الرقم: 38/ك ع إ إج/ن ت ب ع ع خ/2022

المسيلة في: 20 مارس 2022

مستخرج من محضر اجتماع المجلس العلمى للكلية

لجلسة يوم: 12 أكتوبر 2021

بخصوص الموافقة على الكتاب

اجتمع المجلس العلمي للكلية في دورته العادية بتاريخ: 12 أكتوبر 2021 ووافق على الكتاب بعد ورود تقارير الخبرة الايجابية.

- <u>للأستاذ (ة):</u> مامش نجية
- <u>عنوان الكتاب:</u> أساسيات الإحصاء الاجتماعي

سلمت هذه الشهادة للمعني (ة) بطلب منه (ا) لاستعمالها في حدود ما يسمح به القانون

رئيس المجلس العلمي

الأستاذ: صالح لميش
رئيس المجلس العلمي
كلية العلوم الانسانية والاجتماعية

دار المتنبي للطباعة والنشر
العنوان: حي تعاونية الشيخ المقراني/ طريق إشبيليا
مقابل جامعة محمد بوضياف/ المسيلة
الفاكس: 035.35.31.03 / الهاتف: 0668.14.49.75
EMAIL : elmotanaby.dz@gmail.com

EDITION EL MOTANABY
دار المتنبي للطباعة والنشر

المسيلة في: 2022/05/28

شهادة نشر كتاب

تشهد وتتشرف دار المتنبي للطباعة والنشر ب:

بنشر وطباعة كتاب الموسوم ب:

أساسيات الإحصاء الاجتماعي

تأليف

د. نجية مامش

ردمك (ISBN) : 3_ 48 _ 865 _ 9931 _ 978

بتاريخ: 2022/05/28

دار المتنبي
للطباعة و النشر
تعاونية الشيخ المقراني-المسيلة
0773.30.52.82 ر.ت: 28/01 2864147H13

د. نجية مامش

أساسيات الإحصاء الإجتماعي

EDITION EL MOTANABY
دار المتنبي للطباعة والنشر

سنة النشر: 1443هـ/ 2022م

أساسيات الإحصاء الإجتماعي
د. نجية مامش

مقر دار النشر: حي تعاونية الشيخ المقراني، طريق إشبيليا / مقابل جامعة محمد بوضياف-المسيلة
التواصل مع دار النشر: elmotanaby.dz@gmail.com
الهاتف: 0773305282 / 0668144975
فاكس: 035353103

سنة النشر: 1443هـ/ 2022م
ISBN: 978-9931-865-48-3
9 789931 865483
EDITION EL MOTANABY
دار المتنبي للطباعة والنشر

أساسيات الإحصاء الاجتماعي

د. نجية مامش

(جامعة المسيلة)

بسم الله الرحمن الرحيم

- **العنوان:** أساسيات الإحصاء الاجتماعي
- **إعداد: د.** نجية مامش
- **تنسيق داخلي للكتاب:** العربي زغلاش أيوب
- **مقاس الكتاب:** 25/17

EDITION EL MOTANABY

دار المتنبي للطباعة والنشر

- **الناشر:** دار المتنبي للطباعة والنشر
- رقم الإيداع: **ISBN :978 _ 9931 _ 865 _ 48 _3**
- **سنة النشر:** 1443هـ/ 2022م

- **مقر الدار:** حي تعاونية الشيخ المقراني/ طريق إشبيليا مقابل جامعة محمد بوضياف/ المسيلة

- **للتواصل مع الدار:**
- **الإيميل:** elmotanaby.dz@gmail.com
- **الموقع الإلكتروني:** http://motanaby.onlinewebshop.net/

هاتف: 07.73.30.52.82 —— 06.68.14.49.75

فاكس: 0.35.35.31.03

د: نجية مامش

جامعة المسيلة

أساسيات الإحصاء الاجتماعي

2022

الفهرس

الفصل الأول

مفهوم الاحصاء والطريقة الاحصائية

الفصل الثاني

جمع البيانات الاحصائية

الفصل الثالث

تفريغ وتبويب البيانات الاحصائي

الفصل الرابع

التمثيل البياني للجداول الاحصائية

الفصل الخامس

مقاييس النزعة المركزية

الفصل السادس

مقاييس التشتت

الفصل السابع

مقاييس الشكل

الفصل الثامن

الارتباط

الفصل التاسع

الانحدار الخطي البسيط ومعامل التحديد

الفصل العاشر

اختبار الفرضيات

مقدمة

ان الاحصاء كعلم قائم بذاته يساعد على تطوير الاساليب الاحصائية المختلفة لغرض وصف وتحليل وتفسير وحتى التنبؤ بالطواهر في مختلف المجالات العلمية الطبيعية والاجتماعية مهما كان حجم المعلومات المتوفرة حول الظاهرة ومهما كان نوعها، وبالتالي فهو العلم الذي يخدم كل العلوم التي تعتمد على تحليل البيانات ويساعد على تحقيق أهدافها البحثية سواء كانت وصفية، ارتباطية او تنبئية. كما يعد الاحصاء الركيزة الاساسية لاختبار المعرفة العلمية وتطويرها بناء على مستجدات الواقع، وتقديم نتائج علمية دقيقة وموضوعية حول الظاهرة محل البحث، وبناء عليه تظهر حاجة أي باحث اجتماعي الى أهمية الالمام بأساسيات هذا العلم حتى يتمكن من انجاز عمله البحثي على أكمل وجه.

ويهدف هذا المؤلف إلى تعريف طلاب فرع العلوم الاجتماعية بعلم الإحصاء وأهميته ودوره في تسهيل عمل الباحث الاجتماعي وفي مختلف محطات البحث بدءا من أخذ العينات، وكيفية جمع البيانات وتفريغها وتبويبها ووصفها (مقاييس النزعة المركزية والتشتت وأشكال توزيع البيانات) وقياس العلاقات بين المتغيرات لتحديد شدتها واتجاهها، وذلك بهدف إكساب الطالب مجموعة من الخبرات في مجال الإحصاء الاجتماعي وفق خطوات منهج البحث الاجتماعي لمساعدته في عرض نتائج البحوث بصورة كمية محددة واضحة ومختصرة ودقيقة.

وقد احتوى هذا الكتاب على عشرة فصول، خصص الفصل الاول لمفهوم الاحصاء وتطوره كعلم وأهميته في البحث الاجتماعي وينتهي بالإشارة الى المفاهيم الاساسية في المنهج الاحصائي، تطرق الفصل الثاني الى أساليب وأدوات جمع البيانات الاحصائية العينات وأنواعها، أما الفصل الثالث فتم فيه عرض طرق تبويب البيانات الإحصائية في جداول تكرارية، وخصص الفصل الرابع لطرق العرض البياني، وتم في الفصل الخامس التعريف بأهم الأساليب الاحصائية لوصف وتحليل البيانات من خلال مقاييس النزعة المركزية المختلفة مثل الوسط

الحسابي والوسيط والمنوال وتعريف الطالب بطرق حساب كل منها. في حين تضمن الفصل السادس مقاييس التشتت المختلفة مثل المدى والتباين والانحراف المعياري. كما جاء الفصل السابع لشرح طرق تحديد شكل التوزيعات التكرارية وما يتعلق بالالتواء والتفلطح. واهتم الفصل السابع بالارتباط ومعاملاته المختلفة، أما الفصل التاسع فخصص للانحدار الخطي البسيط، وتم التعرض في آخر فصل الى مفهوم اختبار الفرضيات وأهم الاساليب الاحصائية المساعدة على انجازه.

ويتميز هذا الكتاب ببساطة عرض المقاييس الاحصائية المختلفة التي تساعد الباحث الاجتماعي خصوصا في علم الاجتماع على استخدامها بكل سهولة ابتداء من عملية المعاينة وجمع البيانات الى تفريغ البيانات وتبويبها وصفها وصولا الى تحليل العلاقة بين متغيرات البحث. ليقدم بذلك الأساسيات الاولية في الاحصاء التي يحتاجها أي باحث اجتماعي للتعامل مع بيانات بحثه.

الفصل الاول: مفهوم الاحصاء والطريقة الاحصائية

1- تعريف الإحصاء

2- لمحة عن نشأة وتطور علم الإحصاء

3- علاقة علم الإحصاء بمجموعة العلوم الإنسانية

4- حدود الإحصاء

5- الطريقة الإحصائية

6- مفاهيم أساسية في الطريقة الإحصائية

تمارين

1- تعريف الإحصاء

الإحصاء لغة هو العد الشامل؛ ويقال أحصى الشيء بمعنى عده وحفظه وعقله وضبطه. أما تعريف علم الاحصاء: **"هو مجموعة من الطرق والوسائل والقواعد والقوانين والنظريات المبنية على التحليل المنطقي والتي تهدف الى جمع وعرض ووصف وتحليل البيانات واستخدام النتائج في التفسير والتنبؤ واتخاذ القرارات."**

وينقسم الإحصاء إلى فرعين أساسيين يمثلان الوظائف الاساسية لهذا العلم هما:

أ- **الإحصاء الوصفي** يهتم بأساليب وطرق جمع البيانات الإحصائية وتلخيص هذه البيانات وعرضها باستخدام أساليب ومقاييس خاصة .

ب- **الإحصاء الاستدلالي** ويهتم باختبار الفرضيات وطرق اتخاذ القرار وذلك باستخدام أساليب رياضية خاصة.

2- لمحة عن نشأة وتطور علم الإحصاء

ظهرت كلمه إحصاء (statistique) لأول مرة عام1749 وهى مشتقه من الكلمة اللاتينية (status) أو الايطالية (statista) وتعني كلاهما الدولة السياسية. ومن الطبيعي أن تكون الدولة أول من اهتم بجمع البيانات وذلك لإدارة شؤونها خاصة عن السكان لأغراض حربية وضريبية، وامتدت بعد ذلك لتشمل إحصاءات حجم السكان والمواليد والوفيات والإنتاج والاستهلاك والثروة الخ (موسي محمد، 2008).

و تشير كثير من الدلائل على الاهتمام بالإحصاء واستخدامه منذ زمن بعيد (العصور القديمة) حيث أقتصر اهتمام الحكومات منذ القدم بالمعلومات الاجتماعية وذلك لأغراض التنظيم والتخطيط، واستخدم **الإحصاء في عصره الأول** في جمع البيانات عن السكان وحصرهم من قبل الدولة لأهداف معينة كما سبقت الاشارة اليه، ويعد الإمبراطور الصيني "ياو" «YAO» أول من نظم إحصاء المنتجات الزراعية ويعود ذلك الى سنة 2238 قبل الميلاد، كما نظم قدماء المصريين إدارة قوية لتسيير الضرائب والمساحات وكان ذلك حوالي سنة 2500 قبل الميلاد .

في القرن السابع عشر والذي يمكن اعتباره **العصر *الإحصائي الثاني*** تم استخدام الطريقة الرقمية للدلالة على الظواهر موضوع البحث على اعتبار أن هذه الطريقة أدق وأقوى في التعبير عن هذه الظواهر وتركز الهدف من هذه الطريقة في معرفة عدد السكان، عدد المواليد، عدد الوفيات، مقدار الثروة والدخل ومقدار الضرائب المحصلة وكمية الناتج من المحاصيل الزراعية ... باختصار نجد أن مجال الإحصاء قبل القرن العشرين كان مرتبطاً في الغالب بالمجالات الاقتصادية والاجتماعية المتمثلة بتعداد السكان ومعرفة خصائصهم الاجتماعية والاقتصادية، وكانت الأساليب الإحصائية المستخدمة تمتاز بالبساطة بحيث لم توفر للإحصاء الأسس والمقومات الكافية لأن يصبح علماً قائما بذاته.

ويمكن تحديد بداية **العصر الإحصائي الثالث** الذي تطور فيه علم الإحصاء من مجرد فكره الحصر والعد إلى أن أصبح علما له قواعده ونظرياته ويرجع الفضل في ذلك إلى تطور علوم الرياضيات في القرن الثامن عشر وظهور بعض النظريات العلمية الهامة مثل نظرية الاحتمالات على يد كثير من العلماء من أمثال العالم الألماني: « *Carl Friedrich Gauss* » (1777-1855)، والعالم

الفرنسي « *Pierre Simon Laplace* » (1749-1827) صاحب نظرية "التوزيع الطبيعي" « La loi normale »، بالإضافة إلى ما ساهم به الفلكي والرياضي البلجيكي « Adolphe Quételet » 1796-1874 الذي يعد أول من وضع قواعد محددة لعلم الإحصاء، وقد كان لهذه الأعمال الدور الكبير في تطور هذا العلم واكتسابه أهمية كبرى بحيث أصبح علماً مستقلاً وانتشر استخدامه، وبدأ الاهتمام من قبل العلماء في تطبيق نظرياته وطرقه وأساليبه في الكثير من فروع العلم الحديث كالهندسة والطب والصيدلة والزراعة والصناعة والجغرافيا والفلك وعلم النفس باعتباره الطريقة الصحيحة والأسلوب الأمثل إتباعه في البحث العلمي وقد أدى ظهور الحاسبات الآلية وتطورها في وقتنا الحالي بأنواعها المختلفة وبسرعة انجازها ودقتها إلى تمهيد الطريق لاستخدام وتطبيق الأساليب الإحصائية المختلفة في شتى المجالات والميادين (خلف عبد الجواد، 2013).

3- علاقة علم الإحصاء بمجموعة العلوم الإنسانية

بعد التطور التكنولوجي الهائل في كافة الميادين والذي فرض نفسه فجأة اصطحب هذا بتطور في كافة العلوم الإنسانية من حيث استحداث طرق جديدة لمعالجة الموضوعات الاجتماعية والفلسفية والنفسية وغير ذلك وأصبحت العلوم الرياضية من أهم الموارد المساعدة في تنفيذ البحوث الاجتماعية. ولا يمكن إنكار دور علم الإحصاء في هذا التقدم، فالعلم يحتل مكانة كبيرة ويعتبر جزأً غير بسيط من ضمن هذه العلوم إلى الحد الذي نجد فيه فرعاً من فروع علم الإحصاء يسمى بالإحصاء في مجال العلوم الاجتماعية أو الإحصاء الاجتماعي والطريقة الإحصائية والنظريات العلمية هي أهم أدوات الباحث في مجال العلوم الإنسانية.

فالطريقة الإحصائية هي أسلوب عمل لتنفيذ البحوث الاجتماعية ونظرية الاحتمالات والنهاية المركزية وما يشمل ذلك من تطبيقات أساسية لها أهميتها في هذا المجال، كما أن أسلوب إيجاد علاقة الارتباط سواء كان بسيطاً أو متعدداً للظواهر الاجتماعية وغير ذلك من الظواهر التي يهتم الباحثون بتفسيرها ودرستها، وأيضاً تطبيق نظرية وضع الفروض والاختبارات الإحصائية وتحديد انتماء العديد من تلك الظواهر وتبعيتها لأحد التوزيعات الاحتمالية، كل ذلك ضروري وهام في مجال العلوم الإنسانية، وليس بالغريب القول بأن كل باحث متخصص في مجال العلوم الاجتماعية يجب عليه أن يكون ملماً عارفاً لأهم خطوات الطريقة الإحصائية والنظريات المختلفة لهذا العلم والمجالات التطبيقية المتعددة له، إذا كان يريد أن يرقى بأبحاثه ومعلوماته إلى مستوى روح العصر.

وهكذا نستخلص أن الإحصاء هو علم له طرقه العلمية ووظائفه المتطورة وقوانينه ونظرياته المتعددة والتي تعتبر أساساً للكثير من العلوم الأخرى ومنطلق لتطورها. وهو علم له علاقاته الممتدة عبر كل العلوم يؤثر فيها ويتأثر بها ويمثل جزء يكاد يكون عاماً ومشتركاً في كل العلوم تبدأ به وتنهل من طرقه ونظرياته مع اختلاف في درجة الامتداد والتشعب من علم إلى آخر. (بن محمد الصغير، 2001)

4- حدود الإحصاء

إذا كانت هذه هي وظائف علم الإحصاء فلا يجب الانحدار في سلم المغالاة، واعتبار أن علم الإحصاء هو العلم الذي يوصل الى الحقيقة التي لا شك فيها خاصة بالنسبة للعلوم الاجتماعية نظرا لعدم تغلبه على بعض النقائص، مثل حصر عوامل السببية، وترك بعض الفجوات لعوامل الصدفة في تحليل العلاقة بين الظواهر ذلك أن بعض العلاقات بين الظواهر تبدو بسيطة لكنها في

الواقع هي أكثر تعقيد مما تبدو عليه، وبما أن الإحصاء يعد أداة لا تستخدم إلا في العثور على إجابات عن أسئلة تتصل ببيانات يمكن التعبير عنها بصيغ كميه. وفي مجال العلوم الاجتماعية موضوعات لا حصر لها لا يمكن صياغة البيانات الخاصة بها في صورة كميه على نحو دقيق، ومن ثم لا يستطيع الباحث استخدام التحليل الإحصائي في دراستها. عموما لا يجب أن اتخاذ الأسلوب الإحصاء في البحوث الاجتماعية خصوصا كهدف بل يجب استعماله كوسيلة بغية الوصول إلى الحقيقة.

5- الطريقة الإحصائية

تستعمل الطريقة الإحصائية في كل المجالات والتخصصات، في العلوم الاجتماعية والعلوم الدقيقة والاقتصاد والتجارة، وخاصة اذا تعلق الأمر بظواهر قابلة للقياس.

تمكن الدراسة الإحصائية من تبسيط الظاهرة المدروسة خلال فترة زمنية محدودة على شكل جدول _ عرض جدولي _ أو على شكل بيان _ العرض البياني _ أو على شكل عدد _ العرض العددي _

تهدف الطريقة الإحصائية الى دراسة معطيات ظواهر معينة في فترات سابقة لإيجاد حلول مناسبة لمشاكل الحاضر والتنبؤ بقيمها في المستقبل، إذا بقيت نفس الظروف على حالها .

وتتلخص الطريقة الإحصائية (جلاطو، 2002) في الخطوات التلية :

1-5- اختيار مشكلة البحث: يبدأ البحث بمشكلة عامة تتطور أثناء خطوات الإعداد إلى مشكلة محددة تتطلب إجابة مقترحة، واختيار المشكلة وصياغتها صياغة دقيقة هي التي تجعلها قابلة للبحث.

5-2-الفروض: يصاغ الفرض على أنه إجابة محتملة لمشكلة البحث. فعلاقته بالمشكلة علاقة الإجابة بالسؤال الذي تتصدى المشكلة لحله. وتأتي نتائج التجربة لاحقا لتختبر صحة الفرض.

5-3-خطة البحث وجمع البيانات: تشمل خطة دراسة المشكلة على بيان تفصيلي لمصادر المعلومات القائمة، وبهذا تتناول هذه الخطة بيانا تفصيليا عن عينة الأفراد التي تستخدم في التجربة والأسس العلمية لاختيارها والمقاييس التي تجرى .

5-4-التصنيف: وتتمثل في ضبط حلات المتغير

5-5-التبويب: بعد الانتهاء من جمعه البيانات المحدد في الخطة البحث، وتصنيفها يقوم الباحث بتبويبها في جداول حتى تسهل عليه عملية تلخيصها وتحليلها وتفسيرها .

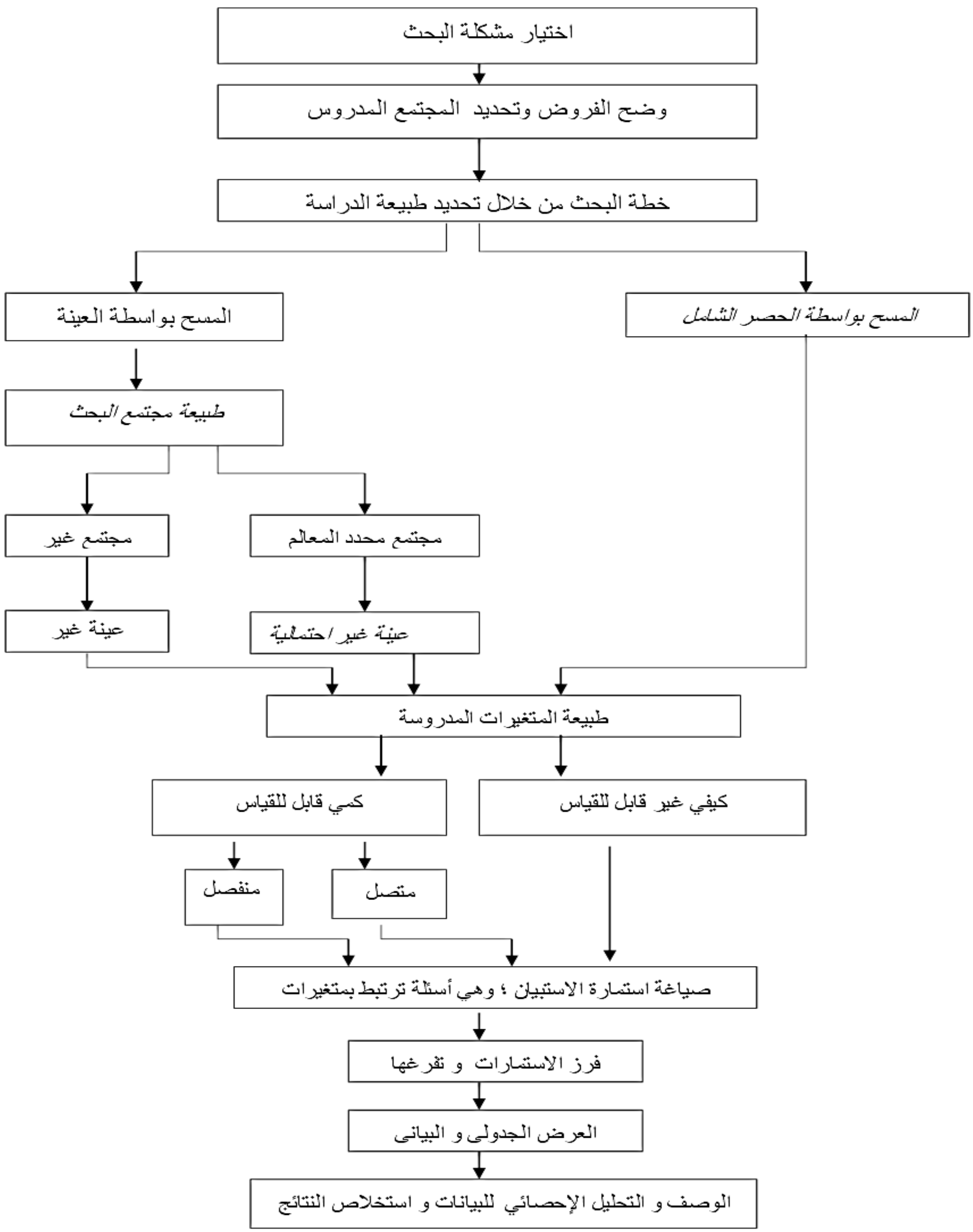
5-6-الوصف الإحصائي: يعتمد الوصف الإحصائي للظواهر المختلفة على الكشف عن مدى تجمع بياناتها العددية أو مدى تشتتها والعلاقة المختلفة التي تربط كل ظاهرة بأخرى والقيمة العددية لهذا الارتباط. ويسمى هذا الميدان من ميادين علم الإحصاء بالإحصاء الوصفي .

5-7-التحليل الإحصائي: يعتمد التحليل الإحصائي على نوع المشكلة، وخصائصها الرقمية وهدف البحث، فالتحليل الذي يصلح لمعالجة مشكلة ما قد لا يصلح لمعالجة مشكلة أخرى. وتمهد الخطوة السابقة المتمثلة في الوصف الإحصائي الشامل، تمهيدا صحيحا للتحليل الإحصائي المناسب لأنه يوضح الخواص الإحصائية للظاهرة. ويسمى هذا النوع من ميادين علم الإحصاء بالإحصاء والتحليلي أو الاستدلالي أو ألاستنتاجي.

5-8-التفسير: ينطوي التفسير على ضرب من ضروب التعميم، مع الحذر ألا يتجاوز هذا التعميم حده أو مداه، ذلك لأن هذه العملية تقوم على إطار تحدده عينة الأفراد الذين أجريت عليهم التجربة والاختبارات التي استخدمت في هذه الدراسة. ولعل من الأخطاء الخطيرة هو تعميم نتائج البحث على كل مجتمع البحث دون توفر الشروط الضرورية لذلك، كعدم استعمال الاحتمالية.

5-9-التقرير: ويقصد به تدوين البحث، اذ يبدأ التقرير من حيث بدأت المشكلة باختيارها وصياغتها، وينتهي إلى حيث انتهت بالتحليل الإحصائي والتفسير النهائي، أي أنه بهذا المعنى يسجل خطوات البحث في تطورها خطوة تلو خطوة ليكون بذلك أقرب الى الموضوعية العلمية والتنظيم المنطقي المتناسق .

الشكل رقم (01): يمثل مخطط يلخص خطوات المنهج الإحصائي



6- مفاهيم أساسية في الطريقة الإحصائية

1-6- الظاهرة المدروسة: (LE PHENOMENE) وتمثل في الموضوع الذي يهتم الباحث بدراسته ومن خلاله يتحدد المجتمع الإحصائي، مثلا: الاستهلاك في الأسرة، وفيات الأطفال، التحصيل الدراسي.

2-6-المجتمع الإحصائي: (POPULATION) هو المجتمع الذي يعكس الظاهرة المدروسة فقد يكون مجموعة من السكان أو مجموعة من الطلبة أو أسر أو إنتاج أو عمال ...

3-6- الوحدة الإحصائية: (UNITE STATISTIQUE) هي الوحدة الأساسية التي يتكون منها مجتمع البحث . مثال فإذا كان مجتمع البحث هو الطلبة فالوحدة الإحصائية هنا هي الطالب .

المتغير: يعرف أنه الخاصية أو السمة المدروسة وهو يختلف باختلاف المفردات التي تمثل مجتمع البحث مثل طول القامة، الذكاء، الاتجاه، التحصيل، وتنقسم المتغيرات الإحصائية إلى قسمين:

1-4-6- المتغيرات الكيفية: (LES VARIABLE QUALITATIVES)

كما تعرف بالمتغيرات الوصفية وهي عبارة عن صفات للمتغيرات المدروسة وهذا النوع من المتغيرات غير قابلة للقياس باعتبارها متغيرات غير رقمية أي لا يعبر عنها بأرقام بل بكلمات، مثل: الجنسية، الألوان، الوظائف ... .

2-4-6- المتغيرات الكمية: (LES VARIABLES QUANTITAVES)

وتعرف بكونها قابلة للقياس ذلك لأنه يعبر عنها بواسطة الأرقام .و هي أكثر المتغيرات استعمالا لان لغة الإحصاء هي لغة الأرقام. ومن أمثلة المتغيرات القابلة

للقياس: الإنتاج ،الاستهلاك، الدخل، الادخار، المواليد، الوفيات، الهجرة ... وينقسم هذا النوع من المتغيرات الى قسمين:

المتغيرات الكمية المنفصلة: (VARIABLES DISCRETES) وهي لمتغيرات التي يمكن تأخذ أرقاما صحيحة لا يمكن تجزئتها مثل عدد الأطفال في العائلة وقطع الغيار المنتجة ...

المتغيرات الكمية المتصلة: (VARIABLES CONTINUES) وهي المتغيرات التي تأخذ كل القيم الممكنة لمجال الدراسة، ونظرا للعدد غير المتناهي لهذه القيم نقسم مال الدراسة الى مجالات جزئية تسمى الفئات .

أسئلة الفصل

<u>السؤال الأول :</u>

س1- ما هو مفهومك لعلم الإحصاء؟

س2- ما المقصود بمفهوم الحصر الشامل والحصر بالعينة ؟

- ما الفرق بينهما وما هي مميزات وعيوب كل منهما ؟

س3- ما الفرق بين نمط العينات الاحتمالية ونمط العينات لا احتمالية ؟

- ما هو العامل الأساسي المتحكم في المفاضلة بين النمطين السابقين ؟

- ما الفرق بين العينة الطبقية والعينة العشوائية البسيطة ؟

س4 – ما هي مصادر جمع البيانات الإحصائية ؟

ما هي انسب وسيلة لجمع البيانات الإحصائية , وما هي التعليمات الواجب أخذها بعين الاعتبار عند إعدادها ؟

س5 – اتيح لك بنك معلومات من خلال شبكة الانترنت فأخذت منه بيانات لدراسة ظاهرة معينة، ما نوع هذه المعلومات من حيث مصدر حصولك عليها؟

<u>السؤال الثاني:</u>

صنف المتغيرات التالية حسب طبيعتها (كيفي، كمي متصل أو كمي منفصل)

الوظيفة، النوع (الجنس)، مكان الإقامة، اللغة الأم، الحالة العائلة، الوزن، الطول، عدد السيارات المملوكة، عدد الأطفال، الجنسية، عدد الغرف في المنزل، عدد الكتب، درجة الحرارة .

السؤال الثالث:

في الدراسات التالية حدد: المجتمع الإحصائي، الوحدة الإحصائية والمتغير الإحصائي وطبيعة هذا المتغير:

- لون المحافظ لدى تلاميذ قسم معين .
- الأجر السنوي لدى المعلمين في مدرسة معينة.
- مدة حياة المصابيح المنتجة من طرف مصنع ما.
- عدد الأصوات التي تحصل عليها كل حزب في الانتخابات .
- المسافة التي تبعد بها كل ولاية عن العاصمة.
- مكان الإقامة بالنسبة لعينة من الطلبة
- عدد الأطفال لدى الأسرة .
- نوع الجريمة المرتكبة لعينة من السجناء.
- الشهادة المحصل عليها بالنسبة لأساتذة كلية العلوم الانسانية

الفصل الثاني: البيانات الاحصائية وطرق جمعها

1- مصادر جمع البيانات الإحصائية

2- أدوات جمع البيانات

3 -طرق جمع البيانات

4- أساليب جمع البيانات

5- شروط العينة

6- أنواع العينات

6-1 العينات الاحتمالية

6-2-العينات لا احتمالية

7-تحديد حجم العينة

8- مصادر الخطأ في العينة

تمهيد

تمر الطريقة الإحصائية بعملية أساسية تتمثل في جمع البيانات وهذه الأخيرة ترتبط بها عناصر فرعية أساسية لابد أن يكون المتبع لهذه الطريقة في إجراء بحوثه على دراية بها، ويمكن تلخيص أهم هذه العناصر فيما يلي:

1- مصادر جمع البيانات الإحصائية

يعتمد الباحثون على مصدرين أساسيين للحصول على المعلومات الإحصائية الخاصة بظاهرة معينة هما (طبية، مبادئ الاحصاء، 2008):

1-1- المصادر الأولية: وفيها يقوم الباحث نفسه بإعداد وسيلة جمع البيانات الخاصة ببحثه وتنفيذها بنفسه أو بواسطة مساعديه، سواء اتبع أسلوب الحصر الشامل أم أسلوب الحصر بالعينة، وذلك عن طريق الاتصال بوحدات أو مفردات المجتمع المبحوث من أجل جمع البيانات.

1-2- المصادر الثانوية: وفيها يتحصل الباحث على المعلومات الإحصائية من الدراسات والتحقيقات، أو بعض الجهات الرسمية المتخصصة في جمع البيانات الإحصائية أو تسجيلها كالحالة المدنية والديوان الوطني للإحصاء، وذلك من أجل استخدامها وفق أهدافه الخاصة من البحث.

