عليه يمكن القول بأنّ الفرق ظاهري بين القيم وهناك توزيع عادل لها.

٢/٢- معرفة مدى التطابق بين آراء عينتين أو متغيرين حول موضوع معيّن :

مثال توضيحي- لو كان لدينا الفرضية الصفرية التالية:

(لا توجد فروق بين آراء مدرسي ومدرسات الجغرافية بشأن التدريس بطريقة المناقشة) وكانت نتائج السؤال كما في الجدول التالي :

الاستجابة للسؤال				
جنس المستجيب	موافق	لا راي لي	أرفض	المجموع
ذكر	٤٢	١٣	٣٣	٨٨
أنثى	٢٠	٨	٢٥	٥٣
	٦٢	٢١	٥٨	١٤١

ومن أجل حساب قيمة (كاي مربّع) نقوم أولاً بحساب التكرار المتوقّع لكل تكرار ظاهري على النحو التالي :

- حساب التكرار المتوقع لذكور موافق = $\frac{\text{٦٢×٨٨}}{\text{١٤١}}$ = ٣٨,٧٠

~~- حساب التكرار المتوقع لذكور لا أدري = $\frac{\text{٢١×٨٨}}{\text{١٤١}}$ = ١٣,١١~~

- حساب التكرار المتوقع لذكور أرفض = $\frac{\text{٥٨×٨٨}}{\text{١٤١}}$ = ٣٦,٢٠

- حساب التكرار المتوقع لإناث موافق = $\frac{\text{٦٢×٥٣}}{\text{١٤١}}$ = ٢٣,٣٠

- حساب التكرار المتوقع لإناث أدري = $\frac{\text{٢١×٥٣}}{\text{١٤١}}$ = ٧,٨٩

- حساب التكرار المتوقع لإناث أرفض = $\frac{\text{٥٨×٥٣}}{\text{١٤١}}$ = ٢١,٨

وبعد ذلك نقوم بحساب قيمة كاي مربّع وفق القانون المعتمد ، كما يلي :

$$\text{كا}^2 = \frac{(\text{٤٢} - \text{٣٨,٧٠})^2}{\text{٣٨,٧}} + \frac{(\text{١٣} - \text{١٣,١١})^2}{\text{١٣,١١}} + \frac{(\text{٢٠} - \text{٢٣,٣٠})^2}{\text{٢٣,٣٠}}$$

$$+ \frac{(\text{٣٣} - \text{٣٦,٢٠})^2}{\text{٣٦,٢}} + \frac{(\text{٨} - \text{٧,٨٩})^2}{\text{٧,٨٩}} + \frac{(\text{٢٥} - \text{٢١,٨})^2}{\text{٢١,٨}} = \text{١,٥}$$

وبالعودة إلى جداول (كاي مربّع) نجد أنّ قيمة (كا^2) الجدولية = (١١,٠٧) عند درجة حرية (٥) ومستوى دلالة (٠,٠٥)

وبما أنّ قيمة (كا^2) الجدولية أكبر من قيمة (كا^2) المحسوبة نرفض الفرض الصفري، ونقول لا توجد فروق جوهرية بين آراء المدرسين والمدرسات حول طريقة المناقشة في تدريس الجغرافية .

مسائل وتمارين

1- بيّن ما هي طبيعة العلاقة ونوع المتغيرات في العلاقات الآتية :

- دخل ونفقات المستهلك .
- دخل ونفقات المنتج .
- دخل ومدخرات الأسرة .
- نفقات الأسرة وعدد أفرادها .
- حجم الإنتاج والنفقات الكلية .
- حجم الإنتاج وكلفة الوحدة الواحدة من الإنتاج .
- حجم الطلب على سلعة ما وسعرها .
- حجم عرض سلعة ما وسعرها .
- مدخرات الأسر والفوائد عليها .
- حجم الاستثمار وعوائد الاستثمار .

2- قدّر متانة العلاقة ونوعها من خلال رسم شكل الانتشار للحالات الآتية :

A-

X	2	3	5	8	12	13	15	16	17	20
Y	5	6	8	11	15	16	18	19	20	23

B-

X	1	2	4	5	7	8	10	11	13	15
Y	3	6	18	27	51	66	102	123	171	227

C-

X	4	5	7	8	10	11	12	15	17	20
Y	29	36	44	51	58	66	68	87	93	108

3- ليكن لدينا البيانات الآتية التي تمثل السماد المستخدم والغلة لـــ 100 قطعة متشابهة من الأرض :

الغلـــة / كمية السماد	غير وفيرة	وفيرة	المجموع
عاليـــة	15	55	70
قليلـــة	25	5	30
المجمـــوع	40	60	100

والمطلوب: إيجاد متانة العلاقة بين كمية السماد والغلة مستخدماً المقياس المناسب .