2- أدوات جمع البيانات

أهم وسيلة تصلح بياناتها للتحليل الإحصائي هي الاستمارة خاصة في العلوم الاجتماعية؛ والاستمارة هي ورقة تشمل مجموعة من الأسئلة التي تعكس الإجابة عليها جملة البيانات التي يحتاجها الباحث في تحليلاته للظاهرة المدروسة

(المتغيرات). ولتصميم الاستمارة الإحصائية يجب على الباحث إتباع التعليمات التالية:

وضع مقدمة إيضاحية تكتب بأسلوب يستميل الشخص الذي تأخذ منه المعلومات ويشجعه للإجابة عن الأسئلة وملء الاستمارة بمعلومات صحيحة، مع الإشارة إلى أن المعلومات التي يدلي بها تبقى سرية ولا تستخدم إلا لأغراض علمية .

القسم الأول من الاستمارة يحتوى على معلومات أولية عن المبحوث دون السؤال عن اسمه، كالسن، المنطقة السكنية، الجنس، الحالة الاجتماعية...

القسم الثاني من الاستمارة يتضمن الأسئلة المتعلقة بالبحث وهذه الأسئلة يجب أن تراعي ما يلي:

* ألا تكون الأسئلة كثيرة حتى لا يملً المبحوث من طول فترة الاستجواب، ولا قليلة بحيث لا تفي بإعطاء جميع البيانات التي يتطلبها البحث .

*لا يسأل المبحوث عن مسائل شخصية كالاسم واللقب .

* أن تكون واضحة المعنى لا لبس فيها ولا غموض فلا يضِل الشخص المبحوث .

* ألا يحتمل السؤال أكثر من إجابة واحدة، ولا يكون موحيا بإجابات معينة

* يفضل طرح الأسئلة المغلقة ⊗ التي تعفي المبحوث من بذل مجهود فكري للإجابة على الأسئلة المفتوحة⊗ ، حتى تسهل على الباحث عملية تفريغ هذه البيانات لاحقا

* ألا تكون الأسئلة محرجة أو مثيرة للغضب .

* أن تتكرر بعض الأسئلة بصيغ مختلفة على أن تكون متباعدة وذلك عندما يراد بيانات دقيقة حول نقطة مهمة في البحث .

3 -طرق جمع البيانات

1-3- الطريقة غير المباشرة وهي أن ترسل الاستمارات الى المبحوثين ليقوموا أنفسهم بملئها، ويمكن إتباع هذه الطريقة في مجتمعات البحث التي تحسن القراءة والكتابة

2-3- الطريقة المباشرة وهي أن يقوم الباحث نفسه بملء الاستمارات من خلال إجراء مقابلة مع وحدات مجتمع البحث، أو من خلال تجنيده لمجموعة من المساعدين وتتبع هذه الطريقة في المجتمعات الأمية.

4- أساليب جمع البيانات

1-4- الحصر الشامل: من أمثلة الحصر الشامل في عملية جمع البيانات: التعداد الوطني، ويقصد به جمع البيانات من كل وحدات مجتمع البحث بدون استثناء وهو ما يتطلب موارد مالية وبشرية وتقنية ضخمة، إلا أنه أكثر دقة وشمولية ومصداقية من حيث النتائج المتوصل إليها (طبية، مبادئ الاحصاء،).

⊗ - الأسئلة المغلقة هي الأسئلة التي تقدم للمستجوب مجموعة من الاقتراحات يقوم باختيار ما يناسبه منها .

⊗ الأسئلة المفتوحة هي التي يجيب عليها البحوث بعباراته الخاصة .

2-4- العينة: (L'ECHANTILLON) هي جزء أو شريحة من المجتمع الإحصائي، يختاره الباحث وفق طرق علمية، يمثل أحسن تمثيل المجتمع الإحصائي ويعكس سماته الأساسية (علام، رسلان، 1987).

وكثيرا ما يلجأ الباحثين إلى اختيار أسلوب العينة لما فيه من مزايا تتمثل في الاقتصاد في الوقت والجهد والمال.

بينما تتحكم هذه العوامل الثلاثة الأخيرة في تحديد حجم العينة إضافة إلى طبيعة الظاهرة المبحوثة، يقوم الباحثون بتقدير نسبة الخطأ التي قد تحتويها نتائج الدراسة بواسطة العينة نظرا لعدم شمولية كل مجتمع البحث بالدراسة، عكس الحصر الشامل.

5- شروط العينة

أ- تكون مفردات العينة ممثلة للمجتمع الذي يجرى عليه البحث تمثيلاً صحيحا وليست ممثلة لمجتمع آخر. بمعني أنه إذا تكررت نفس النتائج على عينات أخرى من نفس المجتمع، كانت العينة التى يجرى عليها البحث عينة ممثلة للمجتمع الأصلي أصدق تمثيل، وبذلك يمكن أن تكون خصائص مفردات العينة (إحصائيات العينة) متقاربة أو متشابهة مع خصائص المجتمع (معالم المجتمع) الذي تنتمي إليه .

ب - ألا تكون المفردات المختارة ممثلة لجزء (قطاع) من أجزاء المجتمع الأصلي بل يجب أن تمثل جميع أجزاء المجتمع(19)

أ- تكون مفردات العينة ممثلة للمجتمع الذي يجرى عليه البحث تمثيلاً صحيحا وليست ممثلة لمجتمع آخر. بمعني أنه إذا تكررت نفس النتائج على عينات أخرى من نفس المجتمع، كانت العينة التى يجرى عليها البحث عينة ممثلة للمجتمع

الأصلي أصدق تمثيل، وبذلك يمكن أن تكون خصائص مفردات العينة (إحصائيات العينة) متقاربة أو متشابهة مع خصائص المجتمع (معالم المجتمع) الذي تنتمي إليه .

ب - ألا تكون المفردات المختارة ممثلة لجزء (قطاع) من أجزاء المجتمع الأصلي بل يجب أن تمثل جميع أجزاء المجتمع (علام، رسلان، أساسيات الاحصاء الاجتماعي،...).

6- أنواع العينات

تنقسم العينات إلى نمطين أساسيين، عينات احتمالية وعينات غير احتمالية:

1-6 العينات الاحتمالية: ويشترط في هذا النوع من العينات ما يلي:

أ- أن يكون لجميع مفردات المجتمع الأصلي فرص متساوية للظهور في عينة البحث.

ب- أن الشرط السابق يستلزم على الباحث الحصول على قائمة بكل الوحدات الإحصائية التي تنتمي إلى المجتمع الإحصائي، الذي يعكس الظاهرة المبحوثة، مع ترقيم تلك الوحدات ترقيما تصاعديا.

و يندرج ضمن هذا النمط من العينات الاحتمالية أنواع فرعية للمعاينة الاحتمالية نذكر منها :

1-1-6- العينة العشوائية البسيطة:

و هي نوع من الطرق المعاينة التي تطبق في المجتمعات المتجانسة أين تكون درجة الاختلاف بين خصائص وحدات المجتمع الإحصائي ضعيفة؛ مثل مجتمع الموظفين الإداريين .

أما عن الخطوات العملية لسحب عينة بهذه الطريقة فهي كما يلي:

أولا: حصر كل وحدات المجتمع الإحصائي في قائمة.

ثانيا: ترقيم كل وحدات المجتمع الإحصائي دون تكرار أو إلغاء لأي وحدة.

ثالثا: القيام بعملية سحب الوحدات التي ستشكل العينة المبحوثة استنادا إلى نسبة السحب أو ما يعرف أيضا بكسر العينة؛ وتجدر الإشارة إلى أن هناك عدة طرق للسحب في هذا النوع من العينات نذكر منها: طريقة القرعة، طرقة الأرقام العشوائية وطريقة السحب الآلي .

من عيوب هذه الطريقة :

صعوبة حصر كل وحدات المجتمع الإحصائي بدقة في بعض الحالات.

6-1-2- العينة العشوائية المنتظمة:

لها نفس شروط العينة العشوائية البسيطة إلا أن طريقة السحب فيها تقوم على أسس موضوعية:

إذ يقوم الباحث بتقسيم مجتمع البحث إلى مجموعات متساوية، وفقا لوحدة أو نسبة السبر (كسر العينة) على شرط أن يكون قد جهز قائمة شاملة لكل وحدات المجتمع الإحصائي مرقمة ترقيما تسلسليا، يتم اختيار أول وحدة فقط من المجموعة الأولى بطريقة عشوائية ومن ثمة يبدأ السحب المنتظم لباقي الوحدات في هذه العينة، (محمد حسن، أساسيات الاحصاء وتطبيقاته، 1992).

مثال: لدينا مجتمع بحث يتكون من 1000 وحدة إحصائية، تم تحديد كسر العينة ب 20/1 أي أن الباحث سيأخذ وحدة إحصائية واحدة من بين 20 وحدة، بهذا فان عدد الوحدات التي ستمثل عينة البحث هو 50 وحدة إحصائية على أساس أن 50=20/1000 . بهذا فان الباحث سيقسم المجتمع الإحصائي مبدئيا إلى 50 مجموعة تتكون كل واحدة منها من 20 وحدة إحصائية ثم يبدأ

بسحب وحدات العينة بأخذ وحدة واحدة فقط من كل مجموعة، إلا أن السحب يكون عشوائيا في أول مجموعة ثم يصبح منتظما ابتداء من المجموعة الثانية إلى آخر مجموعة وذلك بإضافة طول مجال السحب، الذي هو 20 في هذا المثال، لآخر وحدة تم سحبها، كما هو موضح:

أول مجموعة تبدأ من الوحدة ذات الرقم واحد إلى الوحدة ذات الرقم 20، المجموعة الثانية تبدأ من الوحدة ذات الرقم 21 إلى الوحدة ذات الرقم 40 ... المجوعة الخمسين تبدأ من الرقم 981الى 1000، أما عن سحب الوحدات فيكون كالتالي:

بالنسبة للمجموعة الأولى يكون فيها السحب بطريقة عشوائية ولنفترض أن الوحدة ذات الرقم 5 هي أول وحدة تسحب بطريقة عشوائية: (الأرقام المضللة هي التي سيتم اختيارها بتطبيق الطريقةالعشوائية المنتظمة لتشكيل العينة).

المجموعة الأولى: 1، 2، 3، 4، 5، 6، 7، 8، 20

نضيف الى رقم المسحوب طول المجال الذي هو 20 لتكون الوحدة التالية في السحب هي صاحبة الرقم 25

المجموعة الثانية :21، 22، 23، 24، 25، 26، 40

نضيف طول المجال لآخر رقم سحب (25+20=45) فتكون الوحدة التالية هي ذات الرقم:45

المجموعة الثالثة:41، 42، 43، 44، 45، 46، 60

و التالية هي (45+20=65) بعد اضافة طول مجال السحب هي 65:

المجموعة الرابعة: 61، 62، 63، 64، 65، 66، 80

. . . . .

. . . . .

آخر مجموعة 981، 982، 983، 984، 985، 986، ...1000

وبهذا تنتهي عملية السحب المنتظم ويتحصل الباحث على حجم العينة التي سيقوم بدراستها وهي 50 وحدة احصائية في هذا المثال.

3-1-6- العينة الطبقية:

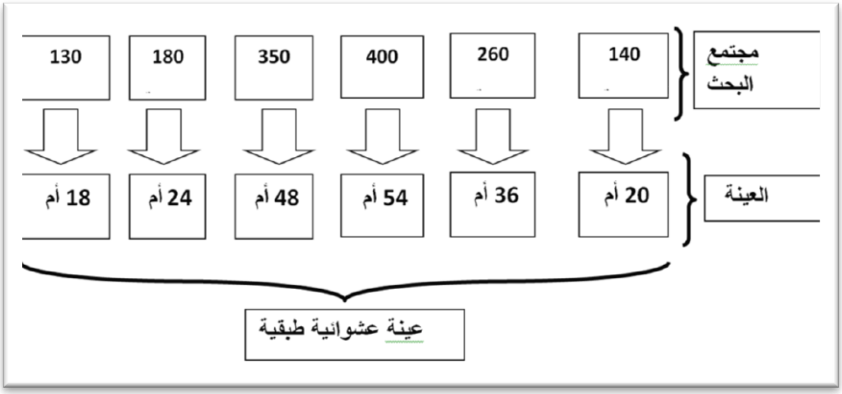
وهي من نوع العينات العشوائية التي يتوافق تطبيقها في حالة المجتمعات غير المتجانسة؛ أي المجتمعات التي يوجد بها اختلافات واضحة بين وحداتها الإحصائية، أما عن طريقة تطبيقها فتستلزم توفر نفس الشروط المذكورة بالنسبة للعينات الاحتمالية السابقة، وسميت بالعينة الطبقية لأنها تعمل على تقسيم أو توزيع مجتمع البحث إلى طبقات فرعية يتحقق بداخل كل واحدة منها نوع من التجانس، ومن ثمة يتم سحب جزء من كل طبقة على شرط أن تكون الحصة المسحوبة من كل طبقة لتشكيل العينة تمثل نفس الحصة الموجودة في مجتمع البحث. والفائدة من استخدام هذه الطريقة هو سحب عينات ممثلة لمجتمعات البحث التي توجد اختلافات كبيرة بين وحداتها(طبية، 2008).

مثال: يمثل التوزيع التالي عدد الأمهات الموظفات حسب القطاعات كالتالي:

القطاع	المجتمع	التكرار النسبي لكل طبقة		عدد الأمهات في كل فئة من العينة		fi
الزراعة	140	=1460/140	0.1	=200*0.1	20	0.1
الصحة	260	=1460/260	0.18	=200*0.18	36	0.18
التربية	400	=1460/400	0.27	=200*0.27	54	0.27
التعليم العالي	350	=1460/350	0.24	=200*0.24	48	0.24
الصناعة	180	=1460/180	0.12	=200*0.12	24	0.12
تجارة	130	1460/130	0.09	=200*0.09	18	0.09
المجموع	**1460**		1		**200**	1

نلاحظ أن حجم المجتمع الكلي من الأمهات العاملات يساوي 1460 كما يتضح من مجموع الخانة الأولى، وأراد الباحث أن يأخذ عينة بحجم 200 وحدة إحصائية من المجموع الكلي للأمهات العاملات كما يتضح من مجموع الخانة الرابعة، فقام أولا بحساب نسبة كل طبقة من هذا المجتمع كما يتضح من الخانة الثالثة، ثم حاول أن يأخذ عينة تشمل كل الطبقات، على ان يحقق نفس النسبة الموجودة في المجتمع الإحصائي بالنسبة لكل طبقة داخل العينة المسحوبة، بحيث تأخذ العينة المسحوبة نفس خصائص المجتمع الإحصائي وتكون بذلك ممثلة له أحسن تمثيل. بحيث لو حسبنا نسبة كل طبقة داخل العينة لوجدناها متساوية مع نسبة كل طبقة داخل المجتمع الإحصائي، كما يتضح من الخانة الأخيرة، وبعد تحديد عدد الوحدات التي يجب سحبها من كل طبقة، يقوم الباحث بعملية

السحب من كل طبقة إما بالطريقة العشوائية البسيطة أو بالطريقة المنتظمة المذكورتين سابقا.



شكل رقم (02)

6-2-العينات لا احتمالية:

وهي العينات التي لا تقوم على مبدأ الاحتمالية، ويصلح تطبيقها في الظواهر التي يصعب على الباحث أن يلم بمعلومات حول المجتمع الذي يقوم بدراسته، أو أن يقوم بحصر حدوده ومعلماته . من أهم أنواعها:

6-2-1-العينة المقصودة: تقوم هذه العينة على الاختيار الكيفي من قبل الباحث للوحدات التي تعكس الظاهرة المبحوثة، استنادا إلى أهداف البحث، كأن يقوم الباحث، في دراسته حول موضوع التدخين، بقصد الأفراد الذين يراهم يدخنون مباشرة ويجمع منهم المعلومات التي يحتاجها في بحثه حول التدخين وأسبابه، إذ أنه في بحث كهذا لا يمكن أن يتحصل الباحث على قائمة بكل الأفراد المدخنين أو حصر حدود ومعلمات المجتمع الذي يقوم بدراسته، حتى في منطقة محدودة، لهذا يكون الحل في الدراسة هو العينة المقصودة.

2-2-6- العينة الحصية: وهي شبيهة بالعينة الطبقية الاحتمالية لكنها من العينات لا احتمالية، ذلك لأنها لا تشترط على الباحث أن يقوم بحصر لكل وحدات المجتمع الإحصائي أو ان يكون على دراية بمعلمات هذا المجتمع، الا انها تتطلب منه معرفة الصفات العامة لهذا المجتمع والتي تخدم أغراض البحث، مثل معرفة التوزيع المئوي لمجموعة سكانية حسب الجنس أو حسب المستوى التعليمي أو حسب المنطقة السكنية، وانطلاقا من ذلك يمكن سحب عينة بالحجم الذي يحدده على أن تتحقق فيها نفس الخصائص الموجودة في مجتمع البحث كما يتضح من المثال التالي:

أراد باحث أن يعرف علاقة المستوى التعليمي بظاهرة التدخين فعرف مسبقا أن المجتمع الإحصائي الذي سيبحثه يتوزع كالتالي، 15% منهم مستوى ابتدائي، 50% منهم بين المستوى المتوسط والمستوى الثانوي، و35% منهم مستوى جامعي، وأراد أن يسحب عينة حجمها 300 وحدة إحصائية، ففي هذه الحالة سيكون عدد الوحدات في كل فئة كما يلي:

الابتدائي: سيأخذ 45 وحدة (مبحوث) انطلاقا من: (300*15)/100=45 وحدة.

الثانوي والمتوسط: سيأخذ 150 وحدة (مبحوث) انطلاقا من: (300*50)/100=150 وحدة.

الجامعي: سيأخذ 105 وحدة (مبحوث) انطلاقا من: (300*35)/100=105 وحدة.

بعد ذلك ينزل الى الميدان ويأخذ العدد المطلوب من كل طبقة حتى يجمع منه البيانات اللازمة وفق العينة المطلوبة.

7-تحديد حجم العينة

يختلف تحديد حجم العينات بين النمطين الاحتمالي وغير الاحتمالي.

7-1- تحديد حجم العينة في النمط الاحتمالي: يتحدد حجم العينة في هذا النمط بقواعد رياضية محددة يمكن اختصارها في النقاط التالية:

*إذا كان مجتمع البحث يقل عن 100 وحدة، يكون من الأفضل أخذها كلها أو بالتقريب 50 % منها على الأقل .

*إذا كان مجتمع البحث يشمل على بعض المئات أو بعض الآلاف من الوحدات، من أفضل أخذ 100 وحدة تقريبا من كل طبقة، أو بشكل عام أخذ 10 % من المجموع الكلي لمجتمع البحث إذا ما اشتمل على بعض الآلاف .

*إذا كان مجتمع البحث يشمل على بعض عشرات الآلاف من الوحدات أو أكثر، يكفي أخذ 1 % من مجموع المجتمع الكلي .

فإذا أردنا سحب عينة عشوائية من التلاميذ من مجتمع يتكون من 000 30 تلميذ من خلال تحديد رياضي مقدر أكثر أو لأقل من 5 % يكفي أخذ 379 تلميذ أي ما يقدر ب 1.26 % وإذا كان أردنا أن نسحب عينة من مجتمع مقدر ب 000 200 1 تلميذ يكفي إضافة 5 وحدات فقط أي ما يقرب 0.032 % من مجموع التلاميذ ليصبح العدد 384 تلميذ . وهو نفس العدد الذي يمكن أخذه إذا ما ارتفع مجموع عدد التلاميذ إلى 000 000 6 أي أخذ 384 تلميذ بنسبة تقدر ب 0.0006 % .

نستنتج أنه كلما كان مجتمع البحث ضيق كلما استدعى ذلك اخذ نسبة كبيرة منه لتشكيل العينة، بينما لا يشترط توسيع حجم العينة إذا كان مجتمع

البحث يفوق المليون وحدة أي يكتفي الباحث بأخذ نسبة قليلة منه تكفي لتمثيل المجتمع الكلي .

8- تحديد حجم العينات غير الاحتمالية

يكفي في العينات غير الاحتمالية وجود عدد كافي من الوحدات حتى يكون في إمكان الباحث إجراء المقارنات اللازمة.

قليلا ما يقوم الباحث عند اتباعه لنوع من العينات غير الاحتمالية تجاوز بعض المئات من الوحدات، الا اذا كان الهدف هو الأخذ بعين الاعتبار العديد من الخصائص أو المتغيرات بين الوحدات المختارة. وعلى العكس من هذا اذا وجدت حالة مختارة بدقة ومبررة، كمؤسسة أو منظمة مهما كانت أو حتى فرد واحد تم ملاحظته لمدة كافية، يمكن أن تمثل المجتمع المقصود من الناحية الكيفية. إن تحديد المشكل بدقة، هو الذي يقرر استعمال العينة غير الاحتمالية وتحديد حجمها.

في البحوث الكيفية، يقوم الموجه الثاني في تحديد حجم العينة على مبدأ تشبع المصادر La saturation des sources. هذا يعني أن الباحث يتوقف عن جمع المعطيات من وحدات المجتمع الأصلي عندما يلاحظ وجود تكرار في المعطيات التي يجمعها، وتصبح عملية الاستمرار في جمع المعطيات، من أجل إضافة شيء مفيد لفهم مشكل للدراسة، أمرا غير مجديا.

9- مصادر الخطأ في العينة

ان الأخطاء التي تقع عند استخدام اسلوب المعاينة كأسلوب لجميع البيانات تسمى أخطاء المعاينة الكلية ويمكن تقسيمها الى نوعين من الأخطاء:

1-9- خطأ المعاينة العشوائي: يمكن القول أن هذا النوع من الأخطاء نجده في البحوث التي تتبع أسلوب الحصر بالعينة دون البحوث التي تتبع أسلوب الحصر الشامل في جمع البيانات، ويمكن تقدير حجم هذا النوع من الخطأ إحصائيا، وينتج هذا النوع من الأخطاء عن(جلاطو، 2002):

- الاختلاف أو التشتت بين قيم الوحدات التي تتكون منها العينة.

- وعن الوحدات التي لم تشأ الصدفة أن ندخلها في العينة.

* ويعتمد الحجم المتوسط لهذا النوع من الخطأ على:

- حجم العينة المسحوبة من المجتمع الإحصائي.

- مدى تشتت مفرداتها (أي درجة تجانس أو اختلاف الوحدات الإحصائية فيما بينها)

- طريقة سحب وحدات العينة (عشوائية بسيطة، عشوائية منتظمة، طبقية ...)

* ويمكن التقليل من حجم خطأ المعاينة العشوائي بمالي:

-زيادة حجم العينة.

-اتباع أسلوب المعاينة المناسب الذي يساعد على التقليل من الاختلاف بين الوحدات المسحوبة لتشكيل العينة (كأسلوب العينة الطبقية أو العينة المنتظمة).

2-9- خطأ التحيز:

يمكن حدوث هذا النوع من الخطأ خلال عملية البحث ككل بعيدا عن أسلوب المعاينة، وقد نجده في كل من البحوث التي تتبع أسلوب الحصر بالعينة أو البحوث التي تتبع أسلوب الحصر الشامل، ويصعب قياس هذا النوع من الخطأ، كما أن الطريقة الوحيدة للتقليل منه هو تحلي الباحث بالموضوعية والدقة في كل

خطوة من خطوات البحث. وقد حدد الأخصائيون ثلاثة مصادر لخطأ التحيز(علام، رسلان، 198 هي:

خطأ التحيز في سحب وحدات العينة وقد تحدث بسبب:

- الاختيار غير العشوائي لوحدات العينة.

- التحيز المقصود من طرف الباحث في عملية السحب.

- استبدال وحدة إحصائية بوحدة أخرى غير وجودة في المجتمع المحدد للبحث.

- عدم التمكن من استكمال وصول كل الاستمارات.

2- خطا التحيز الناتج عن التعريف الخاطئ لوحدة المعاينة.

3- أخطاء أخرى شائعة منها:

- أخطاء عدم الاستجابة (من طرف المبحوثين).

-خطأ التبويب ومعالجة البيانات.

- أخطاء تفسير النتائج.

الفصل الرابع: تفريغ وتبويب البيانات الاحصائية

1- مفاهيم أساسية في الاحصاء
2- مستويات قياس (تصنيف) المتغيرات
3- شروط نظام التصنيف
4- خطوات تفرغ البيانات الإحصائية
5- العرض الجدولي للبيانات الإحصائية
6- طريقة التبويب اليدوي
7- الجداول المزدوجة
8- أنواع التوزيعات التكرارية

تمهيد

يعتبر عرض وتبويب البيانات الإحصائية الخطوة الثانية (بعد جمع البيانات) في مفهوم التحليل الاحصائي، ففي مجال الدراسات والابحاث في العلوم الاجتماعية عامة يحصل الباحث على معلومات ضخمة يصعب عليه وعلى القارئ وصفها أو تفسيرها بسهولة لاستخلاص بعض الحقائق أو المدلولات الاحصائية منها، لذلك يلجأ الباحث الى حصر هذه البيانات وعرضها بطريقة مختصرة تساعد على فهمها وتحليلها احصائيا للتعرف عليها ووصفها ومقارنتها بغيرها من الظواهر، والخروج ببعض المدلولات الاحصائية عن مجتمع أو عينة الدراسة.

1-مستويات قياس (تصنيف) المتغيرات

لغرض استخدام المقاييس والأساليب الإحصائية في تحليل الظاهرة موضوع البحث، فإنه يجب تحديد مستوي القياس او التصنيف للبيانات أو المتغيرات، ويعتبر التصنيف أبسط العمليات الأساسية في أي فرع من فروع العلم فالتصنيف هو تجميع للمفردات أو العناصر أو المعلومات المتشابهة إلي حد كبير المتماثلة في خصائصها مع بعضها في مجموعة، وذلك بهدف المقارنة بين المجموعات المختلفة على أساس الخواص مثال ذلك إذا قمنا بتصنيف عدد من الأفراد إلي مجموعات وفق خاصية العقيدة (مسلم – مسيحي- يهودي)، وبذلك يتم تقسيم مستويات القياس إلي أربعة أنواع هي مستوي القياس الاسمي الترتيبي الفئوي والنسبي وهذه المقاييس تختلف باختلاف طبع المتغيرات ومن حيث كمية المعلومات التي تحتويها، وبالتالي تختلف العمليات الحسابية والإحصائية التي يمكن إجراءها على كل منها، وفيما يلي شرح مبسط للمستويات الاربع السابقة. (الصغير، 2001) .

1-1- المستوى الاسمي:

هو أبسط وأدنى مستويات القياس جميعا، حيث يتم من خلاله تسمية أو تصنيف المبحوثين بالاسم فقط دون تداخل إلى مجموعات متمايزة طبقا لخصائص نوعية. وبالتالي تكون متغيرات هذا التصنيف كيفية أو وصفية مثل: الجنس (ذكر –أنثى)، الجنسية (جزائري – تونسي -...). المستوى التعليمي (ابتدائي، متوسط، ثانوي،...). الحالة الاجتماعية (أعزب، متزوج، مطلق،...).

2-1-المستوى الترتيبي :

و هو أعلى مرتبة من مستوى القياس الاسمي حيث يمكّن استخدامه من ترتيب البيانات الاحصائية ترتيبا تصاعديا أو تنازليا، بينما تستحيل هذه العملية في مستوى القياس الاسمي. وهو ما يستوجب أن تكون متغيرات هذا التصنيف كيفية ترتيبية مثل المستوى المعيشي (مرتفع، متوسط، متدني)، مستوى الذكاء (مرتفع – متوسط –ضعيف)، المستوى التعليمي (ابتدائي، متوسط، ثانوي، جامعي)

3-1- مستوى القياس الفئوي:

ويشمل خصائص القياس الوصفي والترتيبي فضلا على أن الارقام في هذا المستوى تحمل معنى كميا وهو بذلك أرقى من النوعين السابقين، وبناء عليه غالبا ما تكون متغيرات هذا التصنيف ذات طابع كمي متصل. مثل: أطوال مجموعة من الطلاب وتقسم هذه المشاهدات أو البيانات إلى فئات تالية :[110- 120[، [120-130[، [130-140[... .

4-1-المستوى القياس النسبي:

يعد من أرقى مستويات القياس الثلاثة السابقة حيث يتفوق على مستوى القياس الفئوي بأنه يمتلك سمة "الصفر" المطلق الذي يدل على انعدام الخاصية أو السمة. من امثلة المتغيرات التي تقاس على هذا المستوى: عدد الاطفال لدى الاسرة، عدد التلاميذ بالقسم عدد الكراسي بحجرة الدراسة ...إلخ. ففي المثال الاول يمكن ان يكون عدد الاطفال بالأسرة "صفر" وهو صفر مطلق، كما أن الاسر التي لديها ستة ابناء مثلا يكون بحوزتها ضعف عدد الاطفال من الاسر التي لها ثلاثة أطفال.

2- شروط نظام التصنيف

إن أي نظام يختاره الباحث لتصنيف البيانات في جداول احصائية لابد أن تتحقق فيه الشروط التالية:

1-2- عدم التداخل: أي أن لا تتداخل المجموعات التي تقسم إليها البيانات بمعنى أن تنضم كل وحدة إحصائية إلى مجموعة واحدة فقط . فمثلا الوحدة الاحصائية التي لها مستوى ابدائي تصنف في هذه الفئة فقط ولا يمكن أبدا أن تصنف في فئة أخرى.

2-2- الشمول: أن يكون لكل وحدة من البيانات مكانا لها ضمن إحدى مجموعات نظام التصنيف. أي تخصيص فئة لكل حالة من حالات المتغير ودون اهمال أي منها.

3- خطوات تفرغ البيانات الإحصائية

بعد جمع البيانات سواء من مصادرها الأصلية باستخدام الاستمارة أو من مصادر مطبوعة أو منشورة، يكون من الصعب في كثير من الأحيان تفهم أو استنتاج أي معلومات عن الظاهرة او الصفة التي جمعت عنها البيانات، لذلك

نقوم بتبويب تلك البيانات وعرضها في جداول تكون أساسا للتحليل الإحصائي. وهناك طريقتين لتبويب البيانات: الطريقة اليدوية والطريقة الآلية. وتتم عملية التفريغ باتباع الخطوات التالية:

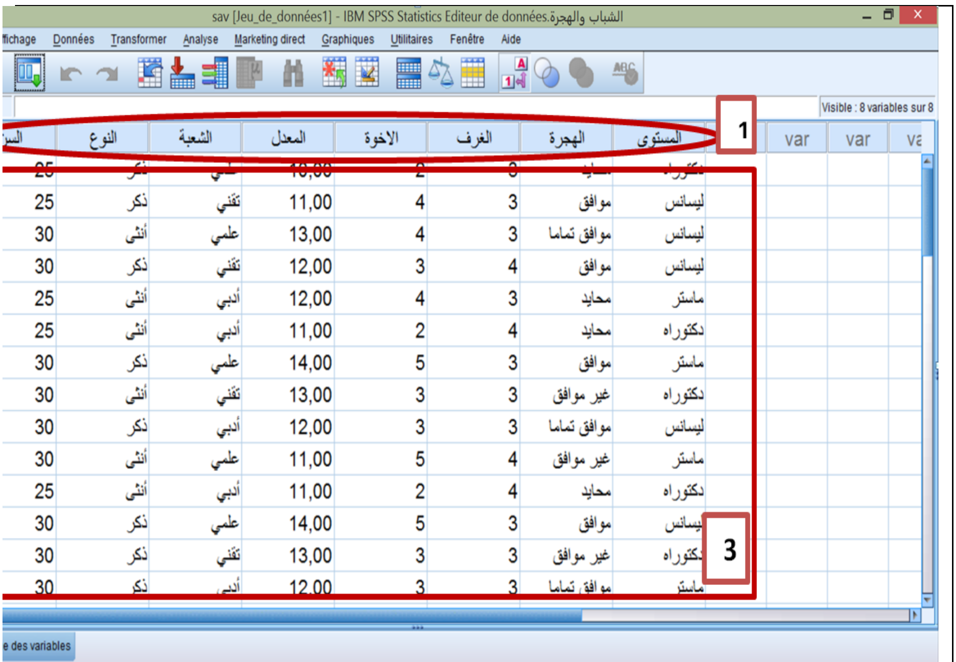
3-1- مراجعة البيانات "مرحلة الفرز": وفيها يتم فحص الاستمارات ذات الإجابات الكاملة واستبعاد الاستمارات ذات الإجابات الناقصة والمتناقضة .

3-2- ترقيم الاستمارات: وذلك بإعطاء رقم متسلسل لكل الاستمارات القابلة للتفريغ، وهي مسألة ضرورية لتسهيل عملية تفريغ الاستمارات قبل الانتقال الى انشاء الجداول الاحصائية وفق متطلبات البحث كما سيتضح خلال الخطوات التالية.

3-3- ترميز المتغيرات: أي إعطاء رمز معين لكل متغير وحالاته حتى تسهل عملية التفريغ، مثال:

المتغير	رمز المتغير	حالات المتغير	رمز حالات المتغير
النوع	A	ذكر	1
		أنثى	2
المستوى التعليمي	B	ابتدائي	1
		متوسط	2
		ثانوي	3
		جامعي	4
الموقف	C	موافق	1
		محايد	2
		غير موافق	3
عدد الاطفال	E	1 2 3	المتغير كمي منفصل يأخذ نفس القيم لأجل العد والعمليات الحسابية اللاحقة
السن	F	20 23 25 30	المتغير كمي متصل يأخذ نفس القيم لأجل تصنيفه في فئات والعمليات الحسابية اللاحقة

4-3- تفريغ البيانات (الاستمارة): وقد يتم ذلك كما سبق ذكره وفق الطرق اليدوية، أو الطريقة الآلية الحديثة باستخدام الحاسب والبرامج المخصصة لذلك الغرض مثل استخدام ورقة EXEL أو SPSS: كما يتضح من خلال الشكل التالي :



السن	النوع	الشعبة	المعدل	الاخوة	الغرف	الهجرة	المستوى
25	ذكر	علمي	10,00	2	3	محايد	دكتوراه
25	ذكر	تقني	11,00	4	3	موافق	ليسانس
30	أنثى	علمي	13,00	4	3	موافق تماما	ليسانس
30	ذكر	تقني	12,00	3	4	موافق	ليسانس
25	أنثى	أدبي	12,00	4	3	محايد	ماستر
25	أنثى	أدبي	11,00	2	4	محايد	دكتوراه
30	ذكر	علمي	14,00	5	3	موافق	ماستر
30	أنثى	تقني	13,00	3	3	غير موافق	دكتوراه
30	ذكر	أدبي	12,00	3	3	موافق تماما	ليسانس
30	أنثى	علمي	11,00	5	4	غير موافق	ماستر
25	أنثى	أدبي	11,00	2	4	محايد	دكتوراه
30	ذكر	علمي	14,00	5	3	موافق	ليسانس
30	ذكر	تقني	13,00	3	3	غير موافق	دكتوراه
30	ذكر	أدبي	12,00	3	3	موافق تماما	ماستر

شكل بياني رقم (01)

1- يمثل التحديد رقم (1) متغيرات الدراسة حيث يأخذ كل متغير عمود مستقل.

2- يمثل التحديد رقم (2) أرقام الاستمارات حيث تأخذ كل استمارة سطر مستقل.

3- أما التحديد رقم (3) فيشير الى الاجابات الواردة في الاستمارة، وتظهر هنا بدون ترميز لتسهيل الفهم.

مع الاشارة الى أن تفريغ الاستمارات على هذا الجدول يتم استمارة تلو الأخرى، فحسب الشكل أعلاه يشير السطر الاول الى الاستمارة رقم (01)، ويوجد على هذا السطر كل الاجابات على أسئلة الاستمارة الخاصة بالوحدة الاحصائية رقم (01)، اذ يبلغ من العمر (25 سنة)، الجنس (ذكر)، الشعبة في الثانوي (علمي)، المعدل المحصل عليه (10.00)، عدد الاخوة (02)، عدد الغرف (03)، الموقف من الهجرة (محايد)، المستوى التعليمي(دكتوراه).

5-3-تبويب البيانات: وذلك بوضعها في جداول احصائية بما يخدم أغراض البحث، ووفق نوع البيانات (المتغيرات).

4- العرض الجدولي للبيانات الإحصائية

1-4- الجداول الإحصائية:

يسمى الترتيب الذي توضع فيه البيانات المفروزة بالجدول الإحصائي، والجدول الإحصائي على أنواع كثيرة ومختلفة يصلح كل منها للاستخدام في حالات معينة. ولكن جميعها تهدف إلى إبراز البيانات وتوضيحها في أضيق حيز وأصغر حجم، ويحتوي الجدول في أبسط الحالات على عمودين وسطرين حيث يبين العمود الأول قيم الظاهرة أو المتغير المدروس الذي يرمز له بحرف "x"والذي قد يأخذ شكل قيم نقطية أو شكل مجالات، أما العمود أو السطر الثاني فيحتوي على تكرارات هذه القيم أو المجالات يرمز لها بالرمز "n" .

2-4- قواعد بناء الجداول الإحصائية :

وضع المختصون في الاحصاء مجموعة من القواعد لبناء الجداول الاحصائية نذكر منها(بن محمد الصغير، مقدمة في الاحصاء الاجتماعي، 2001):

أ- من الناحية الشكلية :

- عنوان واضح في أعلى الجدول، يعطي فكرة عن البيانات التي يحتويها.

- ذكر عنوان كل عمود.

- ذكر مصدر البيانات في حالة كان مصدر البيانات ثانوي (لم تجمع من طرف الباحث نفسه).

- ذكر وحدة القياس المستعملة.

ب- **فيما يخص عدد الحالات أو الفئات التي ينقسم إليها الجدول خاصة في حالة المتغير الكمي المتصل:** عند محاولة تحديد عدد فئات الجدول لا بد من التحكيم بين نوعين من العيوب في اجراء هذه العملية:

- أن وضع عدد قليل من الفئات يؤدي الى إهمال بعض الحقائق هامة حول الظاهرة أو المتغير.

- أن وضع عدد كبير من الفئات يؤدي الى وجود خانات كثيرة في الجدول بتكرارات جد ضعيفة أو معدومة (تكرار صفري)، كما قد يعطي أهمية مفرطة للقيم المتطرفة، (الحوادث العرضية)، ما يعمل على افقاد الجدول وظيفته الاساسية المتمثلة في التلخيص غير المخل للبيانات الاحصائية.

5- طريقة التبويب اليدوي

يمكن تلخيص طريقة التبويب اليدوي في تكوين جداول تفريغ يتكون كل جدول منها من ثلاثة اعمدة؛ يخصص العمود الأول منها لبيان قيم او حالات المتغير (الكمي أو الكيفي) والعمود الثاني لوضع الاشارات نتيجة فرز الاستمارات واحدة تلو الأخرى مقابل الاجابة المناسبة، وبذلك تتحول كل اجابة الى اشارة أو علامة (/) وتسهل العملية بوضع الاشارات على شكل حزم تضم كل منها خمس اجابات بمعنى خمس اشارات ($\cancel{||||}$). وهذا لتسهيل عملية العد، أما العمود الثالث فهو مخصص لوضع عدد الاشارات الموجود مقابل كل اجابة، وهو ما نطلق

عليه اسم التكرار المطلق. ويتضح ذلك من خلال الأمثلة التالية حسب نوع المتغير. (خلف عبد الجواد، الاحصاء الاجتماعي، 2013).

1-5- تفريغ جدول تكراراي في حالة بيانات متغير كيفي:

في هذه الحالة يمكن الاعتماد على عدد الصفات التي يمكن ان يأخذها المتغير الكيفي من أجل تحديد عدد فئات الجدول كما يتضح من المثال التالي:

مثال: تمثل البيانات التالية المعلومات الخام لعينة من 30 فردا حول الحالة الاجتماعية، والمطلوب انشاء جدول توزيع تكراري:

متزوج أرمل أرمل مطلق أعزب أعزب متزوج مطلق أعزب متزوج مطلق متزوج أعزب أرمل متزوج أعزب مطلق أعزب متزوج أرمل أعزب أرمل مطلق متزوج مطلق متزوج أعزب أعزب أعزب متزوج

في هذه الحالة لا بد من ضبط على عدد حالات المتغير وهي اربع حالات في هذا المثال: متزوج، أعرب، مطلق، أرمل، وانطلاقا من ذلك يتم بناء جدول يضم الحالات الأربع، على ان يقابل كل حالة خانة لعد العلامات وهو ما يتمثل في عملية التفريغ، وخانة للتكرارات المطلقة المقابل لها كما يتضح في الجدول التالي:

الجدول رقم 2: توزيع الوحدات المبحوثة حسب الحالة الاجتماعية:

الحالة الاجتماعية	العلامات	التكرار
متزوج	~~////~~ ////	9
أعزب	~~////~~ ~~////~~	10
مطلق	~~////~~	5
أرمل	~~////~~ /	6
المجموع		30

5-2- العرض الجدولي في حالة متغير كمي:

5-2-1- جدول التكرارات في حالة متغير كمي منفصل: سبق أن عرفنا أن المتغير الكمي المنفصل يأخذ عدادا صحيحا ولا يمكن أن يأخذ ارقاما عشرية. مثال: البيانات التالية تمثل عدد أفراد 20 أسرة .

5 6 2 4 5 6 3 4 7 5
7 4 7 5 3 5 7 4 3 4

الخطوات:

أ- إعادة ترتيب البيانات تصاعديا أو تنازليا :

5 4 4 4 4 4 3 3 3 2
7 7 7 7 6 6 5 5 5 5

ب- بناء جدول يوضح كل حالة من حالات المتغير (عدد أفراد الأسرة) وهذا إذا ما كانت الحالات التي يأخذها المتغير قليلة كما هو الحال في هذا المثال:

حجم الأسرة	العلامات xi	التكرارات ni
2	//	2
3	//	2
4	~~////~~	5
5	~~////~~	5
6	//	2
7	////	4
المجموع		**20**

حيث يشمل رأس كل عمود عنوان خاص به ؛ فالعمود الأول خاص باسم المتغير الذي هو حجم الأسرة، ويشمل هذا العمود على حالات المتغير ؛ فالاثنان (2) هي الحالة التي يكون فيها حجم الأسرة فردين، بينما (7) تعكس الحالة التي يكون فيها حجم الأسرة سبعة أفراد وهكذا الخ ... .

العمود الثاني خاص بالعلامات (الاشارات) وترتبط بعلامات تعكس عدد مرات ظهور حالة المتغير في كل خانة من الخانات ويستعين بها الباحثين لتسهيل عملية التفريغ، أما العمود الأخير فسمي بالتكرارات المطلقة وهو يشير إلى الرقم الذي يدل على عدد مرات ظهور حالة المتغير ويرمز له (ni) .و يضاف سطر الى نهاية الجدول يعكس المجموع الكلي للمشاهدات يرمز له بالرمز (sigma Σ).

2-2-5- في حالة متغير كمي متصل: عند دراسة متغير كمي متصل فان مجال الدراسة غالبا ما يضم قيم لا نهائية (تظهر حالات عديدة للمتغير)، ولتعذر وضع جدول بكل حالات المتغير فانه يتم تقسيم هذا المجال إلى مجالات جزئية تسمى الفئات، حيث تحدد هذه الأخيرة حسب حجم العينة وحسب توزيع الوحدات على مجال الدراسة. ولتسهيل هذه العملية وضع الإحصائي ستورج "sturges" طريقة لتحديد طول الفئات تعتمد على مجال الدراسة وحجم المجتمع والعينة تتضح في الصيغة التالية:

صيغة رقم ..		$k = \frac{E}{1 + 3.32 * \log n}$	

حيث:

- n تمثل حجم العينة.

- حيث E يمثل قيمة المدى ويحسب من الصيغة التالية: $E = x_{\max} - x_{\min}$ ؛ ويقوم على الفرق بين اكبر مشاهدة أصغر مشاهدة.
- K هي طول الفئة بالنسبة لهذه الصيغة.
- Log_n هو اللوغاريتم العشري لحجم العينة .
- اما القيمة (1+3.32*log n) فتمثل تقدير عدد الفئات حسب حجم العينة.

ملاحظة: يمكن مما سبق استخلاص أن مضمون هذه الصيغة هو حاصل قسمة المدى على عدد الفئات.

الخطوات :

بعد حساب طول الفئة من الصيغة السابقة او تحديده مسبقا وفق تجربة الباحث.

يتم الانتقال إلى ضبط حدود الفئات: وفيها يضبط الحد الأدنى للفئة Extrémité inférieur ويرمز لها بالرمز "L1"، والحد الأعلى للفئة Extrémité supérieur ويرمز لها بالرمز "L2" .

ملاحظة:

- الفرق بين الحد الأعلى للفئة والحد الأدنى للفئة يسمى **بطول الفئة Amplitude** ويرمز له بالرمز « c ». (جلاطو، الاحصاء، 2002).

$$\boxed{c = L2 - L1}$$

- أما متوسط هذين الحدين يعرف **بمركز الفئة** ويرمز له بالرمز « xi ».

صيغة رقم ..

$$xi = \frac{L1 + L2}{2}$$

مثال: بناء جدول بمتغير كمي متصل :

البيانات الخام التالية تمثل درجات 50 طالب في اختبار الرياضيات:

73 50 62 98 85 55 72 76 87 90 63 70 78
43 68 69 64 68 76 53 74 60 79 74 46 60
77 62 57 86 96 38 73 48 81 41 52 51
49 65 73 81 74 80 56 58 54 65 67 75

أ- كما سبقت الاشارة اليه نبدأ بتحديد أصغر مشاهدة $x_{\min}$ في هذه البيانات والتي هي 38 .

و اكبر مشاهدة $x_{\max}$ وهي 98.

نعلم ان حجم العينة المدروسة هو 50 وحدة احصائية.

ثم نستخدم الصيغة السابقة من اجل إيجاد طول الفئة المناسب كالتالي:

$$\begin{aligned} k &= \frac{98 - 38}{1 + 3.32 * \log 50} \\ &= \frac{60}{1 + 3.32 * 1.69} \\ &= \frac{60}{6.61} \\ &= 9.07 \cong 9 \end{aligned}$$

و منه فان طول الفئة يساوي 9 بالتقريب.

ب- ثاني خطوة هي اعادة ترتيب المشاهدات تصاعديا أو تنازليا كما يتضح:

60 60 58 57 56 55 54 53 52 51 50 49 48 46 43 41 38
74 74 74 73 73 73 72 70 68 68 67 65 65 64 63 62 62
98 96 96 90 87 86 85 81 81 80 79 78 77 76 76 75

ج- بداية عملية التفريغ داخل الجدول:

عنوان الجدول: توزيع الطلبة حسب النقاط المحصل عليها في مادة الرياضيات:

نقاط الطلبة	العلامات	التكرار n_i	مركز الفئة x_i	طول الفئة c_i
]47-38]	////	4	42.5	9
]56-47]	/// $\cancel{////}$	8	51.5	9
]65-56]	//// $\cancel{////}$	9	60.5	9
]74-65]	$\cancel{////}$ $\cancel{////}$	10	69.5	9
]83-74]	// $\cancel{////}$ $\cancel{////}$	12	78.5	9
]92-83]	////	4	87.5	9
]101-92]	///	3	96.5	9
المجموع		50		

يتضح في هذا الجدول ان طول الفئة يساوي تسعة:

وتم حسابه بالطريقة السابقة لإيجاد طول الفئة: $c_1 = 47 - 38 = 9$

$c_2 = 56 - 47 = 9$

$C_{3=65-56=9}$

كما تم حساب مركز الفئة فئة بالصيغة الموضحة أعلاه:

$$x_1 = \frac{47 + 38}{2} = 42.5$$

$$x_2 = \frac{56 + 47}{2} = 51.5$$

ملاحظة: يمكن ايجاد قيمة $x_2, x_3, \dots x_n$ بطريقة أخرى تتمثل في اضافة طول الفئة الى مركز الفئة أي :

$$\boxed{x_2 = x_1 + c_2}$$

فنجد أن: $x_2 = 42.5 + 9 = 51.5$ $x_3 = 51.5 + 9 = 60.5$

6- الجداول المزدوجة

عند دراستنا للظواهر في العلوم الاجتماعية عامة، نهتم عادة بأكثر من بعد واحد لتحليل العلاقة بين المتغيرات. وبذلك نقوم بجمع العديد من البيانات.

الجدول الإحصائي بمتغيرين:

. نفترض انه لدينا عينة بحث جمعنا حولها بيانات خاصة بمتغيرين المتغيرA والمتغيرB .

. المتغير A يمكن ان يأخذ العدد k من الحالات $A_{1,} A_{2,} A_{3,} .....A_k$

. المتغير B أيضا يمكنه ان يأخذ l من الحالات. $B_{1,} B_{2,} B_3 ......B_l$

.المجموع الكلي للعينة المبحوثة يرمز له بالرمز n.

. عدد الوحدات الإحصائية التي تمثل الحالة A_i والحالة B_j يرمزلها بالرمز n_{ij}.

. عدد الوحدات الإحصائية الممثلة للحالة A_i (واي حالة من الحالات B) تكتب

$$n_{i\bullet} = \sum_{j=1}^{l} nij$$

.عدد الوحدات الإحصائية الممثلة للفئة B_j (وأي حالة من الحالات A) تكتب

$$n_{\bullet j} = \sum_{i=1}^{k} nij$$

و يتمثل الشكل النظري للجدول المزدوج فيما يلي :

المتغير B / المتغير A	B_1	B_2	$\dots$	B_j	$\dots$	B_ℓ	المجموع
A_1	n_{11}	n_{12}	$\dots$	n_{1j}	$\dots$	$n_{1\ell}$	$n_{1\bullet}$
A_2	n_{21}	n_{22}	$\dots$	n_{2j}	$\dots$	$n_{2\ell}$	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_i	n_{i1}	n_{i2}	$\dots$	n_{ij}	$\dots$	$n_{i\ell}$	$n_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_k	n_{k1}	n_{k2}	$\dots$	n_{kj}	$\dots$	$n_{k\ell}$	$n_{k\bullet}$
المجموع	$n_{\bullet 1}$	$n_{\bullet 2}$	$\dots$	$n_{\bullet j}$	$\dots$	$n_{\bullet \ell}$	n

المطلوب: إيجاد نسب التكرار لهذا الجدول:

* نسبة الحدث (A_i, B_j) هي: $\boxed{f_{ij} = \frac{n_{ij}}{n}}$

* مجموع نسب السطر i هي: $\boxed{f_{i\bullet} = \sum_{j=1}^{\ell} f_{ij} = \frac{n_{i\bullet}}{n}}$

* مجموع نسب للعمود j هي: $\boxed{f_{\bullet j} = \sum_{i=1}^{k} f_{ij} = \frac{n_{\bullet j}}{n}}$

التوزيعات الهامشية:

* ان مجموع الأسطر تشكل التوزيع الهامشي للمتغير A وكأننا ندرس المتغير A فقط.

* مجموع الأعمدة تشكل التوزيع الهامشي للمتغير B وكأننا ندرس المتغير B فقط.

* نسبة الهامشية، $f_{i\bullet}$ تمثل نسب الفئات A_i.

* النسب الهامشية $f_{\bullet j}$ تمثل نسب الفئات B_j. (Py, statistique descriptive, 2001)

مثال: البيانات الخام التالية تعرض 20 مبحوث تبعا لمتغير الحالة التعليمية والجنس:

متعلم،ذكر أمي، أنثى متعلم،انثى متعلم،ذكر أمي، ذكر متعلم،انثى أمي، ذكر

متعلم،ذكر أمي، أنثى أمي، أنثى أمي، ذكر متعلم،انثى متعلم،انثى متعلم،ذكر

متعلم،ذكر أمي، ذكر متعلم،ذكر متعلم،ذكر أمي، أنثى متعلم،ذكر

- **بناء الجدول:** توزيع عينة من المبحوثين حسب الجنس والحالة التعليمية.

التعليم / الجنس	متعلم		أمي		المجموع
ذكر	卌 ///	8	////	4	12
أنثى	////	4	////	4	8
المجموع	12		8		20

بنفس الطريقة يمكن بناء جداول بأنواع مختلفة من المتغيرات (كمية، أو كيفية).

- **حساب النسب المئوية في الجداول المزدوجة:**

قبل حساب النسب المئوية لا بد من التمييز في الجدول بين المتغير المستقل المؤثر (السبب) والمتغير التابع المتأثر (النتيجة). ففي الجدول السابق نجد أن الجنس هو المتغير المستقل والحالة التعليمية هي النتيجة. وبناء عليه يتم تحديد اتجاه التنسيب (اتجاه حساب النسب المئوية) واتجاه قراءة الجدول، Py, (statistique descriptive, 2001)

قاعدة 1: التنسيب يكون في اتجاه المتغير المستقل. وفي المثال السابق يتبين انه الاتجاه الافقي.

قاعدة 2: القراءة تكون في اتجاه المتغير التابع. وحسب الجدول السابق تكون القراءة بطريقة عمودية.

مثال:

الحالة التعليمي الجنس	متعلم		أمي		المجموع	
ذكر	ni	%	ni	%	ni	%
	8	%66.67	4	%33.33	12	%100
أنثى	4	%50	4	%50	8	%100
المجموع	12	%60	8	%40	%100 2	

اتجاه التنسيب

- قراءة الجدول: مع العلم ان **الجنس** هو المتغير المستقل و**المستوى التعليمي** هو المتغير المستقل:
- **تبدأ القراءة بالغالبية من المجموع الكلي لكل حالة من حالات المتغير التابع**، كالتالي: (60% من مجموع عينة البحث هم من فئة المتعلمين)، **ثم القراءة بالمقارنة لهذه الحالة كالتالي:** حيث ان 66.67% من المتعلمين هم من الذكور مقابل 50% منهم من الاناث.
- **نواصل القراءة دائما نبدأ بالغالبية من المجموع الكلي للحالة الثانية من المتغير التابع المتمثلة هنا في فئة غير المتعلمين** (فنقول ان 40% من عينة البحث هي فئة غير المتعلمة)، **ثم نقارن داخل هذه الفئة بين الذكور والاناث فنكتب** (حيث يتبين ان أكبر نسبة من غير المتعلمين هي من الاناث بـ50% مقابل 33.33% فقط تعود الى الذكور).

- **الاستنتاج: نتوصل مما سبق أن نسبة التعليم لدى الذكور أعلى منها لدى الاناث.**

7- أهمية الجداول التكرارية

1- تلخيص البيانات، حيث يتم عرضها في جداول صغيرة مهما كان عدد البيانات التي تم جمعها.

2- يساعد هذا التلخيص على التعبير على المعلومات بصورة مباشرة وسريعة، مقارنة مع البيانات في حالتها الخام.

3- إمكانية المقارنة بين مجموعتين أو أكثر من البيانات من خلال عرضها في جدول واحد.

4- تسهل عملية حساب مختلف المقاييس الإحصائية.

5- كما تسهل عملية عرض البيانات بيانيا.

8- أنواع التوزيعات التكرارية

8-1- التوزيع التكراري المطلق: الذي رمزنا له بالرمز : n_i ويمثل عدد المرات التي ظهرت فيها كل حالة من حالات المتغير.

8-2- التوزيع التكراري النسبي(الحصي): وهو تكرار كل حالة من حالات المتغير بالنسب للمجموع العام يرمز له بالرمز f_i ، وهو حاصل قسمة التكرار المطلق على المجموع الكلي $\boxed{f_i = \frac{n_i}{\sum ni}}$ كما يتضح في الجدول التالي.

يفيد هذا النوع من التوزيعات التكرارية في عملية المقارنة بين حالات المتغير خاصة إذا كان عدد التكرارات المطلقة كبيرا. يمكن حسابه بالنسبة لكل أنواع المتغيرات الكمية والكيفية.

* مجموع التكرارت النسبية تساوي 1. $\boxed{\sum f_i = 1}$

8-3- التوزيع التكراري المئوي:

لا يختلف التكرار المئوي عن التكرار النسبي الا في أن التكرار المئوي نعتبر أي مجموعة من الحالات المدروسة على انها تمثل 100 حالة، وذلك لتسهيل عملية المقارنة بين مختلف حالات المتغير يرمز له بالرمز p_i ويتم الحصول عليه كالتالي: $\boxed{p_i = \frac{ni}{n} \times 100}$ كما يتضح في الجدول التالي، أيضا يمكن حسابه بالنسبة للمتغيرات الكمية والمتغيرات الكيفية.

* مجموع التكرارات المئوية تساوي 100. $\boxed{\sum p_i = 100}$

الجدول رقم 02: التالي يمثل توزيع 200 فرد حسب عدد سنوات الدراسة:

سنوات الدراسة	التكرار المطلق ni	التكرار المئوي p_i	التكرار الحصي f_i
]6-0]	40	$\frac{40}{200} * 100 = \mathbf{20}$	$\frac{40}{200} = \mathbf{0.2}$
]12-6]	80	$\frac{80}{200} * 100 = \mathbf{40}$	$\frac{80}{200} = \mathbf{0.4}$
]14-12]	50	**25**	**0.25**
]16-14]	30	**15**	**0.15**
المجموع	200	100	1

9 - التوزيعات التكرارية المتجمعة

يستخدم هذا النوع من الجداول التكرارية لمعرفة عدد المفردات في التوزيع الأكبر أو الأقل من قيمة معينة، تتمثل في الحدود الدنيا والعليا للفئات اتي يتكون منها جدول التوزيع التكراري الأصلي، وبالتالي إما أن يكون التوزيع صاعدا او نازلا.

1-9- التكرار المتجمع الصاعد: Fréquence cumulée croissante

التكرار المتجمع الصاعد لأي فئة هو تكرار هذه الفئة مضافا إليه مجموع تكرارات الفئات السابقة كما يتضح في الجدول التالي ويمكن حسابه من التكرار المطلق، التكرار المئوي والتكرار الحصي.

الجدول رقم 03: توزيع عينة من الطلبة حسب الوزن :

الوزن (كلغ)	ni	FCC مطلق		↑Pi مئوي			↑fi حصي		
	التكرار	العملية	النتيجة	التكرار	العملية	النتيجة	التكرار	العملية	النتيجة
60-62	10	=10+0	**10**	09.52	=09.52+0	**09.52**	0.10	=0.10+0	**0.10**
62-64	15	=15+10	**25**	14.29	=14.29+09.52	**23.81**	0.14	=0.14+0.10	**0.24**
64-66	20	=20+25	**45**	19.05	=19.05+23.81	**42.86**	0.19	=0.19+0.24	**0.43**
66-68	25	=25+45	**70**	23.81	=23.81+42.86	**66.67**	0.24	=0.24+0.43	**0.67**
68-70	22	=22+70	**92**	20.95	=20.95+66.67	**87.62**	0.21	=0.21+0.67	**0.88**
70-72	13	=13+92	**105**	12.38	=12.38+87.62	**100**	0.12	=0.12+0.88	**1**
Σ	105			100			1		

- يمكن استخدام الجدول التكراري المتجمع الصاعد من اجل إيجاد تقدير لعدد المفردات التي تحمل قيمة الظاهرة **عند حدود معينة أو أقل منها** وغير ظاهرة بالجدول التكراري كالتالي :

المطلوب معرف :

1- عدد الطلبة الذين **يقل** وزنها عن 66 كلغ ؟

2- عدد الطلبة الذين **يقل** وزنها عن 70 كلغ ؟

للإجابة على السؤال الأول نعود إلى التكرار المتجمع الصاعد المطلق المقابل للفئة **64-66** في الجدول السابق، الذي يتبين منه أن عدد الطلبة الذين **يقل** وزنها عن 66 كلغ هو **45**.

للإجابة على السؤال الثاني نعود إلى التكرار المتجمع الصاعد المطلق المقابل للفئة **68-70** في الجدول السابق الذي يتبين منه أن عدد الطلبة الذين **يقل** وزنها عن 70 كلغ هو **92**.

أما إذا كان المطلوب هو :

1- معرفة نسبة الطلبة الذين **يقل** وزنها عن 66 كلغ.

2- معرفة نسبة الطلبة الذين **يقل** وزنها عن 70 كلغ .

فان ذلك يتحقق بالرجوع إلى التكرار المتجمع الصاعد المئوي ما دام القيم التي نبحث عنها موجودة في حد من حدود الفئات الموضحة في الجدول.

للإجابة على السؤال الأول نعود إلى التكرار المتجمع الصاعد المئوي المقابل للفئة **68-70** في الجدول السابق، الذي يتبين منه أن نسبة الطلبة الذين **يقل** وزنها عن 66 كلغ هو 42.86% .

للإجابة على السؤال الثاني نعود إلى التكرار المتجمع الصاعد المئوي المقابل للفئة **68-70** في الجدول السابق، الذي يتبين منه أن نسبة الطلبة الذين **يقل** وزنها عن 70 كلغ هو 87.62%..

- كما يمكن استخدام الجدول التكراري المتجمع الصاعد من اجل ايجاد تقدير لعدد المفردات التي تحمل قيمة الظاهرة عند حدود معينة أو أقل منها **وغير موجودة بالجدول التكراري**، حيث يمكن حساب ذلك التقدير كالتالي :

تكرار القيمة التي نبحث عنه =التكرار المتجمع الصاعد للفئة المتضمنة+نسبة التغير * (الفرق بين التكرارين المتجمعين الصاعدين المحيطين بالقيمة)

$$n_X = FCC_{<L_1} + (\frac{x - L_1}{L_2 - L_1}) * (FCC_{<L_2} - FCC_{<L_1})$$

حيث :

$\boldsymbol{n_X}$ **:**تكرار القيمة التي نبحث عنها

$FCC_{<L1}$: هو التكرار المتجمع الصاعد الأقل من الحد الأدنى للفئة المتضمنة.

x: هي القيمة المعطاة في السؤال.

L_1: هي الحد الأدنى للفئة المتضمنة للقيمة x المعطاة في الجدول.

L_2: هي الحد الأعلى للفئة المتضمنة للقيمة x المعطاة في الجدول.

$\boldsymbol{FCC_{<L_2}}$ **:**التكرار المتجمع الصاعد الأقل من الحد الأعلى للفئة المتضمنة

مثال: من الجدول السابق اوجد نسبة الطلبة الذين **يقل** وزنها عن 65 كلغ.

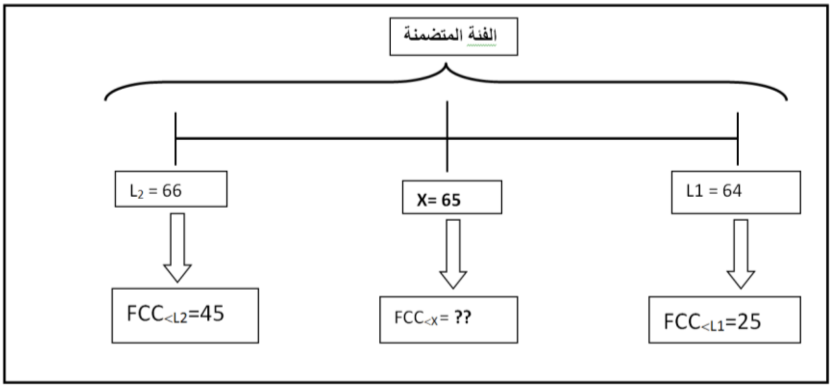
بالرجوع إلى جدول التكرار المتجمع الصاعد يمكن التوصل إلى ما يلي:

- أقل من الحد الأدنى للفئة المتضمنة ($FCC_{<L1}$=64) يقابلها التكرار المتجمع الصاعد 25 .