4- أراد أحد الباحثين دراسة العلاقة بين لون بشرة الإنسان وعدد شعر الرأس فقام بجمع البيانات الآتية عن 60 شخص فتبين أن هناك 40 شخصاً لهم شعر كثيف من بينهم 30 شخصاً لون بشرتهم فاتح، والباقي لهم لون بشرة غامق، أما العشرون الباقون فلهم شعر خفيف بينهم 10 أشخاص لون بشرتهم فاتح، والباقون غامق.

والمطلوب: مساعدة هذا الباحث في تأكيد هذه العلاقة أو نفيها.

5- البيانات الآتية تمثل معدل 200 طالب في امتحانات الفصل الأول والثاني :

المعـــدل	امتحانات الفصل الأول	امتحانات الفصل الثاني
50 – 60	60	55
60 – 80	90	80
80 – فما فوق	50	65
المجموع	200	200

المطلوب: هل توجد علاقة بين معدل هؤلاء الطلاب في كلا الامتحانين.

6- احسب متانة العلاقة بين درجة نضج العنب، والمسافة بين الأشجار في البيانات الآتية التي تم الحصول عليها من إحدى التجارب.

المسافة (بالمتر)	3	4	4.5	5	6	7
درجة النضج	أخضر صغير (حصرم)	أخضر كبير (حصرم)	أخضر كبير منقط بالأحمر	أخضر كبير منقط بالأحمر	أخضر كبير محمر	أحمر كبير

7- الجدول الآتي يبيّن النفقات الشهرية والدخل الشهري لـــ 10 مؤسسات إنتاجية متشابهة :

المؤسسة	1	2	3	4	5	6	7	8	9	10
الدخل (بالآلاف)	8	10	10	12	15	16	17	19	20	25
الإنفاق (بالآلاف)	20	22	24	25	26	30	30	32	35	36

المطلوب:

1- إيجاد معادلة مستقيم الانحدار الممثلة للعلاقة وتفسير ثوابتها.

2- حساب معامل التحديد ومعامل الارتباط وتفسيرهما.

3- حساب الخطأ المعياري للتقدير .

4- قدر دخل مؤسسة تبلغ نفقاتها 23 ألف وحدة .

5- قدر باحتمال 95.5% نفقات مؤسسة تبلغ دخلها 23 ألف وحدة .

8- تبين من سجلات مؤسسة المياه في إحدى الأحياء ما يأتي : إن عدد المرات التي تغير به سعر المتر المكعب هو 14 مرة وأن حجم الاستهلاك اليومي عندما كانت المؤسسة توزع المياه مجاناً للمشتركين هو 20 ألف متر مكعب وأن :

$$\sum x = 252 \ , \ \sum y = 154 \ , \ \sum xy = 1596$$

$$\sum x^2 = 6888 \ , \ \sum y^2 = 2290$$

والمطلوب:

1- بفرض أن العلاقة مستقيمة بين سعر المتر المكعب، وحجم الاستهلاك، أوجد معادلة مستقيم الانحدار .

2- حساب التباين المفسر وغير المفسر وما مدلولهما ؟

3- حساب معامل التحديد وتفسيره واحسب معامل الارتباط واختبر معنويته .

4- قدر باحتمال 95.5% حجم الاستهلاك لهذا الحي عندما يكون سعر المتر المكعب 12 وحدة.

5- ارسم مستقيم الانحدار الناتج وبيّن على الشكل التباين المفسر.

6- أوجد باحتمال 95% حدا لتغير حدي حجم الطلب على السلعة المدروسة .

9- بهدف دراسة العلاقة بين عدد سنوات الخبرة والأجر الشهري للعامل في ورشات صيانة السيارات تم جمع البيانات الآتية عن 12 عاملاً:

العامل	1	2	3	4	5	6	7	8	9	10	11	12
عدد سنوات الخبرة	1	2	4	5	7	9	10	12	15	20	23	25
الأجر الشهري (ألف ليرة)	5	7	10	12	16	20	25	30	40	50	55	60

المطلوب:

1- أوجد متانة العلاقة بين عدد سنوات الخبرة والأجر الشهري.

2- أوجد معادلة مستقيم الانحدار وفسر ثوابتها .

3- أوجد التباين المفسر وغير المفسر والتباين الكلي .

4- احسب معامل التحديد ومعامل الارتباط، وفسرها وتأكد من معنوية معامل الارتباط .