- أقل من x القيمة المعطاة في السؤال 65 كلغ يقابلها التكرار المتجمع الصاعد ؟؟

- أقل من الحد الأعلى للفئة المتضمنة ($FCC_{<L2}$=66) يقابلها التكرار المتجمع الصاعد 45

من خلال التعويض في الصيغة الإحصائية السابقة نتحصل على:



بتطبيق الصيغة السابقة نحصل على:

$$= 25 + \frac{650 - 640}{660 - 640} \times (45 - 25)$$

$$= 25 + \frac{10}{20} \times 20$$

$$= 25 + (0.5) * (20)$$

$$= 25 + 10 = 35$$

$$= 25 + \frac{65 - 64}{66 - 64} \times (45 - 25)$$

$$= 25 + \frac{10}{20} \times 20$$

=25+(0.5)*(20)

=25+10=**35**

ومنه فان عدد الطلبة الذين يقل وزنها عن 65 كلغ هو 35طالب.

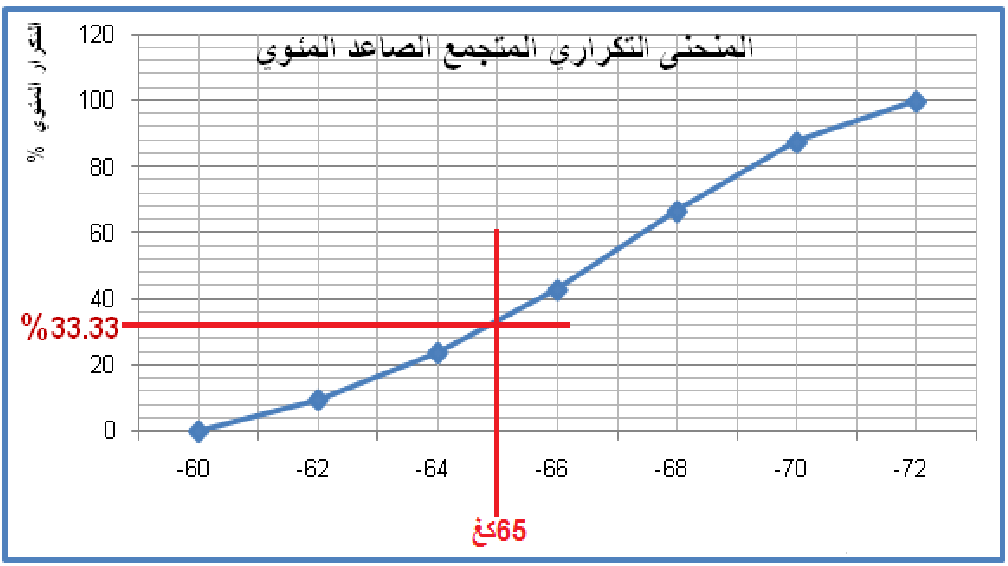
و تكون نسبة الطلبة الذين يقل وزنها عن 65 كلغ هي:

$$\frac{35}{105} \times 100 = 33.33\%$$

أي 33.33% من الطلبة يقل وزنهم عن 65 كلغ.

ومن ناحية أخرى، فانه يمكن التوصل إلى نفس النتيجة من الرسم، وذلك برسم المنحنى التكراري المتجمع الصاعد على ورق الرسم البياني، ثم إيجاد قيمة التكرار المتجمع الصاعد المقابل للقيمة 72 على المحور الأفقي، وذلك من خلال رسم خط عمودي ينطلق من محور السينات من القيمة المعطاة في السؤال(قيمة x) حيث يقطع المنحنى التكراري المتجمع الصاعد، ثم رسم خط مستقيم ثاني يكون أفقيا ينطلق من نقطة تقاطع الخط المستقيم السابق مع المنحنى التكراري المتجمع الصاعد ويقطع محور العينات حيث تعكس نقطة التقاطع تلك القيمة التي نبحث عنها والتي تشير في الشكل البياني التالي الى القيمة 33.33%، وهي نفس القيمة المحصل عليها من العمليات الحسابية السابقة، وهو ما يتضح من الشكل البياني.

شكل بياني رقم 02 :



2-9- التكرار المتجمع النازل: **Fréquence cumulée décroissante**

التكرار المتجمع النازل لأي فئة هو عبارة عن مجموع التكرارات مطروحا منه تكرارات الفئات السابقة؛ كما يتضح في الجدول التالي.

الوزن (كلغ)	ni	مطلق FCD		Pi↓ التكرار المئوي المتجمع			fi↓ التكرار المتجمع الحصي		
	التكرار	العملية	النتيجة	التكرار	العملية	النتيجة	التكرار	العملية	النتيجة
-60 62	**10**	-105 =0	**105**	09.52	=0-100	**100**	0.10	=0-1	**1**
-62 64	**15**	-105 =10	**95**	14.29	-100 =09.52	**90.48**	0.14	-1 =0.10	**0.90**
-64 66	**20**	-95 =15	**80**	19.05	-90.48 =14.29	**76.19**	0.19	-0.90 =0.14	**0.76**
-66 68	**25**	-80 =20	**60**	23.81	-76.19 =19.05	**57.14**	0.24	-0.76 =0.19	**0.57**
-68 70	**22**	-60 =25	**35**	20.95	-57.14 =23.81	**33.33**	0.21	-0.57 =0.24	**0.33**
-70 72	**13**	-35 =22	**13**	12.38	-33.33 =20.95	**12.38**	0.12	-0.33 =0.21	**0.12**
Σ	105			1					

** يمكن استخدام الجدول التكراري المتجمع النازل لحساب تقدير العدد، وبالتالي لنسبة المفردات التي تحمل القيمة عدد حدود معينة فأكثر.

9-2-1- فإذا كان المطلوب معرفة :

1- عدد الطلبة الذين يساوي وزنها أو يزيد عن 66 كلغ ؟

2- عدد الطلبة الذين يساوي وزنها أو يزيد عن 70 كلغ ؟

للإجابة على السؤال الأول نعود إلى التكرار المتجمع النازل المطلق المقابل للفئة **66-68** في الجدول السابق، الذي يتبين منها أن الطلبة الذين التي **يساوي وزنها أو يزيد** عن 66كلغ هو **60**.

للإجابة على السؤال الثاني نعود إلى التكرار المتجمع النازل المطلق المقابل للفئة **70-72** في الجدول السابق الذي يتبين منه أن عدد الطلبة الذين **يساوي وزنهم أو يزيد** عن 70 كلغ هو **13**.

6-2-2- أما في حالة كان المطلوب:

إيجاد تقدير لنسبة المفردات التي تحمل قيمة الظاهرة عند حدود معينة غير موجودة في حدود الجدول، فانه يمكن تقديرها بإتباع الخطوات الموالية.

مثال: المطلوب تقدير **نسبة** الطلبة الذين **يساوي وزنها أو يزيد** عن 65 كلغ :

الحل:

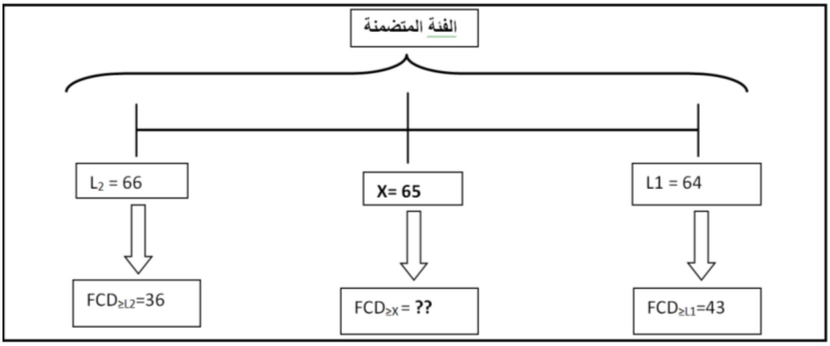
1- الفئة المتضمنة لهذه القيمة هي: 64-66.

2- 64 فأكثر يقابلها التكرار المتجمع النازل 80.

3- 63 فأكثر يقابلها التكرار المتجمع النازل ؟؟

4- 66 فأكثر يقابلها التكرار المتجمع النازل 60

ويمكن توضيح ذلك من خلال المخطط التالي :



و بتطبيق الصيغة التالية :

$$n_X = FCD_{\geq L_1} - (\frac{\mathrm{x} - \mathrm{L}_1}{\mathrm{L}_2 - \mathrm{L}_1}) * \left(FCD_{\geq L_1} - FCD_{\geq L_2}\right)$$

حيث :

n_X :تكرار القيمة التي نبحث عنها

$FCD_{\geq L_1}$:التكرار المتجمع النازل الذي يساوي أو يزيد عن الحد الأدنى للفئة المتضمنة

X: هي القيمة المعطاة في السؤال.

L_1: هي الحد الأدنى للفئة المتضمنة للقيمة x المعطاة في الجدول.

L_2: هي الحد الأعلى للفئة المتضمنة للقيمة x المعطاة في الجدول.

$FCD_{\geq L_2}$:التكرار المتجمع النازل الذي يساوي أو يزيد عن الحد الأعلى للفئة المتضمنة

بتطبيق الصيغة الأخيرة نحصل على :

$$n_X = 80 - \left(\frac{65 - 64}{66 - 64}\right) * (80 - 60) = 70$$

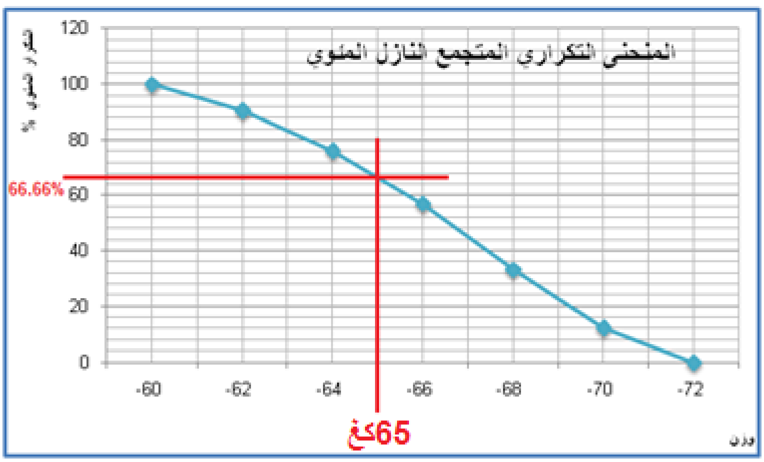
و منه فان عدد الطلبة التي يساوي وزنها 65كلغ فأكثر هو 70.

أما النسبة $\boxed{66.66\%}$ =100*(70/105)

يمكن الوصول إلى نفس النتيجة من خلال الشكل البياني للمنحنى التكراري المتجمع الصاعد :

وذلك برسم المنحنى التكراري المتجمع النازل على ورق الرسم البياني، ثم إيجاد قيمة التكرار المتجمع النازل المقابل للقيمة 65 على المحور الأفقي، وذلك من خلال رسم خط عمودي ينطلق من محور السينات من القيمة المعطاة في السؤال (قيمة $x=65$) حيث يقطع هذا الخط المنحنى التكراري المتجمع النازل، ثم رسم خط مستقيم ثاني يكون أفقيا ينطلق من نقطة تقاطع الخط المستقيم السابق مع المنحنى التكراري المتجمع النازل ويقطع محور العينات حيث تعكس نقطة التقاطع تلك القيمة التي نبحث عنها والتي تشير في الشكل البياني التالي إلى القيمة **66.66%**، وهي نفس القيمة المحصل عليها من العمليات الحسابية السابقة.

شكل بياني رقم 03:



الفصل الرابع: العرض البياني للجداول الاحصائية

1-التمثيل البياني في حالة المتغيرات الكيفية

1-1- الأعمدة البيانية البسيطة

1-2- الدائرة النسبية

1-3- الأعمدة المنزلقة

1-4- الأعمدة المجزأة

2- التمثيل البياني في حالة متغير كمي منفصل

2-1- في حالة جدول بسيط

2-2- الخط البياني

2-3- التمثيل البياني لمتغير كمي متقطع

3- التمثيل البياني في حالة متغير كمي متصل

3-1- المدرج التكراري

3-1-1- المدرج التكراري في حالة جدول منتظم

3-1-2-المدرج التكراري في حالة جدول غير منتظم

3-2- المضلع والمنحنى التكراري

3-2-1- في حالة جدول منتظم

3-2-2- في حالة جدول غير منتظم

تمهيد

تقوم عملية التمثيل البياني للمعطيات الإحصائية على إعادة عرض الجداول التكرارية في شكل رسوم بيانية تختلف باختلاف نوع المتغير الذي يبنى عليه الجدول، وذلك من أجل إعطاء تصور أبسط وأوضح مما تقدمه الجداول التكرارية عن توزيع المتغير المدروس.

1-التمثيل البياني في حالة المتغيرات الكيفية

في حالة كان المتغير كيفيا يمكن استخدام الأنواع التالية من الأشكال البيانية :

1-1- الأعمدة البيانية البسيطة: تستخدم الأعمدة البيانية في الحالات التالية:

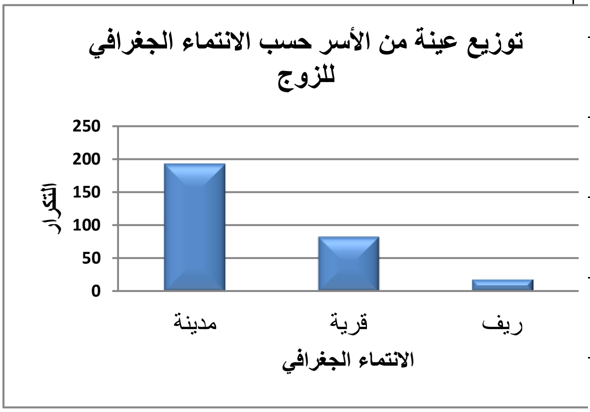
- عندما تكون أجزاء الظاهرة المقارنة كثيرة العدد نسبياً، حيث يصعب تمثيلها بالدائرة المجزأة لأنها تبدو مكتظة لدرجة يصعب معها مقارنة التوزيع النسبي للظاهرة المدروسة.
- عندما تكون الأجزاء المقارنة في فترات زمنية مختلفة .
- عندما نرغب في توضيح قيم الأجزاء المقارنة المختلفة للظاهرة موضع عبر الزمن سواء لظاهرة واحدة أو لعدة ظواهر بين فترات زمنية مختلفة .

مثال: الجدول الإحصائي التالي يبين توزيع عينة من الاسر حسب الانتماء الجغرافي للزوج:

جدول رقم: -4-

المنطقة	التكرار	النسبة
مدينة	193	65,42
قرية	83	28,14
ريف	19	6,44
المجموع	295	100

شكل بياني رقم: -4-



طريقة العمل تتم باختيار مقياس رسم مناسب مع مساحة الورق المخصص للرسم.

- نرسم محورين متعامدين، المحور الأفقي خاص بحالات المتغير والمحور العمودي خاص بقيم هذا المتغير، ومن اجل تنظيم القيم على هذا المحور لابد التمييز بين اكبر تكرار وأصغر تكرار في الجدول، والجدول الذي لدينا نجد ان أكبر تكرار هو 17270 فرد أجير لدى القطاع العام، يمكن في هذه الحالة اختيار سلم رسم يبدأ من 0 إلى 20000 بالنسبة للسنتمتر الواحد بمعنى كل 1 سم يمثل 20000 فرد.
- يبقى تحديد عرض قاعدة كل عمود اختياري.
- أما ارتفاع كل عمود يحدد وفق تكرار كل حالة.

ملاحظة: يمكن رسم الأعمدة البيانية باستخدام التكرار المطلق أو التكرار المئوي.

الشروط الواجب توفرها في كل شكل:

- عنوان واضح للشكل البياني.
- ضبط وحدة القياس.
- وضع الشكل في إطار.
- مصدر البيانات.

2-1- الدائرة النسبية :

تمثل الدائرة مجموع القيم الكلية للظاهرة، فيتم تقسيمها إلى قطاعات جزئية تتناسب مع قيم المجموعات التي تتكون منها الظاهرة، وتميز تلك القطاعات عن بعضها بألوان أو بظلال مختلفة من اجل ضمان الإيضاح.

و يستخدم هذا النوع من الأشكال البيانية في الحالات التالية من المتغير الكيفي:

- عندما يكون الهدف منها مقارنة الأجزاء المختلفة بالنسبة للمجموع الكلي.
- عندما تكون الأجزاء المقارنة قليلة العدد نسبيا.
- كما يمكن استخدامها ايضا لتوضيح التطور النسبي لأجزاء الظاهرة في فترات زمنية مختلفة.

و لتوضيح ذلك نستخدم نفس المثال السابق:

*- ترسم دائرة بمقياس رسم مناسب معتمدين على حساب نصف قطر الدائرة؛

نصف قطر الدائرة = $\boxed{\frac{\sqrt{n}}{2}}$

في مثالنا نجد: نصف قطر الدائرة = $78.62 = \frac{157.25}{2} = \frac{\sqrt{24728}}{2}$

منه نحدد مقياس رسم مناسب وليكن كل واحد سم يساوي 25 اذن نصف قطر الدائرة سيكون 3.93 بالتقريب 4 سم؛ وهي حاصل قسمة نصف قطر الدائرة على مقياس الرسم المختار. (Py, 2010)

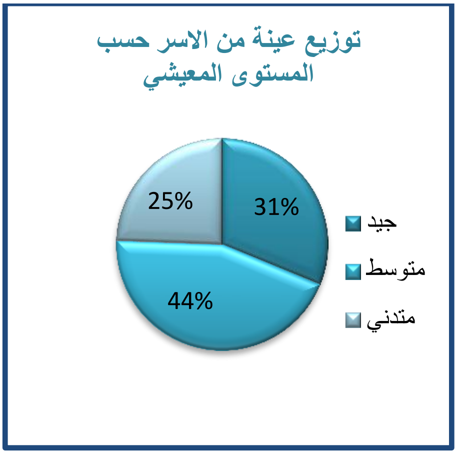
*- نحسب الزاوية المركزية لكل تكرار: والتي نرمزلها بالرمز α ألفا.

$$\boxed{\alpha = \frac{ni}{n} \times 360} \text{ أي } \boxed{\alpha = fi \times 360}$$

جدول رقم 5: توزيع عينة من الاسر حسب المستوى المعيشي:

شكل بياني رقم -5-

المستوى المعيشي	التكرار	النسبة	قيس الزاوية
عالي	92	31.19	112.26
متوسط	131	44.41	159.87
متدني	72	24.41	87.87
المجموع	295	100	360



ملاحظة: في حالة التمثيل البياني بواسطة نصف الدائرة نضرب في 180 بدل 360.

قيم الزاوية المركزية لكل تكرار في الجدول السابق تنعكس في العمود الثالث من الجدول السابق.

*- نقوم برسم الدائرة ونحدد عليها قيمة كل قطاع انطلاقا من درجة الزاوية المركزية.

3-1- الأعمدة المنزلقة: حالة جدول مزدوج (بمتغيرين) :

وتستخدم عندما يكون المطلوب عرض ظاهرة معينة في عدة فترات زمنية أو عدة أماكن جغرافية أو حتى بين منشأتين مختلفتين . ففي هذه الحالة تمثل كل فترة بعدد من الأعمدة المتلاصقة مساوية لعدد أوجه الظاهرة، مع تمييز كل عمود بلون أو تظليل معين.

فبدلاً من رسم شكلين بيانيين منفصلين يصعب مقارنتهما، فإنه يفضل انشاء رسم بياني واحد تكون فيه قيمة الأعمدة أزواجاً متجاورة يمثل كل زوج نفس الخاصية في مجموعتي البيانات. ويبين الشكل التالي هذا النوع من الرسم البياني بالأعمدة المتلاصقة أو المتجاورة .

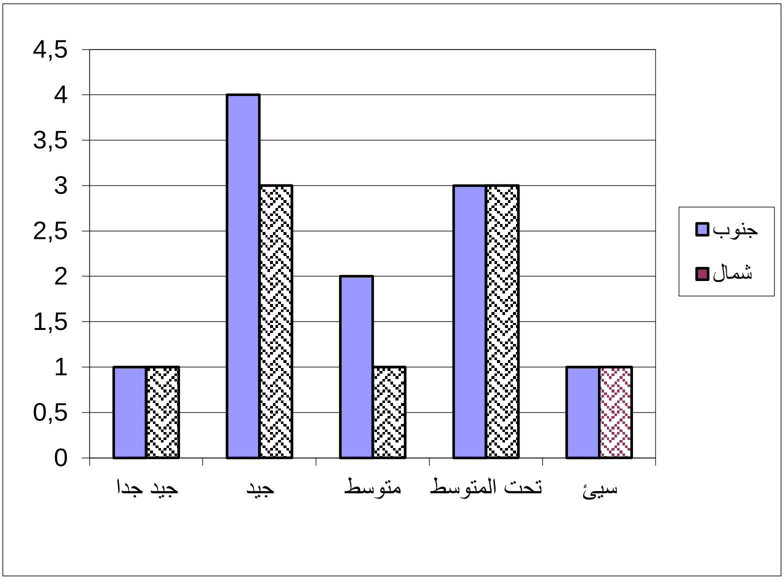
مثال: توزيع عينة من سكان الشمال وعينة من سكان الجنوب حسب رأيهم في خدمة من الخدمات:

جدول رقم -6-

	جنوب	شمال
جيد جدا	1	1
جيد	4	3
متوسط	2	1
تحت المتوسط	3	3
سيئ	1	1

شكل بياني رقم -6- :

أعمدة بيانية تمثل: توزيع عينة من الشمال وعينة من الجنوب حسب رأيهم في خدمة من الخدمات



1-4- الأعمدة المجزأة :

ويتم اللجوء لهذا النوع من الرسوم عندما تكون القيمة الإجمالية للظاهرة موزعة على مجموعات فرعية متميزة، ويشمل البيان عدة فترات زمنية أو جغرافية. يتم اللجوء إلى رسم عمود يمثل قيمة الظاهرة الكلية لكل فترة زمنية أو منطقة جغرافية، ويجزأ هذا العمود إلى أجزاء تتناسب مع قيم المجموعات الجزئية (الفئات) ويميز كل جزء من الأعمدة بألوان معينة أو بأشكال وتظليلات مختلفة لكل مجموعة فرعية على حدة. **يستعمل في حالة جدول بمتغير واحد أو جدول بمتغيرين.**

*- من اجل تسهيل عملية رسم الأعمدة المجزأة يمكن الاستعانة بالتكرار المتجمع الصاعد حتى تسهل عملية ترتيب مساحة كل فئة في شكل طبقات متراصة فوق

بعضها، بحيث يتم تحديد الخط الفصل بين الفئة والأخرى انطلاقا من قيمة التكرار المتجمع المقابل لها. كما يتضح في الأمثلة التالية.

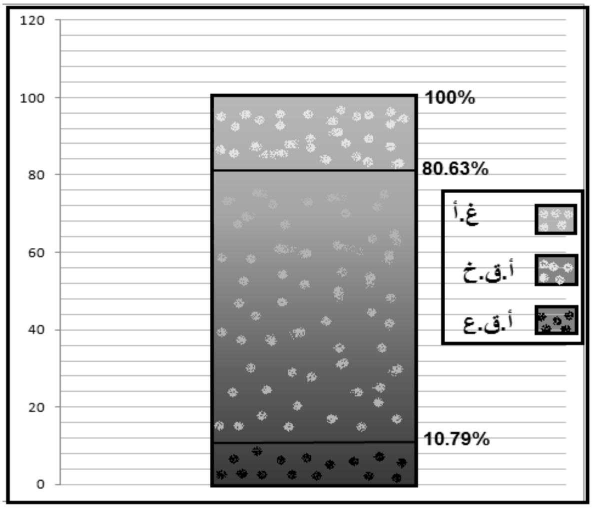
*- نظرا للاختلاف في إجمالي القيم الذي قد يعوق المقارنة إلى حد ما وخصوصاً عندما يكون المقصود هو المقارنة النسبية. يمكن تسهيل هذه المقارنة يمكن اللجوء إلى استخدام النسب المئوية، والتكرار المتجمع المئوي.

1-4-1- في حالة جدول بمتغيرين:

بالرجوع الى الجدول رقم -7- وحساب التكرار المتجمع الصاعد المئوي يمكن تمثيل تلك النسب بواسطة العمود المجزأ كما يتضح في الشكل التالي:

الوضع المهني	التكرار ni	Pi	Pi↗
أجير لدى القطاع العمومي	2669	10.79	**10.79**
أجير لدى القطاع الخاص	17270	69.84	**80.63**
غير أجير	4789	19.37	**100**
المجموع	24728	100	

شكل بياني رقم -7-: عمود مجزأ يمثل توزيع عينة من العمال حسب القطاع المهني:



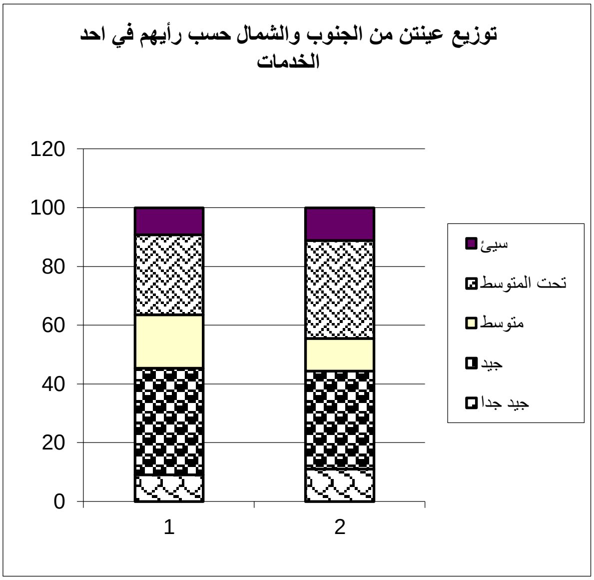
1-4-2- في حالة جدول بمتغيرين:

فبتحويل القيمة الواردة في الجدول رقم -2- السابق إلى نسب مئوية نحصل على :

جدول رقم -8-:

	جنوب	P_i	Pi↗	شمال	P_i	Pi↗
جيد جدا	1	9.09	9.09	1	11.11	11.11
جيد	4	36.36	45.45	3	33.33	44.44
متوسط	2	18.18	63.63	1	11.11	55.55
تحت المتوسط	3	27.27	90.9	3	33.33	88.88
سيئ	1	9.09	99.99	1	11.11	99.99
المجموع	11	100		9	100	

شكل بياني رقم -8- :



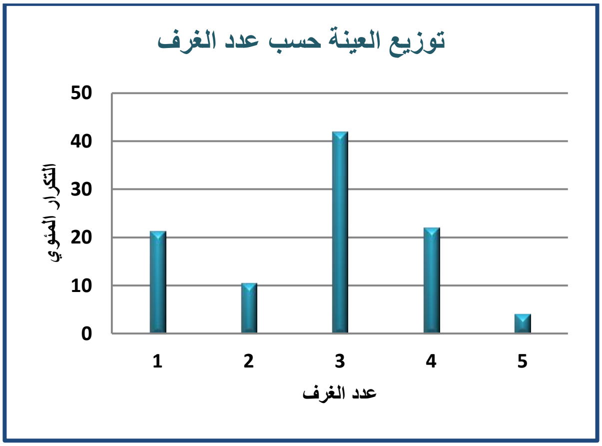
2- التمثيل البياني في حالة متغير كمي منفصل

2-1- في حالة جدولا بسيط :

مثال: الجدول التالي يمثل توزيع عينة من الاسر حسب عدد الغرف :

جدول رقم -9- شكل بياني رقم -9- التمثيل البياني بواسطة العصي :

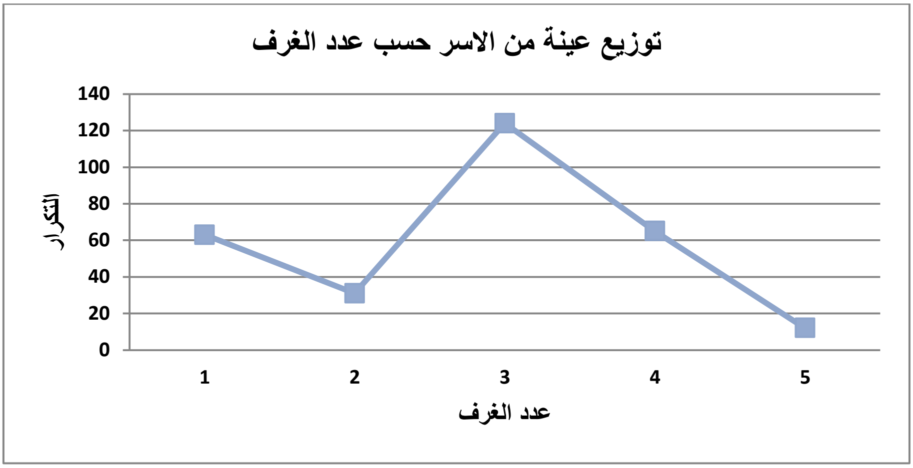
عدد الغرف	التكرار	النسبة
1	63	21,36
2	31	10,51
3	124	42,03
4	65	22,03
5	12	4,07
المجموع	295	100



2-2- الخط البياني :

للتعبير عن العلاقة بين ظاهرتن بخط بياني، يتم رسم محورين متقاطعين في نقطة الأصل، يمثل المحور الأفقي إحدى الظاهرتين والمحور الرأسي الظاهرة الأخرى، وقد يكون الموضوع متعلقاً بتمثيل قيم ظاهرة ما خلال فترة زمنية معينة، وفي هذه الحالة يكون المحور الأفقي ممثلاً للفترات الزمنية المتعاقبة، ويمثل المحور الرأسي قيم الظاهرة في تلك الفترة الزمنية . وتمثل أزواج القيم لكلا الظاهرتين على الرسم ويوصل بين تلك النقاط على الخط البياني. (Py, La statistique (sans formule statistique

شكل بياني رقم -10-: التمثيل البياني بواسطة الخط البياني :



3-2- التمثيل البياني لمتغير كمي متقطع:

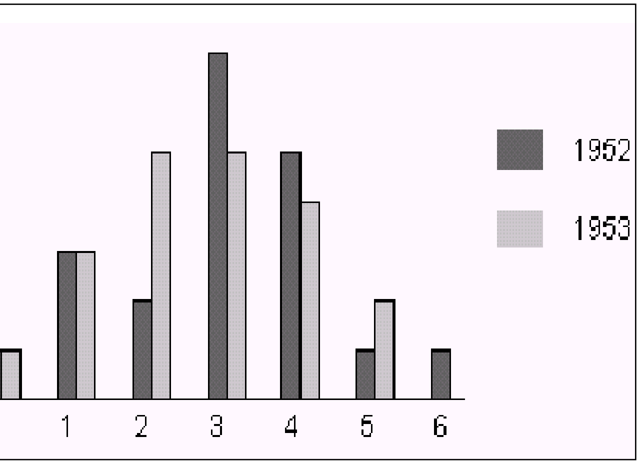
وتستخدم هذه الطريقة لعرض ظاهرة واحدة في عدة فترات زمنية حيث يخصص لكل فترة زمنية عمود مستقل مع ترك مسافات متساوية بين عمود وآخر وبالتالي نحصل على شكل عبارة عن مجموعة أعمدة متجاورة.

مثال: الجدول التالي يمثل توزيع الأسر حسب عدد الأطفال لدى النساء اللاتي ولدن سنة 1952 واللاتي ولدن سنة 1953:

أعمدة بيانية تمثل توزيع الأسر حسب عدد الأطفال لدى النساء اللاتي ولدن سنة 1952 واللاتي ولدن سنة 1953 :

جدول رقم -10- شكل بياني رقم -11- :

عدد الاطفال	1952	1953
0	1	1
1	3	3
2	2	5
3	7	5
4	5	4
5	1	2
6	1	0
المجموع	20	20



3- التمثيل البياني في حالة متغير كمي مستمر

1-3- المدرج التكراري: هو واحد الاشكال البيانية الهامة في تمثيل الجداول التكرارية ذات المتغير الكمي المتصل كالسن، الدخل، الطول والوزن .....، وهو عبارة عن أعمدة متراصة حيث :

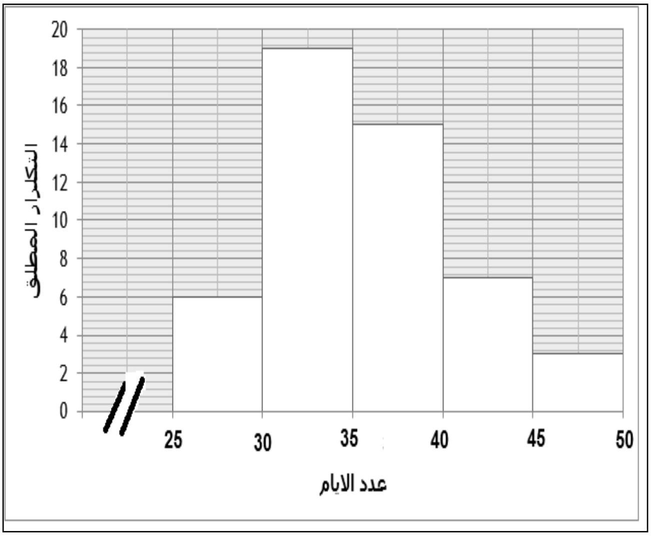
- تكرار الفئة يمثل ارتفاع المستطيل.
- و طول الفئة يمثل قاعدة كل مستطيل.

1-1-3- في حالة جدول منتظم:

مثال توزيع عدد من الأيام حسب حجم المبيعات لآلة معية في اليوم :

جدول رقم -11- شكل بياني رقم -12-

الأيام	التكرار
30-25	6
35-30	19
40-35	15
45-40	7
50-45	3
المجموع	50



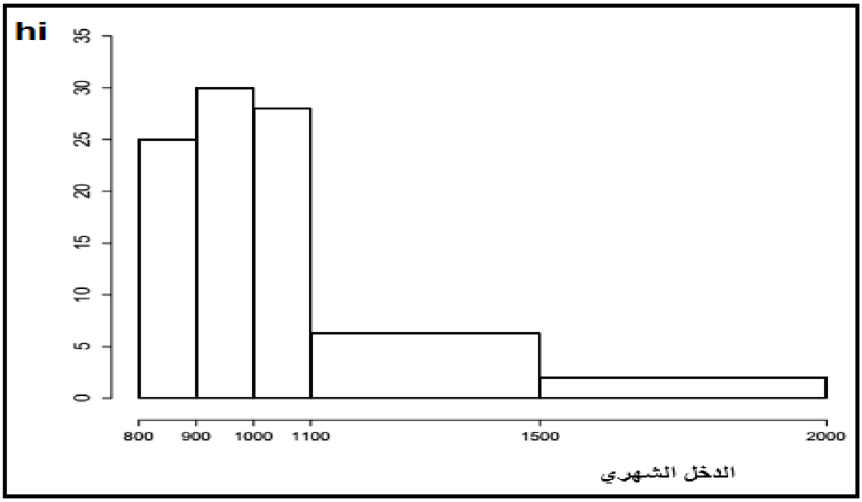
3-1-2-المدرج التكراري في حالة جدول غير منتظم:

مثال: توزيع عينة من الأجراء حسب الدخل الشهري:

جدول رقم -12- :

فئات الدخل	التكرار ni	التكرار الحصي fi	طول الفئة a_i	التكرار المعدل $h_i=(n_i*c_{\min})/c_i$	FCC
800-900	25	21.2	100	(100*25)/100= **25**	25
900-1000	30	25.4	100	(100*30)/100= **30**	55
1000-1100	28	23.7	100	(100*28)/100= **28**	83
1100-1500	25	21.2	400	(100*25)/400= **6.25**	108
1500-2000	10	8.5	500	(100*10)/500= **2**	118
المجموع	118	100			

شكل بياني رقم -13-: مدرج تكراري يوضح توزيع عينة من العمال حسب الدخل الشهري :



3-2- المضلع والمنحنى التكراري:

المضلع التكراري هو خط منكسر يتشكل عن طريق تقاطع فواصل محور السينات المتمثلة في مراكز الفئات مع تراتيب محور العينات المتمثلة في قيمة التكرار، وهو يشترك في طريقة رسمه مع المنحنى التكرار، الا أن الاول (المضلع) يتم فيه إيصال النقاط بواسطة خطوط مستقيمة باستعمال المسطرة بينما المنحنى يتم استعمل اليد لإيصال نقاط التقاطع بواسطة خط منحني.

خطوات رسم المضلع التكراري والمنحنى التكراري (Py, 2010)

- حساب مركز الفئات
- حساب مركز فئة افتراضي في بداية التوزيع بطرح نصف طول الفئة القاعدي (طول الفئة الأصغر في الجدول) من الحد الأدنى، للفئة الأولى في الجدول.

مثال من الجدول رقم (7) يكون مركز الفئة الافتراضي في بداية التوزيع كالتالي (800-(2/100)= 750، وهو مركز الفئة الافتراضي الأول، الذي سيضاف الى محور السينات عند الرسم كما يضح في الشكل البياني.رقم

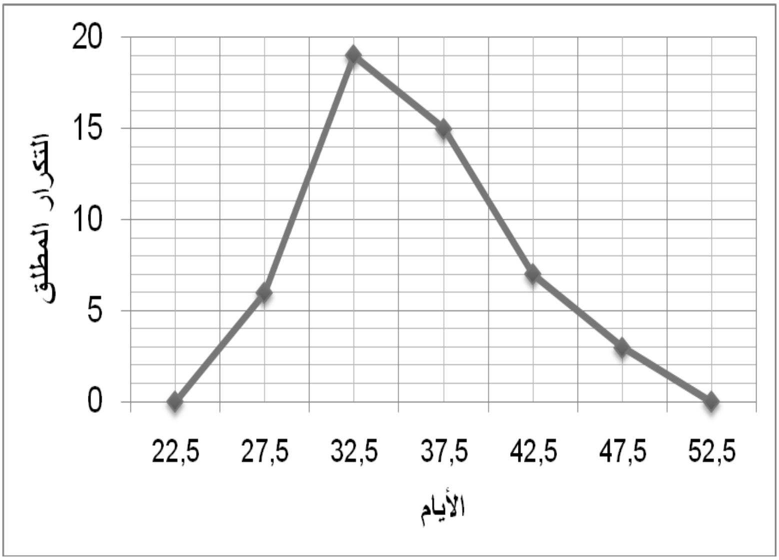
- حساب مركز فئة افتراضي في نهاية التوزيع بإضافة نصف طول الفئة القاعدي في الجدول إلى الحد الأعلى لآخر فئة في الجدول.

مثال من الجدول رقم (7) يكون مركز الفئة الافتراضي في بداية التوزيع كالتالي 2000+(2/100)=2050 وهو مركز الفئة الافتراضي الأخير.

- إعطاء القيمة "صفر" كتكرار لمركزي الفئة الافتراضيين الأول والأخير في الجدول، بهدف ربط المضلع التكراري أو المنحنى التكراري بمحور السينات عند النقطة صفر. كما يضح في كل من الشكلين -11- و-12-، وذلك من أجل تحديد المساحة تحت المضلع أو المنحنى التكريري، التي يجب أن تتساوى مع المساحة التي يحددها المدرج التكراري.

3-2-1- في حالة جدول منتظم (أطوال الفئات فيه متساوية): حسب الجدول رقم -13- السابق يكون الشكل كالتالي :

شكل بياني رقم -14-: مضلع تكراري يمثل توزيع الأيام حسب عدد المبيعات كل يوم :



3-2-2- في حالة جدول غير منتظم:

أي في حال كان أطوال الفئات غير متساوية في الجدول كما هو الحال في الجدول رقم -7- فان المضلع التكراري لا يرسم الا بعد القيام بتعديل التكرارات كما هو موضح في الجدول رقم -7- أي إيجاد قيمة التكرار المعدل الذي يتوافق وأطوال الفئات والذي يرمز له بالرمز "hi" :

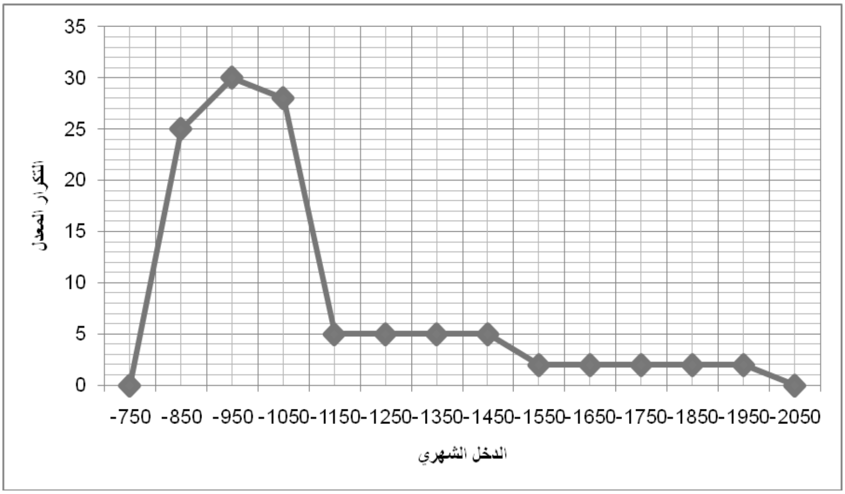
$$h_i = (n_i * c_{\min}) / c_i$$

حيث أن :

- n_i تشير الى التكرار الخاص بكل فئة.
- C_{min} تشير الى طول الفئة القاعدي (أي أصغر طول فئة في الجدول).
- C_i تمثل طول الفئة الخاص بكل فئة.
- يتم الاعتماد على التكرار المعدل "h_i " في عملية رسم المضلع والمنحنى التكراريين:

لتكون طريقة الرسم كما يلي:

شكل بياني رقم -14-: توزيع عينة من العمال حسب الدخل الشهري:



الفصل الخامس: مقاييس النزعة المركزية

<u>تمهيد</u>

تمهيد

إن الأسلوب البياني في تحليل ودراسة الظواهر لتحديد الخصائص والاتجاهات والعلاقات، يعتمد في دقته على دقة التمثيل البياني نفسه وبذلك ربما تختلف الخصائص من رسم إلى آخر لنفس الظاهرة، وعليه فإنه من الأفضل اللجوء إلى طرق القياس الكمي، حيث يستخدم الباحث الطريقة الرياضية في القياس. فالهدف الأساسي من استخدام مقاييس النزعة المركزية هو تلخيص البيانات في محاولة أخرى لوصفها عن طريق التعرف على مركزها ومن خلال مؤشرات النزعة المركزية يمكن للباحث فهم أبعاد الظاهرة قيد الدراسة. وسيم خلال النقاط التالية التعرف على هذه المقاييس بدءا من أبسطها (المنوال) الى ادقها وأهمها المتمثل في المتوسط الحسابي.

1- المنوال: le mode

يعتبر المنوال واحد من مقاييس النزعة المركزية الهامة بل هو المقياس الوحيد الذي يستخدم لحساب المتوسط لظاهرة لا يمكن قياسها بالمقياس الكمي كالمهنة، النوع ،اللون ......و **يعرف المنوال بأنه القيمة الأكثر شيوعا في التوزيع**، أو بمعنى أدق هو النقطة التي تدل على اكثر درجات التوزيع تكرارا . يروم للمنوال بالرمز: «Mo »

1-1-المنوال في حالة بيانات غير مبوبة:

وقد لا يكون للقيم (مجموعة المشاهدات) منوال وقد يوجد أكثر من منوال واحدكما يتضح لاحقا.

مثال 1 - المجموعة 22,5,7,**9,9,9**,10,10,11,12,18 لها منوال واحد وهو 9 التوزيع الذي له منوال واحد يسمى توزيع وحيد المنوال . Unimodal . .

مثال 2 - المجموعة 3,5,8,10,15,16 ليس لها منوال .

مثال 3 - المجموعة 2,3,**4,4,4**,5,5,**7,7,7**,9 لها منوالان وهما 7,4 وتسمى توزيع ذو منوالين . Bimodal

1-2- حساب المنوال في حالة متغير كمي منفصل: يمكن معرفة المنوال بسهولة عندما نقارن تكرار المشاهدات لنبحث عن أكبرها، والجدول التالي يوضح ذلك:

Xi	12 13 14 15 16 17	∑
ni	3 7 10 8 6 2	36

المنوال هنا هوا لدرجة 14 لان تكرارها يساوي 10 وهو اكبر التكرارات، نكتب:

$M_o = 14$

و **يقرأ:** اغلبية العينة تحصلت على النقطة 114.

1-3- حساب المنوال في حالة متغير كمي متصل :

لحساب المنوال في حالة متغير كمي متصل تتبع الخطوات التالية :

أ. **تحديد الفئة المنوالية: يتأثر حساب المنوال بطبيعة الفئات في الجدول، لذلك تحدد الفئة المنوالية حسب الحالات التالية:**

-في حالة جدول منتظم (أطوال فئاته متساوية): الفئة المنوالية هنا هي الفئة مقابلة لاكبر تكرار.

في حالة جدول غير منتظم (اطوال فئاته غير متساوية): فان الفئة المنوالية في هذه الحالة هي المفئة المقابلة لأكبر تكرار بعد عملية تعديل التكرارت"hi".

ب. **تطبيق الصيغة التالية:** $Mo = L1 + \frac{\Delta 1}{\Delta 1 + \Delta 2} * C$

حيث :- L1 هو الحد الأدنى للفئة المنوالية.

-$\Delta 1$ هو الفرق بين تكرار الفئة المنوالية والتكرار السابق لها.

-$\Delta 2$ هو الفرق بين تكرار الفئة المنوالية والتكرار الموالية لها.

- c طول الفئة المنوالية .

مثال: الجدول التالي يمثل توزيع عينة من المبحوثين حسب السن:

السن	50-40	60-50	70-60	80-70	Σ
Ni	6	8	12	9	35

1- أطوال الفئات متساوي إذن الفئة المنوالية هي 70-60 باعتبارها الفئة المقابلة لأكبر تكرار (12).

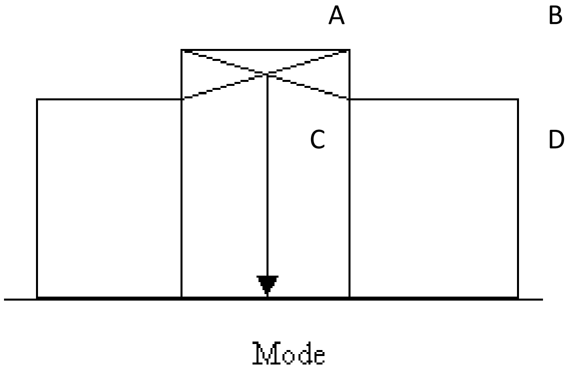
2- نعوض القيم في الصيغة السابقة كالتالي:

$$Mo = 60 + \frac{12-8}{(12-8)+(12-9)} * 10 = 65.71$$

لتكون قيمة المنوال في هذا الجدول تساوي 65.71. **ويقرأ:** أغلبية المبحوثين تقدر اعمارهم بـ 65.71 سنة.

4-1- ايجاد قيمة المنوال بيانيا: وتعتمد عل رسم المدرج التكراري.

لاستخراج المنوال بيانيا :نستعين برسم المدرج التكراري مع مراعاة أن تكون الفئات متساوية وإلا فيتم تعديل التكرارات: فالمستطيل الأكثر ارتفاعا في الشكل يعكس تكرار الفئة المنوالية. ولإيجاد قيمة المنوال نقوم بربط الحد الأدنى للفئة المنوالية التي رمز لها بالرمز (A)بالحد الأدنى للفئة اللاحقة للفئة المنوالية التي رمز لها بالرمز (D) بواسطة خط مستقيم كما هو واضح في الشكل، ثم ربط الحد الأعلى للفئة المنوال التي رمز لها بالرمز (B) بالحد الأدنى للفئة السابقة لها ذات الرمز (C) بواسطة خط مستقيم وبإسقاط نقطة تقاطع هذين الخطين على محور السينات تتبين قيمة المنوال كما يوضح الشكل:



5-1-حساب المنوال من جدول غير منتظم (أطوال فئاته غير متساوية):

البيانات التالية تمثل توزيع عينة من المبحوثين حسب السن :

السن	التكرار ni	طول الفئة ci	التكرار المعدل hi
20-10	9	10	9=(10/10)/9
30-20	13	10	13=(10/10)/13
50-30	22	20	11=(10/20)/22
60-50	10	10	10=(10/10)/10
80-60	12	20	6=(10/20)/12
$\sum$	66		

خطوات حساب المنوال: بما أن جدول غير منتظم ونظرا لتأثر القيمة المنوالية لطول الفئة في الجدول فانه يتم تعديل التكرار المطلق قبل حساب القيمة المنوالية:

1- تعديل التكرارات باستخدام الصيغة الإحصائية التالية: $hi = \dfrac{ni}{c_i / c_{\min}}$

كما يتضح في العمود الثالث من الجدول.

حيث hi هو رمز التكرار المعدل، ni هو التكرار المطلق، C_i هو طول الفئة المقابل لكل تكرار وC_{min} هو اصغر طول فئة في الجدول .

2- **تحديد الفئة المنوالة من عمود التكرارات المعدلة:** وهي في هذه الحالة الفئة المقابلة لأكبر تكرار معدل: المتمثلة في الفئة: 20-30 .

3- التعويض في الصيغة الاحصائية لحساب المنوال:

$$Mo = L1 + \frac{\Delta 1}{\Delta 1 + \Delta 2} * C$$

بالتعويض نجد :

$$M_o = 20 + \frac{13-9}{(13-9)+(13-11)} * 10$$

$$= 20 + \frac{4}{4+2} * 10 = 26.67$$

و يقرأ: أغلبية العينة المبحوثة يقدر سنها بـ 26.67 سنة.

6-1-خصائص المنوال :

1-لا يدخل في حساب قيمة المنوال كل مفردات التوزيع.

2-لا يتأثر المنوال بالقيم المتطرفة سواء كانت كبيرة او صغيرة .

3-لا يمثل المنوال دائما القيمة الوسطى في التوزيع .

4-يمكن حساب المنوال من جداول تكرارية مفتوحة .

5- يتساوى المنوال مع الوسيط والوسط الحسابي في حالة كان التوزيع طبيعي (منحنى متماثل).

6-يتاثر المنوال بأطوال الفئات ؛ مما يتطلب تعديل التكرارات قبل حسابه.

7-قد يشمل التوزيع اكثر من منوال واحدة فيكون التوزيع بمنوالين او متعدد المنوال. كما قد تنعدم قيمة المنوال في حالة كانت كل التكرارات متساوية .

8-يمكن حساب المنوال من معرفة قيمتي المتوسط الحسابي والوسيط كما يلي:

$$Mo = 3.Me - 2x$$

2- المتوسط الحسابي:(Moyenne arithmétique)

يعد المتوسط الحسابي من أكثر المقاييس الإحصائية انتشارا بين الناس لسهولته وفائدته .وهو في ابسط تعريفاته عبارة عن مجموع القياسات الخاصة بظاهرة معينة على عدد هذه القياسات. وتختلف طرق حسابه حسب البيانات .

1-2- حساب المتوسط الحسابي من البيانات غير المبوبة :

ويستعمل في حالة كانت المشاهدات الخاصة بالمتغير المدروس لها نفس المستوى من الأهمية . نفرض ان القياسات هي: $x_1, x_{2,} x_3....x_n$ حسب التعريف السابق فان صيغة حساب المتوسط الحسابي البسيط هي:

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

حيث: - $\bar{x}$ هو رمز المتوسط الحسابي.

- $\sum_{n}^{i=1} x_i$ هو مجموع المشاهدات.

-n هو عدد المشاهدات أو حجم العينة.

مثال: البيانات التالية خاصة بأعمار ستة أطفال: 6 3 5 6 12 10

بالتعويض في الصيغة السابقة فان المتوسط الحسابي لهذه المشاهدات هو :

$$\bar{x} = \frac{10+12+6+5+3+6}{6} = 7$$

و منه: متوسط عمر الأطفال هو 7 سنوات.

2-2- المتوسط الحسابي في حالة بيانات مبوبة:

1-2-2- الوسط الحسابي في حالة بيانات كمية متقطعة مبوبة: يستعان بهذه الطريقة في حالة ما كانت القياسات الخاصة بالمتغير الإحصائي تختلف عن بعضها من حيث الأهمية أو الوزن كما تستعمل في حالة البيانات المبوبة لكل المتغيرات الكمية المتقطعة .

باستعمال الصيغة التالية : $$\boxed{\bar{x} = \frac{\Sigma n_i * x_i}{\Sigma n_i}}$$

مثال: البيانات التالية تمثل عينة من السلع حسب الوزن:

ni*xi	ni	الاوزان xi
15	3	5
28	4	7
4	2	2
8	2	4
∑= 55	∑=11	المجموع

بالتعويض في الصيغة السابقة نتحصل على:

$$\bar{x} = \frac{55}{11} = 5$$

اذن متوسط وزن السلع هو 5كلغ.

2-2-2- حساب المتوسط الحسابي في حالة المتغير الكمي المتصل:

في حالة كان الجدول مبوب **بمتغير كمي متصل** يتم الاستعانة بمراكز الفئات لحسابه كما يتضح في المثال التالي:

مثال: الجدول التالي يوضح توزيع مجموعة سكانية حسب السن:

السن	ni	xi	ni*xi
15-10	2	12.5	25
20-15	8	17.5	140
25-20	6	22.5	135
30-25	12	27.5	330
35-30	27	32.5	877.5
40-35	16	37.5	600
45-40	14	42.5	595
50-45	8	47.5	380
55-50	5	52.5	262.5
60-55	2	57.5	115
$\sum$	100		3460

بتطبيق نفس الصيغة السابقة نتحصل على:

$$\boxed{\bar{x} = \frac{\sum n_i * x_i}{\sum n_i}} \qquad \bar{x} = \frac{3460}{100} = 34.6$$

. اذن متوسط السن بالنسبة للعينة المبحوثة يساوي 34.6 سنة

طرق أخرى لحساب المتوسط الحسابي: أوجدت الطرق التالية لحساب المتوسط الحسابي لغرض تقليص حجم القيم الكبيرة الناتجة عن حاصل جداء ni * xi ومن حاصل جمع كل جداء، لهذا توصل الاحصائيون الى الطرق التالية لأجل تحقيق هذا الغرض.

2-2-1-طريقة الوسط الفرضي:

في هذه الحالة نطبق الصيغة التالية:

$$\bar{x} = x_o + \frac{\sum n_i(x_i - x_o)}{\sum n_i}$$

حيث: - x_o هو الوسط الفرضي: **وهو مركز الفئة المقابل لأكبر تكرار**

-n_i هو التكرار

- x_i هو مركز الفئة

و منه فان x_o في الجدول التالي يتمثل في القيمة: 32.5 ويتم طرحه من كل قيمة من قيم مركز الفئة بغرض تقليص حجم القيم الكبيرة التي يمكن الحصول عليها من حاصل ضرب التكرار المطلق في مركز الفئة المقابل له، وهذه الطريقة باختصار تتمثل في استخدام خاصية من خصائص المتوسط الحسابي: **القائلة أنه اذا تم طرح قيمة ثابتة من كل المشاهدات ثم قمنا بحساب المتوسط الحسابي فان المتوسط الحسابي الأصلي يساوي المتوسط الحسابي الجديد زائد القيمة الثابتة المطروحة قبل حساب المتوسط الحسابي.**

باستعمال الجدول السابق نحصل على:

السن	n_i	x_i	x_i-x_o	$n_i(x_i$-$x_o)$
15-10	2	12.5	-20	-40
20-15	8	17.5	-15	-120
25-20	6	22.5	-10	-60
30-25	12	27.5	-5	-60
35-30	**27**	**32.5**	0	0
40-35	16	37.5	5	80
45-40	14	42.5	10	140
50-45	8	47.5	15	120
55-50	5	52.5	20	100
60-55	2	57.5	25	50
$\sum$	100			210

بالتعويض في الصيغة نحصل على نفس المتوسط الحسابي السابق:

$$\bar{x} = x_o + \frac{\sum n_i(x_i - x_o)}{\sum n_i}$$

$$\bar{x} = 32.5 + \frac{210}{100} = \boxed{34.6}$$

2-2-2-2- طريقة الانحرافات المختصرة:

من أجل تحقيق نفس الهدف المتمثل في اختصار حاصل القيم المحصل عليها من جداء n_i*x_i يتم استعمال هذه الطريقة التي تعتمد على كل من الوسط الفرضي وطول الفئة لتقليص حجم القيم المشاهدة وتسهيل الحساب باستعمل قيم صغيرة قدر الامكان حيث يتم طرح قيمة الوسط الفرضي من المشاهدات (مراكز الفئات في حالة جدول مبوب الى فئات) ثم قسمة الحاصل على طول الفئة

الذي يجب أن تكون قيمته ثابتة. يمكن توضيح ذلك باستعمال نفس قيم الجدول السابق كما يلي:

السن	n_i	x_i	x_i-x_o	$\frac{x_i - x_o}{c}$	$n_i(\frac{x_i - x_o}{c})$	Fcc
15-10	2	12.5	-20	-4	-8	2
20-15	8	17.5	-15	-3	-24	10
25-20	6	22.5	-10	-2	-12	16
30-25	12	27.5	-5	-1	-12	28
35-30	27	32.5	0	0	0	55
40-35	16	37.5	5	1	16	71
45-40	14	42.5	10	2	28	85
50-45	8	47.5	15	3	24	93
55-50	5	52.5	20	4	20	98
60-55	2	57.5	25	5	10	100
$\sum$	100				42	

بتطبيق الصيغة التالية يمكن الحصول على نفس النتيجة مع تحقيق فكرة اختصار القيم والمجموع الكلي لحاصل جداء التكرار في مركز الفئة:

$$\bar{x} = x_o + \left(\frac{\sum n_i * \left(\frac{x_i - x_o}{c_i} \right)}{\sum n_i} \right) * c$$

بالتعويض في هذه الصيغة نحصل على: $\bar{x} = 32.5 + \left(\frac{42}{100}\right) \times 5 = 34.6$

وهي نفس النتيجة السابقة.

3-2-2- خصائص المتوسط الحسابي:

1. المجموع الجبري لانحرافات مجموعة من القيم عن وسطها الحسابي يساوي صفرا.
2. إذا أضفنا أو طرحنا مقدارا ثابتا من قيمة كل مشاهدة في مجموعة من المشاهدات، فان الوسط الحسابي الجديد عبارة عن الوسط الحسابي للقيم الأصلية للمشاهدات (قبل الطرح أو الإضافة) مضافا إليه أو مطروحا منه نفس المقدار الثابت .
3. إذا ضربنا أو قسمنا كل مشاهدة في مجموعة من المشاهدات في مقدار ثابت فان الوسط الحسابي الجديد عبارة عن الوسط الحسابي للقيم الأصلية للمشاهدات (قبل الضرب أو قبل القسمة) مضروبا أو مقسوما في المقدار الثابت .
4. يتأثر الوسط الحسابي بالقيم المتطرفة سواء كانت كبيرة أو صغيرة في التوزيع، باعتباره يدخل في حسابه كل المفردات .
5. لا يمكن حساب المتوسط الحسابي للتوزيعات التكرارية المفتوحة كما لا يمكن ايجاد المتوسط الحسابي بالطرق البيانية .

3-الوسط::La Médian

الوسيط لمجموعة من الأرقام مرتبة حسب قيمها (في ترتيب تصاعدي أو تنازلي) هي القيمة التي تتوسط البيانات، أي التي تقع في المنتصف أو الوسط الحسابي للقيمتان اللتان تتوسطا البيانات أو تقع في منتصف البيانات، أي تقسمها البيانات إلى قسمين متساويين كما يضح في الشكل التالي حيث أن 50%

من المشاهدات تقل قيمتها عن الوسيط و50% منها تزيد قيمته عن قيمة الوسيط.

|--------------|---------------|

x_i أكبر %50 M_e أقل %50 x_n

1-3- الوسيط في حالة بيانات غير مبوبة :

** في حالة كان مجموع الوحدات الإحصائية المبحوثة فرديا: يكون الوسيط عبارة القيمة ذات الترتيب $R_{me} = (N+1)/2$، بعد ترتيب القيم تصاعديا أو تنازليا.

مثال 1 = إذا كان حجم العينة رقم فردي كما في الحالة التالية:

8 -10 -4 -8 -4 -3 -6 -5 -8

يجب اعادة ترتيب المشاهدات ترتيب تصاعدي او تنازلي وتحديد رتبة كل مشاهدة كما يتضح من الجدول التالي:

المشاهدات x	10	8	8	8	6	5	4	4	3
الرتبة R	1	2	3	4	5	6	7	8	9

من الصيغة السابقة نحسب رتبة الوسيط: $R_{me}=(9+1)/2 = 5$ ومنه قيمة الوسيط هي القيمة ذات المرتبة الخامسة أي $M_e=6$ كما يتضح من الجدول.

** **في حالة حجم العينة عدد زوجي:** يقع الوسيط في هذه الحالة بين رتبتين وقيمته تساوي متوسط القيمتين المقابلتين للرتبة الأول R_{me1} والرتبة الثانية R_{me2} كما توضحه الصيغة التالية:

$$M_e=(X_{Rme1} + X_{Rme2})/2$$

حيث: R_{me1} هي رتبة الوسيط الأولى وتساوي: $R_{me1}=N/2$

بينما: R_{me2} هي رتبة الوسيط الثانية وتساوي: $R_{me2}=(N/2)+1$

أما: X_{Rme1} فهي المشاهدة المقابلة لرتبة الوسيط الأولى.

و: X_{Rme2} هي المشاهدة المقابلة لرتبة الوسيط الثانية.

مثال: المشاهدات التالية تمثل اوزان عينة من الاطفال:

7 -12 -15 - 11 – 5 -18 - 9 - 5

لإيجاد الوسيط يجب اعادة ترتيب المشاهدات وتحديد رتبها كما يلي:

5	5	7	9	11	12	15	18	المشاهدة x
8	7	6	5	4	3	2	1	الرتبة R

المرتبة الأولى $R_{me1} = N/2=8/2=4$

المرتبة الثانية $R_{me2}=(N/2)+1=(8/2)+1=5$

و منه الوسيط يساوي متوسط القيمتين المقابلتين لكل من الرتبة الرابعة وهي 11 والرتبة الخامسة وهي 9

أي: $Me =(11+9)/2=10$

و منه فان 50% من الاطفال يقل وزنهم عن 10كلغ و50% من الأطفال يزيد وزنهم عن 10 كلغ.