5- قدر باحتمال 95.5% الأجر الشهري لعامل عدد سنوات الخبرة لديه 18 عاماً .

6- ارسم مستقيم الانحدار الناتج وارسم حوله مدى من خطأين معياريين..

10- البيانات الآتية تمثل دخل الأسرة ونفقاتها لعينة عشوائية حجمها 12 أسرة من إحدى المدن :

الأسرة	الإنفاق (ألف وحدة)	الدخل (ألف وحدة)
1	5	6
2	6	7
3	9	9
4	10	12
5	8	10
6	7	9
7	7	8
8	5	6
9	16	18
10	11	15
11	10	12
12	6	8
المجموع	100	120

والمطلوب :

1- احسب متانة العلاقة بين الدخل والإنفاق لهذه الأسر.

2- هل يمكن الاعتماد على قيمة معامل الارتباط الذي حصلت عليه في الطلب السابق في تقدير متانة العلاقة في المجتمع؟ برهن ذلك.

3- أوجد معادلة الانحدار الممثلة للعلاقة بين الإنفاق والدخل وفسر ثوابتها.

4- قدر باحتمال 95.5% إنفاق أسرة يبلغ دخلها 17 ألف وحدة .

11- بهدف دراسة تأثير نوع من السماد على إنتاج القمح تم استخدام كميات مختلفة من ذلك السماد في 7 قطع متشابهة الظروف :

القطعة	1	2	3	4	5	6	7	المجموع
كمية السماد (كغ)	0	1	2	4	5	7	9	
الغلة (كغ)	200	210	235	265	260	325	325	

والمطلوب :

1- حساب معامل الارتباط بين كمية السماد والغلة وتفسيره وهل قيمته معنوية ؟

2- إيجاد معادلة مستقيم الانحدار بطريقة المربعات الصغرى وتفسير ثوابتها.

3- حاول إيجاد معادلة مستقيم الانحدار بطرائق أخرى معتمداً على خصائص مستقيم الانحدار.

4- احسب معامل التحديد وفسره .

5- قدر باحتمال 95.5% غلة القطعة إذا تم رش 4 كغ من السماد في القطعة.

12- (مسألة دورة امتحانية 2000-2001 الفصل الأول) :

بهدف دراسة إنفاق الفرد الشهري في الأسر ذات الدخل المحدود قام أحد الباحثين بجمع بيانات عن عدد الأفراد في 100 أسرة، وإنفاق الفرد الشهري في كل أسرة فحصل على الجدول الآتي :

عدد الأسر	عدد أفراد الأسرة	نفقات الفرد الشهرية
12	7	750 – 850
18	6	850 – 950
40	5	950 - 1050
18	4	1050 – 1150
12	3	1150 – 1250

والمطلوب :

1- احسب متوسط إنفاق الفرد شهرياً في تلك العينة واحسب متوسط عدد أفراد الأسرة .

2- أيهما أكثر تمثيلاً لبياناته: متوسط عدد أفراد الأسرة أو متوسط إنفاق الفرد شهرياً ؟ دعم إجابتك بالحسابات اللازمة.

3- حدد طبيعة التوزيع الذي تخضع له بيانات إنفاق الفرد الشهري وطبيعة التوزيع الذي تخضع له بيانات عدد أفراد الأسرة

كما رغب الباحث أيضاً بدراسة العلاقة بين الدخل الشهري لهذه الأسر وإنفاقها الشهري على الغذاء بصرف النظر عن عدد أفرادها، وجمع لذلك البيانات اللازمة وحصل على المجاميع الآتية :

$$\sum x = 6000 \qquad \sum x^2 = 364000 \qquad \sum xy = 297000$$

$$\sum y = 4900 \qquad \sum y^2 = 242575$$

حيث x تشير إلى الدخل وتتراوح قيمته بين 40 و 70 وحدة و y الإنفاق، وتتراوح قيمته بين 36 و 56 وحدة، وكانت وحدة القياس لهذه البيانات هي مئة الليرة السورية، والمطلوب:

4- احسب معادلة مستقيم الانحدار وفسر ثوابتها.

5 – احسب الخطأ المعياري للتقدير واشرح مدلوله.

6- ارسم مستقيم الانحدار وارسم حوله مدى من خطأ معياري واحد.

7- احسب معامل التحديد وفسره وبيّن الأهمية الإحصائية لمعامل الارتباط.

8- تبين أن أسرة يبلغ دخلها الشهري 6600 ليرة تنفق منها على الغذاء 5850 ليرة شهرياً، فهل تعتقد أن إنفاقها على الغذاء اعتيادياً ؟ برر إجابتك إحصائياً .