2-3- حساب الوسيط في حالة بيانات مبوبة بمتغير كمي متصل:

في هذه الحالة نطبق الصيغة التالية: $$M_e = l_1 + \frac{\frac{\Sigma n_i}{2} - F_{M_{e-1}}}{n_{Me}} \times c$$

حيث: - M_e هي رمز الوسيط.

- l_1 هو الحد الأدنى من الفئة الوسيطية التي يتم تحديدها من رتبة الوسيط والتكرار المتجمع الصاعد.

- $\frac{\Sigma n_i}{2}$ تمثل رتبة الوسيط.

- F_{Me} هو التكرار المتجمع الصاعد السابق لفئة الوسيط.

- n_{Me} هو التكرار المطلق لفئة الوسيط .

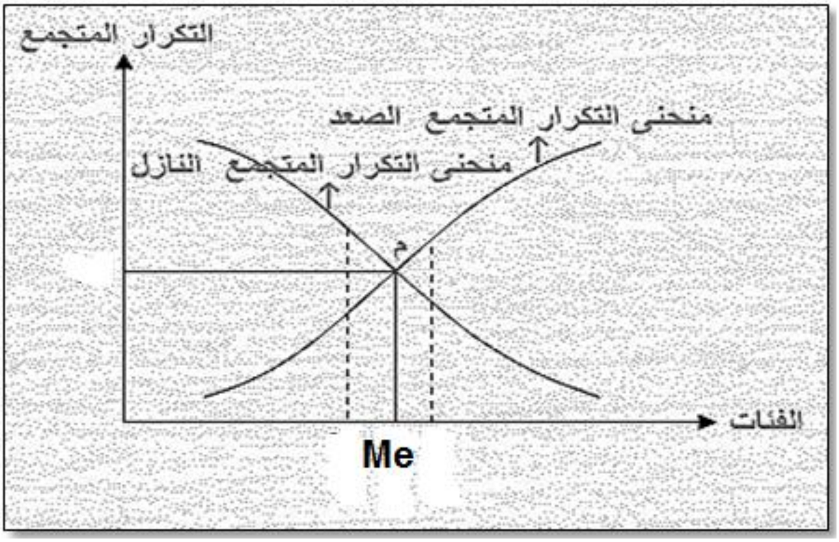
بما أن رتبة الوسيط تساوي 50 أنطلاقا من قسمة مجموع التكرارات على 2 بالنسبة للجدول السابق.، فان الفئة الوسيط هي الفئة 30-35 وهي الفئة المقابلة للتكرار المتجمع الصاعد المقابل لفئة الوسيط، وانطلاقا منها يمكن تعوض كل القيم في الصيغة كالتالي:

$$M_e = 30 + \frac{\frac{100}{2} - 28}{27} \times 5 = 34.07$$

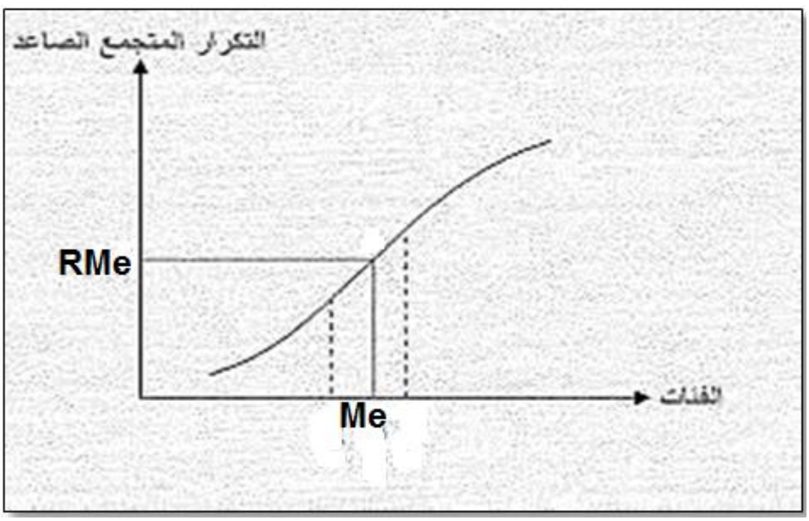
ومنه يمكن القول ان ال50% من مجموع المشاهدات في الجدول تساوي أو تزيد عن القيمة 34.07، و50% من مجموع القيم تساوي أو تقل عن القيمة 34.07. بمعنى هي القيمة التي تقسم المشاهدات الى قسمين متساويين.

3-3- استخراج الوسيط بيانيا: ويمكن ذلك بثلاث طرق مختلفة:

أ- طريقة تقاطع المنحنى التكراري المتجمع الصاعد والمنحنى التكراري المتجمع النازل: حيث يتم اسقاط نقطة تقاطعهما على محور السينات ليتم الحصول بذلك على قيمة تقريبية وآلية للوسيط كما يوضح الشكل النظري التالي:



ب- الاعتماد على المنحنى التكراري المتجمع الصاعد أو النازل بالإضافة الى رتبة الوسيط R_{Me} لاستخراج قيمة هذا الاخير بيانيا، كما يتضح من خلال الشكل النظري التالي :



4-العلاقة التجريبية بين المتوسطات الثلاثة

$$\boxed{\bar{x} - M_o = 3(\bar{x} - M_e)}$$

يصلح استخدام هذه العلاقة خاصة في حالة التوزيعات المتناظرة بحيث تقدم نتائج دقيقة، اذ تسمح باستنتاج قيمة واحد من متوسطات النزعة المركزية الثلاثة من خلال معرفتنا لقيمة المتوسطين الآخرين كالتالي:

1- $\bar{x} = \dfrac{3M_e - M_o}{2}$ لإيجاد قيمة المتوسط الحسابي.

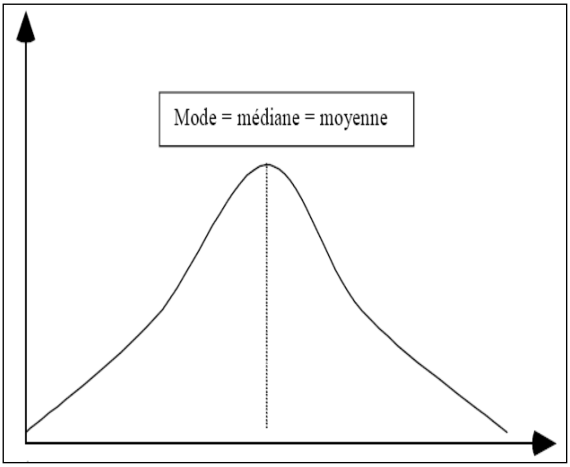
2- $M_e = \dfrac{M_o + 2\bar{x}}{3}$ لإيجاد قيمة الوسيط.

3- $M_o = 3M_e - 2\bar{x}$ لإيجاد قيمة المنوال.

5- تحديد شكل التوزيع من خلال موضع المتوسطات الثلاث

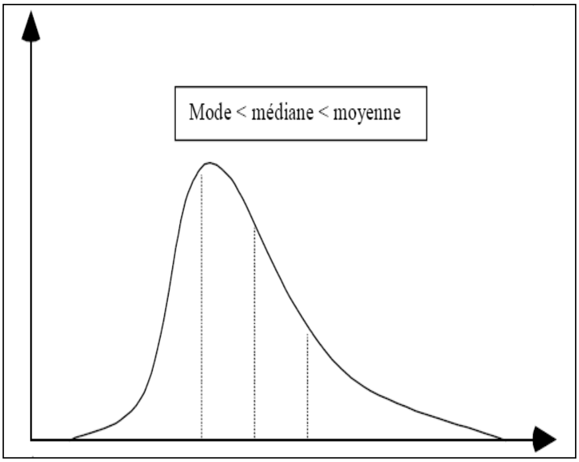
توضح الأشكال (2) و(3) أدناه الموضع النسبي للوسط والوسيط والمنوال للمنحنيات التكرارية الملتوية إلى اليمين وإلى اليسار، أما كل المنحنيات المتماثلة فيتطابق الوسط والوسيط والمنوال (الشكل1)

الشكل الأول:



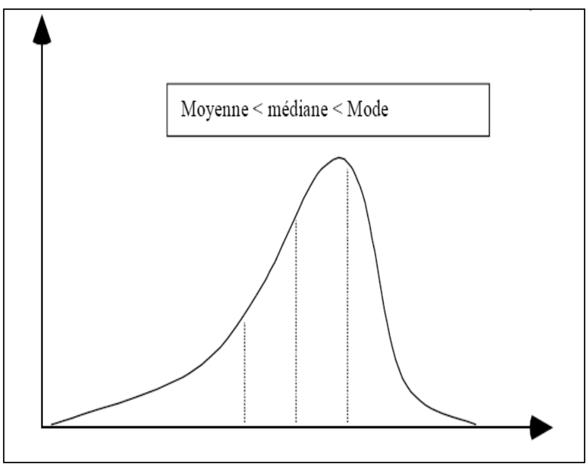
توزيع متماثل (لا يوجد التواء)

الشكل الثاني:



يكون المنوال على يسار الوسيط وبالتالي فالتشتت قوي من ناحية اليمين وهو موجب الالتواء

الشكل الثالث:



يكون المنوال على يمين الوسيط وبالتالي فالتشتت قوي من ناحية اليسار وهو سالب الالتواء.

6-مقاييس المكانة (الموضع) Les caractéristiques de position

6-1-الربيعات: les quartiles

- إن **الربيعات هي القيم** التي تقسم أي مجموعة من المشاهدات (سلسلة إحصائية) إلى أربع أقسام فرعية متساوية حيث أن كل قسم يساوي 25% من مجموع المشاهدات، كما يتضح من المخطط التالي.

%25 %25 %25 %25

|---------|---------|---------|--------|

Q_1 Q_2 Q_3

- **الربيع الأول (Q_1)** يقسم سلسلة المشاهدات إلى قسمين بحيث يكون 25% من المشاهدات أقل من قيمة Q_1 و75% من مجموعها يكون اكبر من قيمة Q_1 كما يوضح المخطط التالي.

- **الربيع الثاني(Q_2)** هو نفسه الوسيط الذي يقسم السلسلة الإحصائية إلى قسمين متساويين 50% أقل من قيمة (Q_2) و50% أكبر من قيمة (Q_2).

|--------------------------|----------|

70% Q_3 25%

- **الربيع الثالث(Q_3)** يقسم سلسلة المشاهدات الى قسمين بحيث يكون 75% من المشاهدات أقل من قيمة Q_3 و25% من مجموعها يكون اكبر من قيمة Q_3.

|-----------|-----------------------|

75% 25% Q_1

من اجل حساب واحد من القيم السابقة لابد من حساب رتب هذه القيم حتى يكون في الإمكان تحديد الفئة التي تتضمنها كالتالي:

$$R_{Q3} = \frac{3\Sigma n_i}{4} \text{ ،} R_{Q2} = \frac{2\Sigma n_i}{4} \text{ ،} R_{Q1} = \frac{1\Sigma n_i}{4}$$

ثم يمكن تطبيق الصيغة التالية لحساب أي منها:

$$\boxed{Q_x = l_1 + \frac{\frac{x\Sigma n_i}{4} - F_{Qx-1}}{n_{Qx}} \times c}$$

حيث: - Q_x تمثل الربيع المراد حسابه؛ فقد يكون الربيع الأول $\mathbf{Q_1}$، أو الربيع الثاني $\mathbf{Q_2}$، أو الربيع الثالث $\mathbf{Q_3}$.

- l_1 تمثل الحد الأدنى لفئة الربيع المراد حسابه.

- $\frac{x\Sigma n_i}{4}$ تمثل رتبة الربيع المراد حسابه حيث x تختلف من 1 الى 3 حسب رقم الربيع.

- F_{Qx-1} تمثل التكرار المتجمع الصاعد السابق لفئة الربيع المراد حسابه.

- n_{Qx} تمثل التكرار المطلق المقابل لفئة الربيع المراد حسابه دائما.

- c تمثل طول الفئة الخاصة بالربيع المراد حسابه.

- أما x فتمثل رقم الربيع الذي نريد حسابه.

مثال: الجدول التالي يمثل مجموعة من البيانات والمراد ايجاد قيم الربيعات الأول، الثاني والثالث.

الفئات	n_i	Fcc
25-30	6	6
35-30	19	25
40-35	15	40
45-40	7	47
50-45	3	50
	50	

أول خطوة هي حساب قيمة التكرار المتجمع الصاعد Fccكما يتضح في الجدول، ثاني خطوة إيجاد رتبة كل ربيع ومنه تطبيق الصيغة الإحصائية السابقة لإيجاد القيم.

- رتبة الربيع الأول: تحسب وفق الصيغة التالية $R_{Q1} = \frac{50}{4} = 14.5$ ومنه فان فئة الربيع الأول هي الفئة 30-35 باعتبار القيمة 14.5 متضمنة في التكرار المتجمع الصاعد المقابل لهذه الفئة .
- بتطبيق الصيغة السابقة نحصل على : $Q_1 = 30 + \frac{12.5 - 6}{19} \times 5$

$$= 30 + 1.71 = 31.71$$

ومنه فان قيمة الربيع الأول تساوي 31.71، بمعنى أن 25% من مجموع المشاهدات في هذا الجدول تساوي أو تقل عن 31.71 و75% من مجموع المشاهدات تساوي أو تزيد عن القيمة 31.71.

- بنفس الخطوات يمكن الحصول على كل من الربيع الثاني والثالث:

فرتبة الربيع الثالث هي: $R_{Q3} = \frac{3 \times 50}{4} = 37.5$

- و انطلاقا من التكرار المتجمع الصاعد فان فئة الربيع الثالث هي: 35-40، بالتعويض في الصيغة السابقة :

$$Q_3 = 35 + \frac{37.5 - 25}{15} \times 5$$
$$= 35 + 4.17 = 39.17$$

ومنه فان قيمة الربيع الثالث تساوي 39.17، بمعنى أن 75% من مجموع المشاهدات تساوي أو تقل عن 39.17 و25% من مجموع المشاهدات تساوي أو تقل من 39.17.

2-6-العشيرات les déciles: هي قيم المتغير (x_i) التي تقسم السلسلة الاحصائية الى 10 اقسام فرعية متساوية حيث يساوي كل قسم منها 10% بذلك يمكن حساب قيم تسع عشيرات. كما يتضح من المخطط التالي :

10% 10% 10% 10% 10% 10% 10% 10% 10% 10%

D_6 D_7 D_8 D_9 D_5 D_3 D_4 |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

D_2 D_1

يتم حساب العشيرات بحساب أولا رتبة العشير المراد حسابه بالطريقة التالية:

$$R_{D_1} = \frac{1 \times \Sigma n_i}{10}$$

$$R_{D2} = \frac{2 \times \Sigma n_i}{10}$$

.

.

$$R_{D9} = \frac{9 \times \Sigma n_i}{10}$$

بعد إيجاد الرتبة تتبع نفس خطوات إيجاد الربيعات من أجل إيجاد العشيرات بتطبيق الصيغة التالية مع تغيير رقم العشير المراد حسابه في كل مرة:

$$\boxed{D_x = l_1 + \frac{\dfrac{x\Sigma n_i}{4} - F_{Dx-1}}{n_{Dx}} \times c}$$

مثال: حساب العشير الثاني من الجدول السابق: نجد أن رتبته تساوي:

$$R_{D2} = \frac{2 \times 50}{10} = 10$$

و من التكرار المتجمع الصاعد نجد أن فئة هذا العشير هي: 30-35

بتطبيق الصيغة السابقة نجد :

$$D_2 = 30 + \frac{10-6}{19} \times 5$$
$$= 30 + 1.05 = 31.05$$

و منه نقول أن 20% من مجموع البيانات تقل 31.05 و80% من مجموع البيانات تساوي أو تزيد عن 31.05.

بنفس الطريقة نحسب باقي الشعيرات.

3-6-المئينيات les centiles:

وهي قيم المتغير الاحصائي التي تقسم السلسلة إلى مئة قسم فرعي متساوي، ليكون عددها بذلك يساوي 99 مئين.

$P_1, P_2, P_3, \ldots P_{99}$

من اجل حساب المئين يتم أولا إيجاد رتبة المئين المراد حسابه بالطريقة التالية :

$$\boxed{\begin{array}{l} R_{P_1} = \dfrac{1 \times \Sigma n_i}{100} \\ R_{P2} = \dfrac{2 \times \Sigma n_i}{100} \\ . \\ . \\ R_{P99} = \dfrac{9 \times \Sigma n_i}{100} \end{array}}$$

ثم يتم تطبيق الصيغة التالية من أجل الحصول على المئين المراد معرفته مع الاستعانة بالتكرار المتجمع الصاعد لتحديد فئته:

$$\boxed{P_x = l_1 + \frac{\frac{x\Sigma n_i}{4} - F_{Px-1}}{n_{Dx}} \times c}$$

مثال: حساب المئين الخمسون: P_{50}

أولا: نحسب رتبته وهي بالنسبة للجدول السابق(50*50)/100=25، ومن التكرار المتجمع الصاعد نجد ان فئته هي الفئة 30-35، بتطبيق الصيغة الإحصائية الاخيرة نجد ان قيمة هذا المئين تساوي:

$$p_{50} = 30 + \frac{\frac{50*50}{100} - 6}{19} \times 5 = 31$$

و منه فان 50% من المشاهدات تقل عن 31 و50% من المشاهدات تساوي أو تزيد عن القيمة 31.

الفصل السادس: مقاييس التشتت

تمهيد

1- المدى

2-المدى الربيعي

3-الانحراف الربيعي

4- الانحراف المتوسط

5-التباين والانحراف المعياري

5-1- الانحراف المعياي في حالة بيانات غير مبوبة

6-مقاييس التشتت النسبي

6-1- معامل الاختلاف

6-1-1- معامل الاختلاف النسبي

6-1-2-معامل الاختلاف الربيعي

6-2- الدرجة المعيارية

تمهيد

إذا كانت مجموع من البيانات متباعدة أو متباينة عن بعضها يقال انها مشتتة أما اذا كانت البيانات متجانسة وغير متباعدة فيقال أنها غير مشتتة. (طبية، 2008، ص75)

و بالتالي لا تعتبر مقاييس النزعة مركزية كافية لوصف مجموعة من البيانات وصفاً كاملاً وخصوصا المقارنة فيما بينها، فقد تتساوى بعض العينات في الوسط الحسابي بالرغم من اختلاف توزيع بياناتها حول مركزها (درجة تجانس البيانات). فالعينات في المثال التالي ذات وسط حسابي واحد يساوي -8- ولكنها بلا شك تختلف عن بعضها اختلافا واضحا حيث أن العينة الأولى يوجد بينها تشابه تام في القيم، بينما العينة الثانية يلاحظ فروق كبيرة بين قيمها الصغرى -3- والكبرى -16-.

جدول رقم 14 :

عينة 1	8	8	8	8	8
عينة 2	4	3	6	16	11

فالوسط الحسابي يمثل مركز البيانات لكنه لا يبين مدى التفاف أو بعثرة البيانات حول هذا وسط، ولهذا لا بد من وجود مقياس آخر مع مقاييس النزعة المركزية لقياس درجة التجانس أو التشتت في داخل هذه البيانات.

إن الدرجة التي تتجه بها البيانات الرقمية للانتشار حول قيمة وسطى – المتوسط الحسابي مثلا - تسمى تشتت أو توزيع البيانات. ومن أهم مقاييس التشتت المدى والتباين والانحراف المعياري والانحراف المتوسط .

1- المدى

المدى هو الفرق بين أكبر قيمة وأصغر قيمة.

1-1- حساب المدى من البيانات غير مبوبة :

المدى = أكبر قيمة – أصغر قيمة

مثال1: احسب المدى للبيانات التالية :

80 – 350 – 100 – 150 – 90 – 110 – 300 – 250 – 200 – 95

الحل :

نرتب القيم أولاً: (80-90-95-100-110-150-200-250-300-350)

المدى = 350 – 80 = 270

2-1- حساب المدى من البيانات المبوبة :

المدى = الحد الأعلى للفئة الأخيرة – الحد الأدنى للفئة الأولى.

مثال2: احسب المدى للجدول التالي :

جدول رقم 15 :

الفئات	50-40	60-50	70-60	80-70
n_i	6	8	12	9
x_i	45 **min**	55	65	75 **max**

المدى في هذه الحالة هو الفرق بين مركز الفئة الأخير (الأكبر) ومركز الفئة الأول (الأصغر) كما يتضح في الصيغة التالية:

$$\boxed{\mathbf{E} = x_{i\,max} - x_{i\,min}}$$

حيث أن :

- $x_{i\,max}$ هو مركز الفئة الأكبر.
- $x_{i\,min}$ هو مركز الفئة الأصغر.

و منه يمكن حساب المدى في الجدول السابق كالتالي:

$$E = 75 - 45 = 30$$

و بالتالي المدى في العينة المبحوثة هو 30 أي يقدر طول مجال المشاهدات بـ 30 سنة.

مزايا المدى:

- و بسيط وسهل الحساب.
- يستخدم في مراقبة الجودة.
- يكثر استخدامه في مجال المناخ والجو مثل معرفة المدى الحراري والرطوبة والضغط الجوي...

عيوب المدى:

- يعتمد في حسابه على قيمتين فقط في التوزيع.
- يتأثر بالقيم المتطرفة.

2-المدى الربيعي

للتخلص من القيم المتطرفة الشاذة التي يتأثر بها المدى المطلق لجأ الاحصائيون الى ما يعرف بالمدى الربيعي الذي يقيس التشتت داخل مجال أضيق من مجال المدى المطلق، حيث يلغي 25% من القيم المتطرفة الصغرى و25% من القيم المتطرفة الكبرى ليقيس قيمة المدى داخل 50% من قيم التوزيع الوسطى وذلك بتطبيق الصيغة التالية:

صيغة رقم (02)

$$EQ = Q_3 - Q_1$$

و يتضح من الصيغة أنها تقوم على ايجاد الفرق بين الربيع الثالث والربيع الاول، بعد حساب كل منها، وهو يشير الى درجة التشتت الموجود داخل 50% من القيم الوسطى للتوزيع، ذلك لأن طريقة حسابه تلغي 25% من القيم المتطرفة الصغرى والكبرى،و يفيد أكثر في المقارنة بين مجموعتين.

مثال3: جدول رقم 03 :

المقاييس	المجموعة الأولى	المجموعة الثانية
الربيع الاول Q_1	25	130
الربيع ثالث Q_3	35	155

$$EQ = Q_3 - Q_1$$

من الصيغة نجد أن المدى الربيعي الاول يساوي :EQ_1 =35-25= 10

و المدى الربيعي الثاني يساوي: EQ_1 =155-130= 25

ومنه فان درجة التشتت في المجموع الثانية اقوى من درجة التشتت في المجموعة الاولى.

3-الانحراف الربيعي

هو متوسط المدى الربيعي، دائما من أجل تحقيق الغرض المتمثل في التخلص من القيم المتطرفة، يقوم الانحراف الربيعي على ايجاد قيمة التشتت داخل مجال يقدر بـ 25% من القيم التي تتموضع وسط التوزيع، ويحسب بالصيغة التالية:

صيغة رقم (03)

$$DQ = \frac{Q_3 - Q_1}{2}$$

مثال$_4$: من الجدول رقم 15:

نجد أن الانحراف الربيعي في المجموع الاولى من الصيغة رقم .. يساوي :

$DQ_1=(35-25)/2=5$

$DQ=(155-130)/2=12.5$

يتم الوصول الى نفس النتيجة السابقة أي أن التشتت في المجموع الاولى أقل من التشتت في المجموعة الثانية.

4- الانحراف المتوسط

يعرف الانحراف المتوسط بانه متوسط الانحرافات عن وسطها الحسابي $\overline{x}$ ويرمز له بالرمز « MD ». والسبب في أخذ القيم المطلقة للانحرافات هو أن مجموع انحرافات القيم عن وسطها الحسابي يساوي الصفر.(Py,2001).

3-1- الانحراف المتوسط في حالة بيانات غير مبوبة:

يحسب في هذه الحالة بالصيغة الاحصائية التالية:

الصيغة رقم (04)

$$\boldsymbol{MD = \frac{\sum|x_i - \overline{x}|}{n}}$$

مثال$_5$:

اذا كانت الطاقة التصديرية لخمس محطات لحلية المياه بالمليون متر مكعب كما يلي:

7، 10، 2، 5، 4.

فان حساب الانحراف المتوسط يستوجب:

أولا: ايجاد قيمة المتوسط الحسابي:

$$\boldsymbol{\overline{x} = \frac{\sum x_i}{n}}$$

$$\boldsymbol{= \frac{7+5+2+10+7}{5} = \frac{28}{5} = 5.6}$$

ثانيا :تطبيق الصيغة السابقة من خلال خطوات الجدول التالي:

جدول رقم 16 :

الطاقة التصديرية	حساب الانحرافات	حساب الانحرافات المطلقة
X_i	$x_i - \overline{x}$	$\lvert x_i - \overline{x} \rvert$
4 **5** **2** **10** **7**	1.6-=5.6-4 0.6- 3.6- 4.4 1.4	1.6 0.6 3.6 4.4 1.4
$\sum$	0	11.6

ثالثا: التعويض في الصيغة

$$MD = \frac{\sum \lvert x_i - \overline{x} \rvert}{n}$$

$$= \frac{11.6}{5} = 2.32$$

و منه فان متوسط الفروق بين القيم السابقة يقدر بـ 2.32 مليون م^3 وهو فرق هام يشير الى وجود اختلاف واضح بين الشركات المبحوثة فيما يخص طاقتها التصديرية.

الانحراف المتوسط في حالة بيانات مبوبة:

2-4-1- في حالة متغيركمي منفصل :

يعطى الانحراف المتوسط في هذه الحالة من خلال العلاقة التالية :

صيغة رقم (05)

$$\boxed{\mathbf{M.D} = \frac{\sum \lvert x_i - \overline{x} \rvert * n_i}{\sum n_i}}$$

حيث:

ni: هو التكرار

x_i: هو مركز الفئة

$\overline{x}$: هو المتوسط الحسابي

$\sum n_i$ هو مجموع التكرارات في الجدول

مثال6: الجدول التالي يمثل الانفاق الشهري لعينة من الأسر:

جدول رقم 05:

الانفاق (1000)		الخطوة 1	الخطوة 2	الخطوة 3	الخطوة 4	الخطوة 5
	n_i	x_i	$n_i * x_i$	$\overline{x} = \frac{\sum x_i * n_i}{\sum n_i}$	$\lvert x_i - \overline{x}\rvert$	$n_i\lvert x_i - \overline{x}\rvert$
2-5	1	3.5	3.5	***حساب المتوسط الحسابي***	7.2	7.2
5-8	8	6.5	52	$\overline{x}=\frac{428}{40}=10.7$	4.2	33.6
8-11	13	9.5	123.5		1.2	15.6
11-14	10	12.5	125		1.8	18
14-17	8	15.5	124		4.8	38.4
$\sum$	40		428			112.8

الخطوة 6: التعويض في الصيغة

$$M.D = \frac{112.8}{40} = 2.82$$

و منه فان: متوسط انحراف القيم عن وسطها الحسابي يقدر بـ 2.8 أي 2800 دج

5-التباين والانحراف المعياري

لما كان من الصعب التعامل رياضيا (تحليليا) مع الانحراف المتوسط بسبب القيم السالبة التي تنتج عن طريقة حسابه، دعت الحاجة الى استخدام مقياس آخر بنفس قوة الانحراف المتوسط، لكي يكون من السهل التعامل معه تحليليا، وبما أن الفكرة هي التخلص من الاشارات للانحرافات فان تربيع الانحرافات يخلصنا من الاشارة، ولهذا فان الانحراف المعياري يعرف عن طريق التباين والذي يعرّف على انه متوسط مربع انحراف القيم عن وسطها الحسابي ويرمز له بالرمز "σ^2 "، والجذر التربيعي للتباين ينتج عنه مقياس من أهم وأدق مقاييس التشتت وهو ما يسمى بالانحراف المعياري الذي يرمز له بالرمز"σ" (موسى محمد،2007).

1-5-التباين والانحراف المعياري في حالة بيانات غير مبوبة:

تختلف نوعا ما صيغة حسابه في حالة كان الباحث يعمل على عينة أو على مجتمع :

التباين في حالة بيانات غير مبوبة :

في حالة عينة	في حالة مجتمع
$$\sigma^2 = \frac{\sum(x_i - \overline{x})^2}{n-1}$$ **صيغة رقم (06)** أو $$\sigma^2 = \frac{\sum {x_i}^2}{n-1} - \overline{x}^2$$ **صيغة رقم (07)**	$$\sigma^2 = \frac{\sum(x_i - \overline{x})^2}{n}$$ **صيغة رقم (08)** أو $$\sigma^2 = \frac{\sum {x_i}^2}{n} - \overline{x}^2$$ **صيغة رقم (09)**

التباين في حالة بينات مبوبة :

	في حالة عينة		في حالة مجتمع
-1	$\sigma^2 = \frac{\sum(n_i(x_i - \overline{x})^2)}{n-1}$ صيغة رقم (10)	-1	$\sigma^2 = \frac{\sum(n_i * (x_i - \overline{x})^2)}{\sum n_i}$ صيغة رقم (12)
	أو		أو
-2	$\sigma^2 = \frac{\sum\left(n_{i*}({x_i}^2)\right)}{n-1} - \overline{x}^2$ صيغة رقم (11)	-2	$\sigma^2 = \frac{\sum\left(n_{i*}({x_i}^2)\right)}{N} - \overline{x}^2$ صيغة رقم (13)

-1- الانحراف المعياي في حالة بيانات غير مبوبة :

الانحراف المعياري: هو الجذر التربيعي لانحراف القيم عن وسطها الحسابي.

	في حالة عينة		في حالة مجتمع
-1	$\sigma = \sqrt{\frac{\sum(x_i - \overline{x})^2}{n-1}}$ صيغة رقم (14)	-1	$\sigma = \sqrt{\frac{\sum(x_i - \overline{x})^2}{n}}$ صيغة رقم (15)
-2	أو $\sigma = \sqrt{\frac{\sum {x_i}^2}{n-1} - \overline{x}^2}$ صيغة رقم (15)	-2	أو $\sigma = \sqrt{\frac{\sum {x_i}^2}{n} - \overline{x}^2}$ صيغة رقم (16)

مثال7:

لدينا المشاهدات التالية خاصة بأوزان عينة من الأطفال حيث حجم العينة يساوي 8 :

9 7 6 5 4 4 3 2

1- يجب حساب قيمة المتوسط الحسابي :

$$\overline{x} = \frac{2+3+4+4+5+6+7+9}{8} = 5$$

$$\sigma^2 = \frac{1}{n}\sum(x_i - \overline{x})^2$$

$$= \frac{\sum(x_i - \overline{x})^2}{n}$$

$$= \frac{1}{n}[(2-5)^2 + (3-5)^2 + (4-5)^2 + (4-5)^2 + (5-5)^2 + (6-5)^2 + (7-5)^2]$$

$$= \frac{1}{8}[9+4+1+1+0+1+4+16]$$

$$= \frac{36}{8}$$

$$= 4.5$$

يمكن أيضا استخدام الصيغة الثانية لحساب الانحراف المعياري من خلال الخطوات التالية :

$$\sigma^2 = \frac{1}{n}\sum x_i^2 - \overline{x}^2$$

$$= \frac{1}{8}(2^2 + 3^2 + 4^2 + 4^2 + 5^2 + 6^2 + 7^2 + 9^2) - 5^2$$

$$= \frac{1}{8}(4+9+16+16+25+36+49+81) - 25$$

$$= \frac{236}{8} - 25$$

$$= 29.5 - 25 = 4.5$$

5-2- الانحراف المعياري في حالة بيانات مبوبة بمتغير كمي منفصل::

	في حالة عينة		في حالة مجتمع
-5	$\sigma = \sqrt{\frac{\sum(n_i * (x_i - \overline{x})^2)}{n-1}}$ صيغة رقم (17)	-7	$\sigma = \sqrt{\frac{\sum(n_i * (x_i - \overline{x})^2)}{n}}$ صيغة رقم (19)
	أو		أو
-6	$\sigma = \sqrt{\frac{\sum(n_i * (x_i)^2)}{n-1} - \overline{x}^2}$ صيغة رقم (18)	-8	$\sigma = \sqrt{\frac{\sum\left(n_{i*}(x_i{}^2)\right)}{n} - \overline{x}^2}$ صيغة رقم (20)

حيث :

n_iهي التكرار، x_i هي مراكز الفئات، $\overline{x}$ هو المتوسط الحسابي، n أو N هو حجم العينة

مثال8: البيانات التالية تمثل توزيع عينة من الأسر حسب عدد الاطفال:

جدول رقم 17 :

الطريقة الاولى

الطريقة الثانية

عدد الاطفال	ni	ni*xi	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$	$n_i(x_i - \overline{x})^2$	X_i^2	$(ni*x_i)^2$
1	10	10	-2,69	7,2361	72,361	1	10
2	15	30	-1,69	2,8561	42,8415	4	60
3	20	60	-0,69	0,4761	9,522	9	180
4	25	100	0,31	0,0961	2,4025	16	400
5	18	90	1,31	1,7161	30,8898	25	450
6	15	90	2,31	5,3361	80,0415	36	540
المجموع	103	380			238,06		1640

خطوات الحساب :

1-حساب المتوسط الحسابي:

$$\overline{\boldsymbol{x}} = \frac{\sum \boldsymbol{n_i} * \boldsymbol{x_i}}{\sum \boldsymbol{n_i}}$$

$$= \frac{380}{103} = 3.69$$

2-ايجاد التباين بتطبيق الصيغة الاولى :

$$\overline{\boldsymbol{x}} = \frac{\sum \boldsymbol{n_i} * \boldsymbol{x_i}}{\sum \boldsymbol{n_i}}$$

$$= \frac{1496}{40} = 37.4$$

$$\boldsymbol{\sigma^2} = \frac{\sum(\boldsymbol{n_i} * (\boldsymbol{x_i} - \overline{\boldsymbol{x}})^2)}{\boldsymbol{n}}$$

$$=\frac{238.06}{103} = \mathbf{2.31}$$

3-ايجاد قيمة الانحراف المعياري الذي هو الجذر التربيعي للتباين:

$$\boldsymbol{\sigma} = \sqrt{\boldsymbol{\sigma^2}} = \sqrt{\mathbf{2.31}} = \mathbf{1.51}$$

و منه فان قيمة الانحراف المعياري تقدر بـ $\boldsymbol{\sigma} = \mathbf{1.51}$ وهو تشتت ضعيف.

ايجاد التباين بتطبيق الصيغة الثانية :

$$\boldsymbol{\sigma} = \sqrt{\frac{\sum\left(\boldsymbol{n_{i*}}(\boldsymbol{x_i}^2)\right)}{\boldsymbol{n}} - \overline{\boldsymbol{x}}^2}$$

$$= \frac{1640}{103} - 3.69^2$$

$$= 15.92 - 13.62$$

$$= \mathbf{2.3}$$

4- التباين بتطبيق الصيغة الاخيرة يساوي 2.3 وهي نفس النتيجة الصيغة السابقة. وبالتالي فان قيمة **الانحراف المعياري** ستكون نفسها:

$$\boldsymbol{\sigma} = \sqrt{\boldsymbol{\sigma}^2} = \sqrt{\mathbf{2.31}} = \mathbf{1.51}$$

مثال9: البيانات التالية تمثل أوزان عينة من التلاميذ:

الطريقة الاولى   الطريقة الثانية

جدول رقم 07 :

الاوزان	ni	xi	ni*xi	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$	$n_i(x_i - \overline{x})^2$	x_i^2	$n_i * x_i^2$
34-32	4	33	132	-4,4	19,36	77,44	1089	4356
36-34	7	35	245	-2,4	5,76	40,32	1225	8575
38-36	13	37	481	-0,4	0,16	2,08	1369	17797
40-38	10	39	390	1,6	2,56	25,6	1521	15210
42-40	5	41	205	3,6	12,96	64,8	1681	8405
44-42	1	43	43	5,6	31,36	31,36	1849	1849
المجموع	40		1496			**241,6**		**56192**

خطوات الحساب :

1- **حساب المتوسط الحسابي :**

$$\overline{\boldsymbol{x}} = \frac{\sum \boldsymbol{n_i} * \boldsymbol{x_i}}{\sum \boldsymbol{n_i}}$$

$$= \frac{1496}{40} = 37.4$$

2- **ايجاد التباين بتطبيق الصيغة الاولى :**

$$\boldsymbol{\sigma}^2 = \frac{\sum(\boldsymbol{n_i} * (\boldsymbol{x_i} - \overline{\boldsymbol{x}})^2)}{\boldsymbol{n}}$$

$$=\frac{241.6}{40} = 6.04$$

ايجاد قيمة الانحراف المعياري الذي هو الجذر التربيعي للتباين:

$$\boldsymbol{\sigma = \sqrt{\sigma^2} = \sqrt{6.04} = 2.46}$$

و منه فان قيمة الانحراف المعياري تقدر بـ $\boldsymbol{\sigma = 1.51}$ وهو تشتت ضعيف.

3- ايجاد التباين بتطبيق الصيغة الثانية :

$$\boldsymbol{\sigma = \sqrt{\frac{\sum\left(n_{i*}(x_i{}^2)\right)}{n} - \overline{x}^2}}$$

$$= \frac{56192}{40} - 37.4^2$$

$$= 1404.8 - 1398.76$$

$$\boldsymbol{= 6.04}$$

التباين بتطبيق الصيغة الاخيرة يساوي 2.3 وهي نفس النتيجة الصيغة السابقة.
وبالتالي فان قيمة الانحراف المعياري ستكون نفسها:

$$\boldsymbol{\sigma = \sqrt{\sigma^2} = \sqrt{6.04} = 2.46}$$

اذن الانحراف المعياري لأزان عينة من التلاميذ هو: 2.46.

4-5- خصائص الانحراف المعياري :

1- الانحراف المعياري للمقدر الثابت يساوي صفر، أي اذا كان لدينا القراءات التالية :

X: a, a, a, a,…a، حيث أن a مقدار ثابت فان ($\boldsymbol{\sigma = 0}$)أي الانحراف المعياري يساوي صفر.

2- اذا اضيف مقدرا ثابت الى كل قيمة من قيم المفردات، فان الانحراف المعياري للقيم الجديدة (بعد الاضافة) تساوي الانحراف المعياري للقيم الاصلية (القيم بعد الاضافة)، فاذا كانت القيم الاصلية $x_1, x_2, x_3, x_4, \ldots x_n$، وتم اضافة مقدار ثابت a الى كل قيمة من قيم x، فان الانحراف المعياري للقيم الجديدة y: حيث $(y= x+a)$; $x_1 +a, x_2 +a, x_3 +a, \ldots x_n +a$ تساوي الانحراف المعياري للقيمة الاصلية.

3- اذا ضرب كل قيمة من قيم المفردات في مقدار ثابت، فان الانحراف المعياري للقيم الجديدة يساوي الانحراف المعياري للقيم الاصلية مضروبا في المقدار الثابت، أي اذا كانت x هي القيم الاصلية، وكانت القيم الجديدة هي $y=(a)(x)$، حيث a مقدار ثابت، فان $\sigma_y=(a)(\sigma_x)$

مزايا وعيوب الانحراف المعياري :

-أنه أكثر مقاييس التشتت استخداما.

-يسهل التعامل معه رياضيا.

-عند حسابه تؤخذ كل القيم في الاعتبار.

-من عيوبه يتأثر بالقيم الشاذة.

5-مقاييس التشتت النسبي

1-5- معامل الاختلاف:

هو واحد من المقاييس المستخدمة لقياس درجة التشتت، وفيه يحسب التشتت كنسبة مئوية من قيم مقياس النزعة المركزية، ومن ثمة يفضل استخدام معامل الاختلاف عند مقارنة درجة تشتت بيانات مجموعتين أو أكثر مختلفة لها وحدات قياس مختلفة، بدلا من الانحراف المعياري او الانحراف الربيعي، لان معامل الاختلاف يعتمد على التغيرات النسبية في القيم عن مقياس النزعة المركزية، بينما

يعتمد الانحراف المعياري أو الانحراف الربيعي على التغيرات المطلقة للقيم، فعند مقارنة درجة تشتت بيانات الاطوال بالسنتمتر، وبيانات الاوزان بالكيلوغرام لا يمكن الاعتماد على الانحراف المعياري في هذه المقارنة، انما يستخدم معامل الاختلاف ومن ثم يطلق عليه معامل الاختلاف النسبي، وفيما يلي بعض هذه المعاملات:

5-1-1- معامل الاختلاف النسبي:

يرمز له بالرمز: « CV » ويحسب بتطبيق المعادلة التالية :

صيغة رقم (21)

$$\boxed{cv = \frac{\sigma}{\overline{x}} * 100}$$

مثال10: البيانات التالية تمثل المتوسط الحسابي والانحراف المعياري لعينتين من الطلبة في الرياضيات:

المقاييس	المجموعة الاولى	المجموعة الثانية
$= \overline{x}$	11	14
$= \sigma$	5	3

الحل :

معامل الاختلاف النسبي للمجموعة الاولى :

$$cv = \frac{\sigma}{\overline{x}} * 100 = \frac{5}{11} * 100 = 45.45\%$$

معامل الاختلاف النسبي للمجموعة الثانية :

$$cv = \frac{\sigma}{\overline{x}} * 100 = \frac{3}{14} * 100 = 21.43\%$$

يلاحظ أن درجة تشتت نقاط المجموعة الثانية أقل من درجة تشتت نقاط المجموعة الأولى.

2-1-5-معامل الاختلاف الربيعي :

له نفس أهداف وخصائص معامل التشتت النسبي يرمز له بالرمز:

$$cv_q = \frac{(Q_3 - Q_1)}{M_e} * 100$$

صيغة رقم (22)

2-5- الدرجة المعيارية:

يرمز لها بالرمز: « Z » وتقيس الانحرافات عن الوسط الحسابي بوحدات من الانحراف المعياري يسمى "المتغير المعياري"، وتحسب الدرجة المعيارية على اساس حساب الفرق بين القيمة والمتوسط الحسابي مقسوما على الانحراف المعياري كما يتضح من الصيغة التالية:

صيغة رقم (23)	$Z = \frac{x - \overline{x}}{\sigma}$

حيث :

X هي قيمة معطاة

$\overline{x}$هو المتوسط الحسابي

σ هو الانحراف المعياري.

- و الدرجة المعيارية على ذلك قد تساوي صفرا في حالة تساوي القيمة بالمتوسط.
- و قد تكون الدرجة المعيارية موجبة الاشارة اذا كانت القيمة أعلى من المتوسط.
- كما يمكن للدرجة المعيارية أن تكون سالبة الاشارة إذا كانت القيمة أقل من المتوسط.

تمكن الدرجة المعيارية من اجراء مقارنات بين القيم من مجموعات مختلفة في المتوسط الحسابي والانحراف المعياري، حيث أن استخدام القيم الخام في هذه الحالة لأجل المقارنة بين عدد من المجموعات لا تعطي معنى أو دلالة بل يؤدي الى

نتائج مضللة، وقد يساعد المتوسط الحسابي للمجموعة على معرفة مدى تقد الفرد أو تأخره بالنسبة لمجموعته، فعلى سبيل المثال لو أن طالب تحصل على العلامة (16) في احدى المواد الدراسية وكان المتوسط الحسابي لطلاب الفصل في في تلك المادة هو (15) يمكن من الوهلة الاولى اعتبار أن هذا الطالب تحصيله اعلى من المتوسط، لكن حتى بذلك لا يمكن معرفة مركز هذا الطالب في القسم، أي مدى بعد العلامة (16) عن المتوسط بالنسبة للمجموعة. لذلك يحتاج الباحث لأي مقارنة مدى ارتفاع القيمة او انخفاضها عن المتوسط بمقياس من مقاييس التشتت والتي من أفضلها الانحراف المعياري وهو ما يستدعي استخدام القيم المعيارية كما يتضح من المثال التالي:

الاختبار	علوم طبيعية	رياضيات	فزياء
علامة الطالب	16	13	15
المتوسط	17	11	12
الانحراف المعياري	2	1	3

بالنظر الى الانحرافات المعيارية لعلامات الاختبارات الثلاث يتبين أن الانحراف المعياري يشير الى أن متوسط تشتت علامات امتحان الفيزياء عن المتوسط هو 3 نقاط، ما يعني أن بعض الدرجات تزيد أو تقل عن المتوسط بأكثر أو أقل من 3 نقاط.

أما متوسط تشتت اختبار مادة الرياضيات عن المتوسط فهو "1" نقطة لذلك فان علامة 13 للطالب في هذه المادة تقع أعلى المتوسط بقدر 2 نقطة أو انحرافين معياريين **من المحتمل أن تكون أعلى الدرجات لأنها أعلى من المتوسط بكثير.**

بهذا يتبين أن الدرجات الخام تعطي صورة مضللة عند المقارنة في مثل هذه الحالات، حيث اذا قارنا علامات الطالب بباقي زملائه في الفصل مع الاخذ بعين الاعتبار كل من المتوسط العام ودرجة التشتت نجد أن أفضل العلامات هي علامة الرياضيات تليها علامة الفيزياء وأضعفها علامة العلوم الطبيعية.

بذلك فان العلامات الخام 16، 13، 15، لا يمكن مقارنتها بطريقة مباشرة لأن التوزيع التكراري لكل اختبار مختلف عن الاخر من حيث المتوسط والانحراف المعياري، ما يؤدي الى اختلاف وحدات القياس في كل منها. وللتغلب على هذه المشكلة نلجأ الى تحويل البيانات (العلامات) الخام في كل اختبار الى ميزان مشترك ومتفق في المتوسط والانحراف المعياري، وبذلك نتمكن من اجراء عملية المقارنة السليمة، مع الاشارة الى أن هذا التحويل لا يغير من شكل التوزيع التكرار للبيانات الخام.

بناء على ما سبق نصل الى أن الدرجات المعيارية لا تتوزع توزيعا اعتداليا بمجرد تحويل القيم الاصلية الى درجات معيارية، اذ لا يتحقق ذلك الا اذا كان توزيع القيم الاصلية اعتداليا، أي قبل تحويلها الى درجات معيارية. بالإضافة الى أنه يمكن اشتقاق الدرجات المعيارية من البيانات ذات القياس الفئوي أو النسبي.

مثال11: في جدول توزيع عينة من الاطفال حسب الاوزان: والمطلوب ايجاد الدرجة المعيارية للوزن 35 كلغ والدرجة المعيارية للوزن 45 كلغ. مع العلم أن المتوسط الحسابي قدر بـ 37.4، والانحراف المعياري بـ 1.35.

بتطبيق الصيغة (z) نتحصل على ما يلي:

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

$$= \frac{35 - 37.4}{1.35}$$

$$= -1.78$$

و منه يمكن القول أن الوزن 35 كلغ يقل عن المتوسط الحسابي بـ 1.78 درجة معيارية.

أما القيمة المعيارية للقيمة 45 فهي كما يلي:

$$= \frac{\mathbf{45 - 37.4}}{\mathbf{1.35}}$$

$$= \mathbf{5.63}$$

أي أن الوزن 45 كلغ يبتعد عن الوسط الحسابي بـ 5.63 درجة معيارية.

تحويل الدرجات المعيارية الى القيم الاصلية:

يمكن اعادة القيم المعيارية الى القيم الاصلية وذلك بتطبيق الصيغة التالية :

صيغة رقم .. $\boldsymbol{x = \overline{x} \mp Z * S}$

في المثال السابق ما هي القيمة المقابلة للدرجة المعيارية (+2)،

ملاحظة المقصود بالدرجة المعيارية (+2) هو أن القيمة الخام تزيد عن المتوسط الحسابي بمقدار 2 انحراف معياري .

و بهذا تكون القيمة المقابلة للدرجة المعيارية (+2) تساوي :

$$x = 37.4 + (2 * 1.35) = 40.1$$

الفصل السابع: مقاييس الشكل

تمهيد

رأينا في الفصل السابق انه عند تمثيل بيانات الظاهرة في شكل منحنى تكراري، فان هذا المنحنى ياخذ أشكال مختلفة، فقد يكون هذا المنحنى متماثل بمعنى له قمة في المنتصف، واو اسقطنا عمودا من قمته على المحور الافقي لشطره نصفين متماثلين، مثل منحنى التوزيع الطبيعي.

وانه في حالة كان المنحنى متماثل، تتساوى قيم كل من المنوال، المتوسط الحسابي والوسيط، أي كلها تقع على نقطة واحد، ولكن في كثير من الحالات يكون هناك قيم كبيرة في البيانات تجذب اليها كل من المنوال والوسيط، وهذا ما يجعل المنحى التكرار يتمدد جهة اليمين، ومشيرا الى وجود التواء ناحية اليمين. كما يمكن أن يحدث العكس حيث تظهر القيم الصغرى بقوة في التوزيع فتجذب اليها المنوال والمتوسط الحسابي مما يجعل المنحنى التكراري يتمدد جعة اليسار مشيرا بذلك الى وجود التواء جهة اليسار. كما يمكن من خلال الشكل البياني معرفة ما اذا كان توزيع البيانات منبسطأ ومدبب، هذا من الناحية البيانية، الا انه يمكن احصائيا وبطرق حسابية تحديد نوع الشكل ان كان كان متماثلا أو ملتوي، أو حالة الشكل ان كان مدبب أو مفلطح من خلال ما يعرف بمقاييس الشكل التي يهتم هذا الفصل بعرضها والتفصيل فيها.

1-مقاييس الالتواء

توجد مقاييس كثيرة لقياس الالتواء في التوزيعات الاحصائية، وتحديد نوع الشكل ان كان متماثلا، شكل بالتواء موجب أو شكل بالتواء سالب منها ما يلي:

1-1-طريقة "بيرسون" « Pearson » :

تأخذ هذه الطريقة في الاعتبار العلاقة بين الوسط الحسابي والوسيط والمنوال، في حالة ما اذا كان التوزيع قريب من التماثل وليس شديد الالتواء وذلك من خلال العلاقة التالية:

صيغة رقم 1

$$\beta_1 = \frac{3(\bar{x} - Me)}{\sigma}$$

حيث: $\bar{x}$ المتوسط الحسابي

- Me الوسيط
- σ الانحراف المعياري

و يمكن من خلال الاشارة التي يأخذها هذا المعامل الحكم على شكل الالتواء، كما يلي:

قيمة معامل (ألفا)	نوع المنحنى
$\beta_1 = 0$	منحنى التوزيع التكراري متماثل
$\beta_1 > 0$	منحنى التوزيع التكراري سالب الالتواء
$\beta_1 < 0$	منحنى التوزيع التكراري موجب الالتواء

2-1-معامل الاتواء الربيعي لـ "يول" "Yule":

تعتمد هذا المعامل في حسابه على قيم الربيعات الاول، الثاني والثالث، يفضل استخدام هذا المعامل في حالة البيانات التي تحتوي على قيم شاذة، وأيضا البيانات التي لا نعرف لها توزيع محدد، ويمكن ايجاد قيمته من الصيغة التالية: (Boukella-Bouzouane, 2001)

صيغة رقم 2 :

$$\boxed{C_y = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)}}$$

تفسر نتائجه بنفس الطريقة بالنسبة لمعامل "بيرسون".

مثال: البيانات التالية تمثل أوزان 8 أفرد:

66 85 52 78 91 74 58

المطلوب:

1- حساب معامل الالتواء بطريقة بيرسون.

1- حساب معامل الالتواء الربيعي

الحل:

1- حساب معامل الالتواء بيرسون:

و في هذه الحالة يتطلب تطبيق الصيغة الاحصائية رقم (1). التي تستدعي معرفة قيمة كل من المتوسط الحسابي، الوسيط والمنوال بالإضافة الى قيمة الانحراف المعياري ؛ بعد عرض البيانات السابقة في جدول لتسهيل العمليات الحسابية.

المبحوث	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	$\sum$
الوزن x	58	74	91	80	78	52	85	66	584
X^2	3364	5476	8182	6400	6084	2704	7225	4356	43890
ترتيب القيم	52	58	66	74	78	80	85	91	
رتبة الوسيط	1	2	3	4	5	6	7	8	

1- حساب قيمة المتوسط الحسابي:

$$\overline{x} = \frac{\sum x_i}{n}$$

$$= \frac{584}{8} = 73$$

المتوسط الحسابي يساوي: $\overline{x} = 73$

2- استخراج قيمة الوسيط M_e

بعد ترتيب المشاهدات ترتيبا تصاعديا وحساب رتبة الوسيط يتبين الن قيمة الوسيط هي :

و بما أن عدد المشاهدات زوجي فان رتبة الوسيط تتموضع بين رتبتين الرتبة الرابعة والرتبة الخامسة، ومنه فان قيمة الوسيط هي متوسط القيمتين المقابلتين للرتبة الاولى والرتبة الثانية.

$$M_e = \frac{74 + 78}{2} = 76$$

$$M_e = 76$$

الوسيط يساوي: $M_e=76$

2- حساب قيمة الانحراف المعياري :

$$\sigma = \sqrt{\frac{\sum x_i^2}{n} - \overline{x}^2}$$

$$= \sqrt{\frac{43890}{8} - 76^2}$$

$$= \sqrt{157.25} = 12.54$$

الانحراف المعياري $\sigma = 12.54$

3- حساب قيمة معامل الالتواء "بيرسون":

$$\beta = \frac{3(\overline{x} - Me)}{\sigma}$$

$$\beta = \frac{3(73-76)}{12.54} = -0.72$$

ومنه فان الشكل سالب الالتواء لان قيمة معامل بيرسون للالتواء كانت قيمة سالبة .

ثانيا حساب معامل الالتواء الربيعي لبيرسون:

يحتاج هذا المعامل كما سبق تعريفه الى قيم كل من الربيع الاول والربيع الثاني والربيع الثالث بعد حساب الرتب الخاصة بها تم التوصل الى النتائج التالية:

Q_1=62, Q_2=M_e=76, Q_3 =82.5

بالتعويض في صيغة الخاصة بمعامل الالتواء الربيعي نجد :

$$C_y = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)}$$

$$C_y = \frac{(82.5-76) - (76-62)}{(82.5-62)}$$

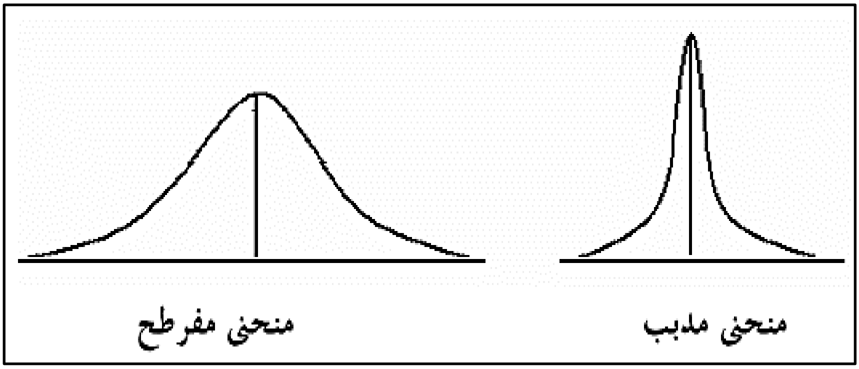
$$= \frac{6.5-14}{20.5} = \frac{-7.5}{20.5}$$

$$= -0.37$$

ومنه فان الالتواء بهذه الصيغة كان سالب الالتواء وهي نفس النتيجة المتوصل اليها من معامل الالتواء الاول .

3-معامل التفلطح "بيرسون"

عند تمثيل التوزيع التكراري في شكل منحنى تكراري قد يكون هذا المنحنى منبسط، أو مدبب، فعندما يتركز عدد كبير من القيم بالقرب من منتصف المنحنى، ويقل في طرفيه، يكون المنحنى مدببا، وعندما يتركز عدد أكبر من القيم على طرفي المنحنى، ويقل بالقرب من المنتصف يكون المنحنى مفرطحا، أو منبسط ويمكن اظهار ذلك في الشكل التالي :



ويمكن قياس التفرطح باستخدام عدد من الطرق، منها طريقة العزوم، حيث يحسب معامل التفرطح (K) بتطبيق الصيغة التالية :

صيغة رقم 3	$\mathbf{k} = \dfrac{\frac{1}{n}\sum(x_i - \overline{x})^4}{\sigma^4}$

حيث أن :

- $\frac{1}{n}\sum(x_i - \overline{x})^4$ هو العزم الرابع حول الوسط الحسابي.
- σ^4 هو الانحراف المعياري مرفوع الى القوة اربعة .

ومن النتائج المتوصل اليها يمكن وصف منحنى التوزيع من حيث التفرطح أو التدبب كالتالي:

- اذا كان K=3 يكون منحنى التوزيع معتدلا.
- اذا كان K>3 يكون منحنى التوزيع مدببا.
- اذا كان K<3 يكون منحنى التوزيع منبسطا (مفرطحا).

لو انطلقنا من المثال السابق نجد أن قيمة التفلطح في تلك المجموعة بعد حساب معرفة مجموع فرق كل مشاهدة من تلك المشاهدات عن وسطها الحسابي يقدر ب:

ا- حساب مجموع فروق كل قيمة عن وسطها الحسابي:

المبحوث	1	2	3	4	5	6	7	8	$\sum$
الوزن x	58	74	91	80	78	52	85	66	584
x_i-x	-7	12	-21	5	7	18	1	-15	0
$(x_i\text{-}x)^4$	2401	20736	194481	625	2401	104976	1	50625	376246

اذن قيمة $(x_i\text{-}x)^4 = 376246$

و مما سبق نعرف أن قيمة الانحراف المعياري تقدر بـ $\boldsymbol{\sigma = 12.54}$

بالتعويض في صيغة حساب معامل التفرطح نتحصل على ما يلي:

من هذه النتيجة يمكن القول ان شكل توزيع البيانات مفرطح.

$$\boxed{\mathbf{k} = \frac{\frac{\sum(x_i - \overline{x})^4}{n}}{\sigma^4} = \frac{47030.75}{24728.07} = 1.901}$$

تمارين

التمرين الاول: البيانات التالي تمثل توزيع عينة من المساكن حسب عدد الغرف:

عدد الغرف	1	2	3	4	5	6	7	8	$\sum$
ni	2	10	9	30	15	18	10	6	100

المطلوب :

حساب معاملات الالتواء والتفلطح.

التمرين الثاني:

البيانات التالية تمثل المكافئات الفصلية لعينة من الاطارات مقدرة بـ دج.

قيمة المكافئة	ni
1000-800	23
1100-1000	30
1200-1100	70
1300-1200	7
1500-1300	10
المجموع	140

المطلوب :

1-حساب الانحراف المعياري.

2- حساب معامل الالتواء.

3- حساب معامل التفلطح.

الفصل الثامن: مفهوم الارتباط ومعاملاته

1- مفهوم الارتباط

2- معامل بيرسون

3- معمل سبيرمان

4- معمل الاقتران والتوافق

5- معامل الارتباط النقطي

6- معامل لامبدا

7- معمل إيتا

1-مفهوم الارتباط

تركز عدد من البحوث الاجتماعية على تحليل العلاقة بين أكثر من متغير حيث يهتم الباحث بتحديد كيف وإلى أي مدى يرتبط متغيرات أو أكثر، والإحصاءات المستخدمة في التحليلات ثنائية المتغير، فالمنطق متشابه إلى حد كبير وإن كانت الإحصاءات المستخدمة في دراسة العلاقات متعددة المتغير تتسم بدرجة كبيرة من التعقيد.

وعند تحليل العلاقة بين متغيرين يهتم الباحث بالإجابة عن ثلاثة تساؤلات هل ترتبط هذه المتغيرات ؟ وما هو اتجاه وشكل الارتباط الموجود ؟ هل هناك احتمال أن يكون الارتباط الذي تمت ملاحظته بين حالات العينة أحد خصائص المجتمع البحثي أم أن هذا الارتباط هو نتاج لصغر حجم العينة التي قد تكون غير ممثلة للمجتمع البحثي ؟

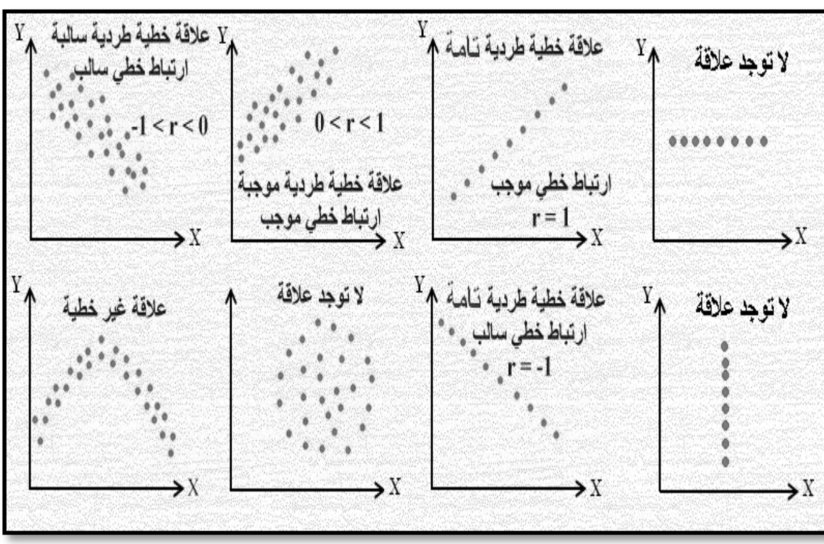
يمكن تحديد الارتباط بين متغيرين من خلال استخدام مجموعة من الإحصاءات تعرف باسم معاملات الارتباط ومعامل الارتباط هو رقم يلخص التحسن في تخمين القيم على متغير واحد لأي حالة على أساس معرفة قيم المتغير الثاني، فكلما ارتفع المعامل قوي الارتباط، ومن ثم تحسنت قردتنا التنبؤية أو التفسيرية. وتتراوح معاملات الارتباط بين صفر وواحد (-1، 0، +1)، وتشير القيم التي تقترب من 1 إلى وجود ارتباط قوي نسبياً أما تلك التي تقترب من صفر فتشير إلى ارتباط ضعيف نسبياً. ويتطلب كل مستوى قياس أنواع مختلفة من الحسابات وبالتالي فلكل من هذه المستويات اختبارات ارتباط مختلفة.

إضافة إلى حجم الارتباط يهتم الباحث بمعرفة اتجاه العلاقة بين المتغيرين فهل هي علاقة طردية أو عكسية، وتجدر الإشارة هنا إلى أن مفهوم

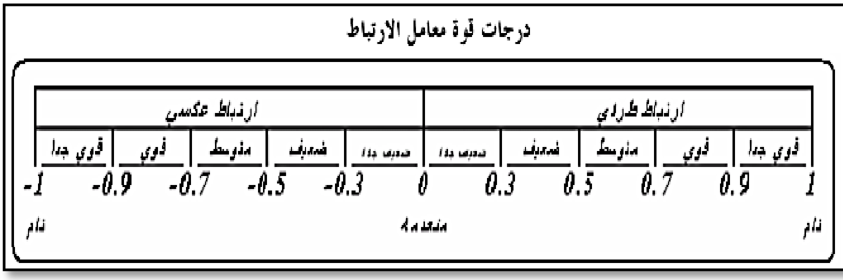
الاتجاه ليس له معنى على مستوى القياس الأسمى، حيث إن الأرقام على هذا المستوى من القياس مجرد عناوين للفئات، وبالتالي لا تتغير إشارات معاملات الارتباط الاسمية فكلها موجبة وتشير إلى مدى قوة الارتباط، أما على مستوى قياس الفترة فإن الإشارات تتغيرولها دلالات هندسية على درجة عالية نسبياً من التعقيد.

وأخيراً يهتم الباحث باختبارات الدلالة الإحصائية وهي الاختبارات التي توضح احتمال أن تكون العلاقات التي يلاحظها الباحث نتاج التحيز في عملية الاختبار بدلاً من أن تعكس علاقات موجودة فعلاً في مجمع البحث. (علام اعتماد، رسلان يسرى، 1987)

2-نوع وشدة العلاقة الارتباطية



بالطبع عرفنا أن قيمة معامل الارتباط محصورة فى الفترة المغلقة [-1، 1] وتتحدد نوعية الارتباط من الجدول التالى :



3- تفسير نتائج معامل الارتباط

تتراوح قيم معامل الارتباط بين +1 و-1، وتكون درجة العلاقة قوية كلما اقترب مقداره من هذين الحدين (أي +1 و-1) ويكون العلاقة تامة اذا بلغت تمام القيمة (1) سواء كانت موجبة او سالبة. كما تضع العلاقة بين المتغيرين كلما اقترب مقدار معامل الارتباط من الصفر، في حين تنعدم العلاقة اذا قدر معامل الارتباط بصفر، هذا من حيث شدة العلاقة بين المتغيرين، أما بالنسبة لاتجاه العلاقة بين المتغيرين فتمثله الاشارة التي يحملها مقدار معامل الارتباط فاذا كانت النتيجة موجبة (أي طردية) أي كلما زادت قيم احد المتغيرين زادت قيم المتغير الآخر، أما العلاقة السالبة فتعني أن المتغيرين يسيران في اتجاهين متعاكسين ؛ كلما زاقيم أحد المتغيرين تراجعت قيم المتغير المقابل له والعكس بالعكس. (محمد شامل بهاء الدين، 2005، صفحة 538)

4- التباين المشترك

ان ابسط طريقة للبحث فيما اذا كان المتغيرين مرتبطين هي معرفة ما اذا كانا يتغيران معا، أي بطريقة مشتركة وهو ما يعرف بالتباين المشترك، ولفهم هذا الاخير لابد من الرجوع الى مفهوم التباين variance الذي يمثل المقدار الوسطي لتغير البيانات عن قيمة وسطها الحسابي كما يتضح من صيغة حسابه التالية:

$$v = \frac{\sum(x_i - \overline{x})^2}{n-1}$$

حيث تمثل الصيغة حاصل قسمة الفرق بين كل مشاهدة ومتوسطها الحسابي على حجم العينة. واذا كان اهتمامنا منصب على معرفة اذا كان المتغيران مرتبطين، فإننا نهتم بمعرفة اذا كانت تغيرات احدهما تقابلها تغيرات مشابهة في المتغير الثاني. لذلك فعندما يبتعد أحد المتغيرين عن القيمة المتوسطة فإننا نتوقع أن ينحرف المتحول الاخر عن قيمته المتوسطة بصورة مشابهة. ويمكن حساب التباين المشترك باستخدام الصيغة التالية :

$COV = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{n-1}$	**الصيغة رقم (01)**

أو الصيغة التالية :

$COV = \frac{\sum(x_i - y_i) - (n * \overline{x} * \overline{y})}{n-1}$	**الصيغة رقم (02)**

يتضح ان هذه الصيغ انها شبيهة بصيغة حساب التباين لكن بدلا من تربيع الفروق نقوم بضرب الفروق للمتغير الاول بمقابلاتها في المتغير الثاني. ويسمح حساب التباين المشترك بتقييم فيما اذا كان المتغيرين مرتبطين ببعضهما البعض. وتشير القيمة الموجبة للتباين الى ان انحراف أحد المتغيرين عن قيمة المتوسط يعني انحراف المتحول الآخر عن قيمته المتوسطة بنفس الاتجاه. أما القيمة السالبة فتدل على ان انحراف احد المتغيرين عن القيمة المتوسطة باتجاه ما (كالزيادة مثلا) يعني انحراف المتغير الاخر عن قيمته المتوسطة بالاتجاه المعاكس (النقصان). كما أنه كلما زادت قيمة التباين المشترك دل ذلك على قوة التبعية بين المتغيرين x و y. أي كلما كانت قيمة التباين عالية كلما كانت العلاقة بين

المتغيرين قوية والعكس كلما اقتربت قيمة التباين من الصفر يمكن توقع ارتباط ضعيف بين المتغيرين.

لكن يطرح استخدام التباين المشترك في تحديد العلاقة بين متغيرين مشكلة الاختلاف في وحدات قياس المتغيرين، التي تؤدي الى نتائج مختلفة بين مجموعتين، لكل منها وحدة قياس مختلفة خصوصا في العلوم الاجتماعية مثل قياس وجهات النظر المختلفة والتي يستحيل فيها القيام بعملية توحيد وحدة القياس كما هو الحال في العلوم الأخرى مثل التحويل الى الكلغ أو المتر وهو ما يصعب معها المقارنة بين المجموعات لمعرفة حجم التباين الخاص بكل منها. وللقيام بعملية التوحيد القياسي للمتغيرات المراد دراستها ادخل الانحراف المعياري للمتغيرين على صيغة التباين المشترك لتصبح لدينا صيغة تحقق هذا الهدف تعرف بمعامل ارتباط بيرسون تقوم على قسمة التباين المشترك على حاصل جداء الانحراف المعياري للمتغيرين x, y . (لجنة التأليف والترجمة، 2007، الصفحات 118-122) لتصبح الصيغة الاحصائية لقياس العلاقة بين المتغيرات كالتالي:

5- معامل ارتباط بيرسون

يستخدم معامل ارتباط بيرسون لحساب قيمة معامل الارتباط عندما يكون المتغيران المراد قياس الارتباط بينهم متغيرات كمية ويشترط تساوى عدد حالات كل من المتغير المستقل والمتغير التابع ونستخدم القانون التالي لحساب قيمة معامل ارتباط بيرسون الذي يرمز له بالرمز (r) :

الصيغة رقم (03): تعتمد في حسابها على قيمة التباين المشترك والانحراف المعياري كما سبقت الاشارة اليه.

$$r = \frac{Cov(x, y)}{(\sigma_x)(\sigma_y)}$$

الصيغة رقم (04)

$$r=\frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum(x_i-\overline{x})^2}*\sqrt{\sum(y_i-\overline{y})^2}}$$

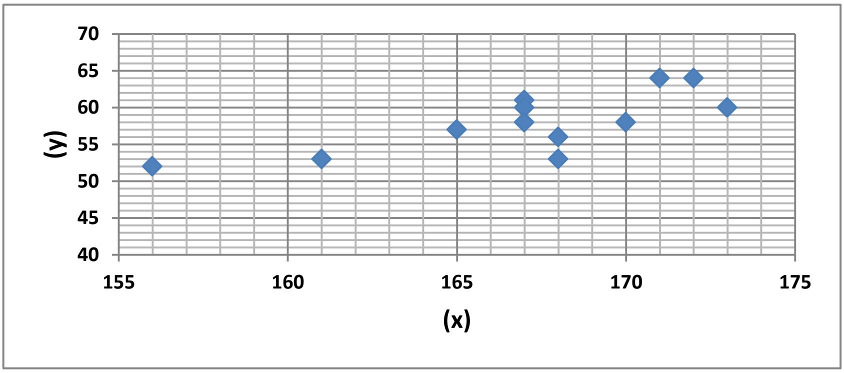
مثال: البيانات التالية تمثل توزيع عينة من الطلبة حسب الطول والوزن :

المطلوب حساب قيمة التباين المشترك والارتباط بين المتغيرين بصيغتين مختلفتين بعد التأكد من خطية العلاقة بين المتغيرين برسم سحابة النقاط:

			العمود (1)	العمود (2)	العمود (3)	العمود (4)	العمود (5)	العمود (6)
رقم الطالب	x	y	$x_i * y_i$	$x_i - \overline{x}$	$(y_i - \overline{y})$	$(y_i - \overline{y})*(x_i - \overline{x})$	$(x_i - \overline{x})^2$	$(y_i - \overline{y})^2$
1	167	61	10187	0,08-	3	0,24-	0,01	9
2	167	60	10020	0,08-	2	0,16-	0,01	4
3	171	64	10944	3,92	6	23,52	15,37	36
4	173	60	10380	5,92	2	11,84	35,05	4
5	156	52	8112	- 11,08	6-	66,48	122,77	36
6	168	56	9408	0,92	2-	1,84-	0,85	4
7	172	64	11008	4,92	6	29,52	24,21	36
8	165	57	9405	2,08-	1-	2,08	4,33	1
9	167	58	9686	0,08-	0	0	0,01	0
10	168	53	8904	0,92	5-	4,6-	0,85	25
11	161	53	8533	6,08-	5-	30,4	36,97	25

0	8,53	0	0	2,92	9860	58	170	**12**
180	248,92	157			116447	696	2005	$\sum$
$\boldsymbol{\sigma_y = 4.04}$	$\boldsymbol{\sigma_x = 4.76}$					$\bar{y}$=58	$\bar{x}$=167,08	الحسابي

أولا: التأكد من خطية العلاقة بين المتغيرين من خلال رسم سحابة النقاط:



ثانيا :ايجاد قيمة التباين المشترك:

أ- بناء على بيانات الجدول أعلاه وبالتعويض في الصيغة رقم (1) نجد قيمة التباين المشترك كالتالي:

$$COV = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{N-1}$$

$$COV = \frac{157}{11} = 14.27$$

ب- باستخدام الصيغة رقم (2) لمعامل للتباين المشترك نجد بعد التعويض:

$$\boldsymbol{COV = \frac{\sum(x_i - y_i) - (n * \bar{x} * \bar{y})}{N-1}}$$

$$COV = \frac{116447 - (12)(167.08)(58)}{12-1}$$

$$= \frac{116447 - 116287.68}{11}$$

$$COV = \frac{159.32}{11} = 14.48$$

ثالثا: ايجاد قيمة معامل الارتباط بيرسون:

أ- باستخدام **الطريقة الاولى**، الصيغة رقم (3):

$$\boldsymbol{r} = \frac{\boldsymbol{Cov(x, y)}}{(\boldsymbol{\sigma_x})(\boldsymbol{\sigma_y})}$$

تتطلب الصيغة ايجاد التباين المشترك وقد تم حسابه بالإضافة الى قيمة الانحراف المعياري لكل من المتغير (x) و(y)

**الانحراف المعياري للمتغير(x) من العمود رقم05 في الجدول تساوي :

$$(\sigma_x) = \sqrt{\frac{(\boldsymbol{x_i} - \overline{\boldsymbol{x}})^2}{n-1}} = \sqrt{\frac{248.92}{11}}$$

$$\sqrt{22.63} = 4.76$$

**الانحراف المعياري للمتغير(y) من العمود رقم06 في الجدول تساوي :

$$(\sigma_y) = \sqrt{\frac{(\boldsymbol{y_i} - \overline{\boldsymbol{y}})^2}{\boldsymbol{n-1}}} = \sqrt{\frac{180}{11}}$$

$$\sqrt{16.36} = 4.04$$

بالتعويض في صيغة معامل الارتباط رقم (3) السابقة نجد :

$$r = \frac{14.27}{(4.76)(4.04)}$$

$$\frac{14.27}{19.23} = 0.74$$

ب- استخدام **الطريقة الثانية** لحساب معامل الارتباط بيرسون الصيغة رقم (4):

$$\boldsymbol{r} = \frac{\sum(\boldsymbol{x_i} - \overline{\boldsymbol{x}})(\boldsymbol{y_i} - \overline{\boldsymbol{y}})}{\sqrt{\sum(\boldsymbol{x_i} - \overline{\boldsymbol{x}})^2} * \sqrt{\sum(\boldsymbol{y_i} - \overline{\boldsymbol{y}})^2}}$$

$$= \frac{157}{\sqrt{248.92} * \sqrt{180}}$$

$$= \frac{157}{15.78 * 13.42}$$

$$= \frac{157}{211.77}$$

$$= \frac{157}{211.77} = 0.74$$

وهي نفس النتيجة السابقة وتشير الى وجود علاقة قوية بين المتغيرين

6- معامل الارتباط المعدل $r_{ajusté}$

- يتأثر معامل الارتباط بيرسون بحجم العينة حيث تزيد قيمته طرديا مع زيادة حجم العينة، في حين يتحيز معامل الارتباط اذا كان حجم العينة صغيرا بحيث يعطي قيمة مضللة لا تعبر عن قيمة الارتباط الحقيقي بين المتغيرين المدروسين « y », « x »، لذا أوجد الاحصائيون طريقة لتعديل القيمة المحصل عليها تتلخص في تطبيق الصيغة التالية:

الطريقة الاولى:

$$r_{aj=}\sqrt{1-\frac{(1-r^2)(n-1)}{n-2}}$$

صيغة: رقم (5)

الطريقة الثانية:

$$r_{aj}=\frac{r(1-r^2)}{2n}$$

صيغة رقم (6)

ان المعامل الناتج يدعى أيضا بالتقدير غير المتحيز لمعامل الارتباط ويستخدم اذ قل حجم العينة عن 30وحدة احصائية. أما اذا كان حجم العينة كبير ابتداء من 100 وحدة احصائية في حين أن الحجم المثالي للعينة فهو 200 وحدة احصائية. فإننا لا نتوقع وجود اختلافات بين معامل الارتباط بالطريقة العادية وحاصل معامل الارتباط المعدل[1].

مثال توضيحي :

في دراسة للعلاقة بين متغيرين (y,x) قدرت قيمة معامل الارتباط بـ r= 0.96، من عينة حجمها 10 وحدات احصائية، المطلوب ايجاد قيمة معامل الارتباط المعدل.

---

[1] الاحصاء التربوي، ص 130.

1- **بناء على الصيغة رقم (5)**

$$r_{aj=}\sqrt{1-\frac{(1-0.96^2)(10-1)}{10-2}}$$

$$=\sqrt{1-\frac{(1-0.92)(9)}{8}}$$

$$=\sqrt{1-\frac{0.72}{8}}=\sqrt{0.91}=0.95$$

قيمة معامل الارتباط بيرسون المعدل: r_{aj} 0.95=

2- **بناء على الصيغة رقم (6) :**

$$\boldsymbol{r_{(aj)}=r-\frac{r(1-r^2)}{2n}}$$

$$=0.96-\frac{0.96(1-0.96^2)}{2(10)}=0.96-\frac{0.96-0.075}{20}$$

$$=0.96-0.0037=0.95$$

قيمة معامل الارتباط بيرسون المعدل: r_{aj} 0.95=

وهي نفس القيمة السابقة.

7- شروط استخدام معامل ارتباط بيرسونr

- يستخدم معامل ارتباط بيرسون في حالة كان المتغيرين من مستوى القياس الفئوي، أو كلاهما من مستوى القياس النسبي أو أحدهما من مستوى القياس فئوي والاخر من مستوى القياس نسبي.
- وجود علاقة خطية بين المتغيرين.

- اعتدالية التوزيع.

8- وظائف معامل بيرسون

- تحديد قوة الارتباط بين متغيرين، أي بيان ما اذا كان الارتباط قويا، متوسطا أم ضعيفا.

- تحديد اتجاه العلاقة بين المتغيرين، أي بيان ما اذا كانت العلاقة طردية (موجبة)، أم عكسية (سالبة).

- يساعد معامل الارتباط في عمليات التنبؤ خاصة عندما تكون قيمته كبيرة أو يقارب الواحد الصحيح.

- تساعد معاملات الارتباط عموما على التحقق من صدق وثبات الاستبيان والاختبارات والمقاييس المختلفة. (محمد شامل بهاء الدين، 2005، صفحة 537)

9- اختبار الدلالة الاحصائية لمعامل الارتباط بيرسون

ان اختبار دلالة الارتباط تعني اختبار اذا ما كان معامل الارتباط مهما، أم أن له وجود مختلف عن الصفر. ويكون اختبار دلالة الارتباط (بأنه مساوي للصفر أو لأي قيمة معينة) يكون بناء على نظرية معينة أو بحوث سابقة أو كليهما. فاذا كان المتوقع أن العلاقة بين المتغيرين متوسطة فلا يجوز ان نختبر اختلافها عن الصفر. فاذا اقترحت الادبيات أن العلاقة بين درجات المستوى الثقافي للاسرة وتحصيل الطالب هي 0.4 (في المتوسط). فان الفرض المناسب للاختبار يكون اختبار معامل الارتباط 0.4 عن اختبار مساواته للصفر.

يهدف هذا الاختبار الى معرفة اذا ما كانت قيمة معامل الارتباط المحسوبة من تطبيق الصيغ الرياضية السابقة يدل احصائيا على وجود علاقة فعلية بين المتغيرين ام أن النتيجة المتوصل اليها تعود الى الصدفة فقط:

للقيام بهذا العمل يشترط تحقق بعض الشروط الاساسية المتمثلة فيما يلي:

- العشوائية في اخيار وحدات العينة، أي استخدام احدى الطرق العشوائية لسحب العينة.

- التوزيع الاعتدالي لدرجات المتغيرين أي أن يكون التوزيع متماثل، ويمكن التعرف على ذلك باستخدام مقاييس الشكل.

- تؤكد بعض المراجع على ضرورة أن يكون حجم العينة ≤ 30 وحدة احصائية، اذا كان أقل من ذلك يجب تعديل معامل الارتباط بيرسون.

ولاختبار الدلالة الاحصائية لمعامل الارتباط "بارسون" تتبع نفس خطوات اختبار الفرضيات الاحصائية كالتالية :

1- صياغة الفرضية الصفرية H_0 والفرضية البديلة H_1.

طرق صياغة الفرضيات حسب نوع الاختبار:

1.1. في حالة **اختبار ذو اتجاهين** تصاغ الفرضية الصفرية والفرضية البديلة كالتالي :

– الفرضية الصفرية $H_o=0$ لا توجد علاقة بين المتغيرين y , x.

– الفرضية البديل $H_1\neq0$ توجد علاقة بين المتغيرين y , x.

1.2. **في حالة الاختبار ذو اتجاه واحد موجب (ناحية اليمين):** تصاغ الفرضية الصفرية بناء على اطار نظري سابق أي من معلومة سابقة حول العلاقة بين المتغيرين: ولتكن مثلا أن العلاقة بين علامة الاحصاء وعلامة الرياضيات هي r= 0,75

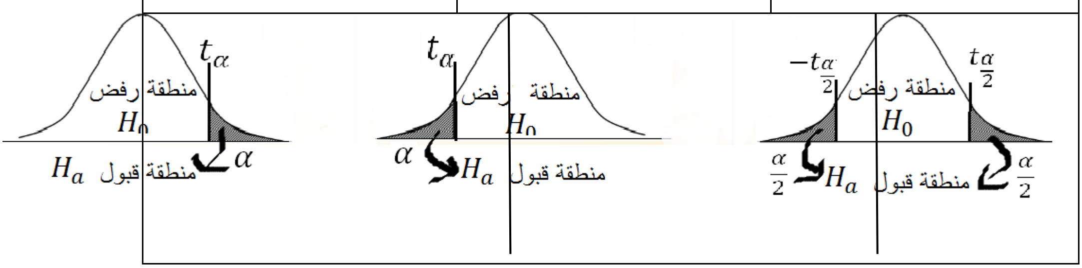
- الفرضية الصفرية $H_0: r = 0,75$ ، العلاقة بين المتغيرين x , y تساوي 0,75
- الفرضية البديل $H_1: r>0,75$ ، العلاقة بين المتغيرين x , y **أكبر** من 0,75

1.3. **في حالة الاختبار ذو اتجاه واحد سالب (ناحية اليسار):** تصاغ الفرضية الصفرية بناء على اطار نظري سابق أي من معلومة سابقة حول العلاقة بين المتغيرين: ولتكن مثلا أن العلاقة بين علامة الاحصاء وعلامة الرياضيات هي r= 0,75

- الفرضية الصفرية $H_0: r = 0,75$ ، العلاقة بين المتغيرين x , y تساوي 0,75
- الفرضية البديل $H_1: r<0,75$ ، العلاقة بين المتغيرين x , y **أصغر** من 0,75

و يمكن تلخيص ما سبق في الجدول التالي :

اختبار ذو اتجاهين Teste Bilatérale	**اختبار ذو اتجاه واحد ناحية اليسار** Teste unilatérale vers la gauche	**اختبار ذو اتجاه واحد ناحية اليمين** Teste unilatérale vers la droite
الفرضية الصفرية $H_0: \rho = 0$ الفرضية البديلة $H_a: \rho \neq 0$	الفرضية الصفرية $H_0: \rho = r$ الفرضية البديلة $H_a: \rho < r$	الفرضية الصفرية $H_0: \rho = r$ الفرضية البديلة $H_a: \rho > r$



2- صياغة الفرضيات الاحصائية؛ الفرضية الصفرية والفرضية البديلة، كما سبق توضيحه.

3- تحديد نوع الاختبار؛ ذو اتجاهين أو ذو اتجاه واحد وناحية اتجاهه.

4- تحديد مستوى الدلالة مثلا alpha=5% أو α= 0,05 غالبا في العلوم الاجتماعية.

5- حساب القيمة الحقيقية لمعامل الارتباط وفق الصيغ « r ».

6- ايجاد قيمة دالة الاختبار:

أ. **حالة العينات الصغيرة الحجم n <30**: نحسب قيمة « t » بتطبيق الصيغة التالية :

الصيغة رقم 01	$t = \dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}$

- حيث » « r هي قيمة معامل الارتباط بيرسون.
- و" « n » هي حجم العينة.

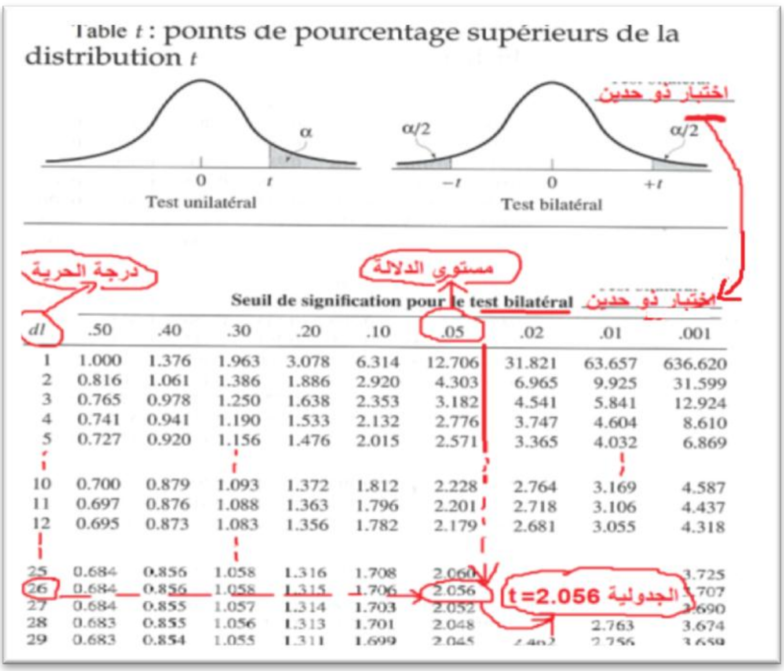
ب. **حالة العينات الكبيرة الحجم n ≥30**: نحسب قيمة « z » بتطبيق الصيغة التالية :

الصيغة رقم 2	$z = \dfrac{\sqrt{n-3}}{2} L_n \dfrac{(1+r)(1-P_0)}{(1-r)(1+P_0)}$

حيث: $L_{n(v)}$ هو اللوغاريتم الطبيعي للمقدار (v) وأن (Z) تتبع تقريبا التوزيع الطبيعي المعياري، ويتم تعيين مجال رفض الفرضية الصفرية H_0 من جدول التوزيع الطبيعي(Z).

7- حساب درجة الحرية (dl) مثلا في اختبار « t » هي (n-2) أي نطرح القيمة (2) من حجم العينة.

8- ايجاد القيمة الجدولية الخاصة باختبار « t » بناء على درجة الحرية (n-2) ومستوى الدلالة 0,05، وذلك من خانة التقاطع بين هتين القيمتين على الجدول الاحصائي لاختبار« t. test » في حالة العينات الصغيرة أقل من 30 وحدة احصائية. او اختبار« Z » في حالة $n \geq 30$. كما يتضح في الشكل البياني التالي:



Seuil de signification pour le test bilatéral

dl	.50	.40	.30	.20	.10	.05	.02	.01	.001
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.620
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
25	0.684	0.856	1.058	1.316	1.708	2.060	[illegible]	[illegible]	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	[illegible]	[illegible]	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	[illegible]	[illegible]	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	[illegible]	2.763	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	[illegible]	2.756	3.659

9- اختبار الفرض الصفري من خلال المقارنة بين قيمة « t » المحسوبة وقيمة « t » الجدولية. حيث يرفض الفرض الصفري في حالة كانت قيمة « t » المحسوبة أكبر من قيمة « t »الجدولية.

10- اتخاذ القرار بقبول أو رفض الفرضية الصفري بناء على النتيجة.[2]

معلومة حول طريقة تحديد القيمة الحرجة:

- تحديد المنطقة الحرجة التي يرفض عندها الفرض الصفري للاختبار عند مستوى الدلالة α، ويكتب كالتالي.

$$R.C.: |t| > t_{1-\frac{\alpha}{2}}(n-2)$$

اين تمثل القيمة $t_{1-\frac{\alpha}{2}}(n-2)$ قيمة $1-\frac{\alpha}{2}$ من قانون "ستيودنت" الى $(n-2)$ درجة الحرية ويتعلق الامر هنا باختبار ذو اتجاهين.[3]

أولا: مثال توضيحي في حالة عينة صغيرة الحجم في دراسة حول العلاقة بين هيكل السيارة وقوتها لعينة من 28 سيارة، وجد ان معامل بيرسون المعدل قدر بـ 0.94، ونريد اختبار الدلالة الاحصائية لهذه النتيجة، لمعرفة اذا كان هذا ينطبق على المجتمع أم ان النتيجة تعود الى الصدفة، في هذه الحالة يجب حساب القيم التالية:

1. حساب قيمة الاختبار (t) بتطبيق الصيغة أ:

$$t = \frac{0.95}{\sqrt{\frac{1-0.95^2}{28-2}}} = 15.18$$

2. اتخاذ القرار بناء على قيمة « t » المحسوبة كما يتضح أعلاه، و« t » المجدولة التي يمكن الحصول عليها من الجدول الاحصائي الخاص بقيم التوزيع الاحتمالي « t » من خانة التقاطع بين مستوى الدلالة 0.05 والقيمة درجة

---

2 http://grasland.script.univ-paris-diderot.fr/STAT98/stat98_6/stat98_6.htm#SIGNIFICATIVITE%20D%27UNE%20RELATION .

3 Malala, Analyse de corrélation,....

الحرية 26 degré de liberté الناتجة من (n-2) وذلك حسب طبيعة الفرضية والاختبار ان كان ذو اتجاهين كما هو الحال في هذا المثال، والمقدرة هنا بـ

3. **تحديد القيمة الحرجة**: اذا كان مستوى الدلالة 0.05 فان القيمة الحرجة في اختبار t.test بالصيغة التالية:

$$t_{1-\frac{\alpha}{2}}\,(n-2)$$

$$t_{1-\frac{0.05}{2}}\,(28-2)$$

$$= t_{0.975}\,(26)$$

$$, t_{0.975}(26) = 2.056\ \alpha = 0.05$$

ومنه فان « t » المجدولة (النظرية) تساوي **2.056**، وبناء على مقارنة النتيجتين يتبين أن القيمة المحسوبة أكبر من القيمة المجدولة أي $t = 15.18 > t_{0.975}(26)$، عند مستوى الدلالة α=0.05

4- **اتخاذ القرار**: بناء على ما سبق يمكن **اتخاذ القرار** التالي: رفض الفرض الصفري لان قيمة الاختبار المحسوبة جاءت أكبر من القيمة المجدولة، أي أن قيمة t المحسوبة جاءت في منطقة رفض الفرض الصفري ما يؤكد وجود علاقة بين متغير هيكل السيارة ومتغير السرعة بنسبة 95% مع احتمال وجود 5% من الخطأ في هذه الحقيقة.

ثانيا: مثال وضيحي الدلالة الاحصائية لمعامل ارتباط بيرسون « r » في حالة عينة كبيرة الحجم.

تتبع نفس الخطوات السابقة لكن بدل استخدام اختبار « t » نستخدم اختبار « z » بتطبيق الصيغة الرياضية التالية: مع توزيع (Z)

$$z = \frac{\sqrt{n-3}}{2} L_n \frac{(1+r)(1-P_0)}{(1-r)(1+P_0)}$$

حيث: $L_{n(v)}$ هو اللوغاريتم الطبيعي للمقدار (v) وأن (Z) تتبع تقريبا التوزيع الطبيعي المعياري، ويتم تعيين مجال رفض الفرضية الصفرية H_0 من جدول التوزيع الطبيعي(Z).

مثال توضيحي :

علامات 32 طالب في كل من مادة الرياضيات والاحصاء التي قدرت درجة العلاقة بينهما بـ r=0.935.

المطلوب اختبار الفرضية الصفرية التالية:

$$H_0 = P = 0.8$$

$$H_a > 0.8$$

عند مستوى الدلالة: α=0.05

الخطوة الثانية: حساب قيمة دالة الاختبار حيث أن $P_0 = 0.8$

$$z = \frac{\sqrt{n-3}}{2} L_n \frac{(1+r)(1-P_0)}{(1-r)(1+P_0)}$$

$$= \frac{\sqrt{32-3}}{2} L_n \frac{(1+0.935)(1-0.8)}{(1-0.935)(1+0.8)}$$

$$= \frac{5.39}{2} L_n \frac{(1.935)(0.2)}{(0.065)(1.8)}$$

$$= \frac{5.39}{2} L_n \frac{0.387}{0.117}$$

$$= \frac{5.39}{2} L_n \frac{0.387}{0.117}$$

$$= \frac{5.39}{2} L_n 3.307$$

$$= (2.69)(1.196)$$

$$z = 3.22$$

انطلاقا من الجدول » Z « وبناء على مستوى الدلالة 0.05، فان قيمة Z الجدولية أو النظرية كما يحددها جدول القيم الزائية أدناه:

جدول القيم الزائية حسب نوع الاختبار ومستوى الدلالة :

	مستوى الدلالة 0.01	مستوى الدلالة 0.05
اختيار ذو نهاية واحدة	2.33	1.64
اختيار ذو نهايتين	2.58	1.96

وبناء عليه فان قيمة الاختبار الجدولية باعتبار الاختبار ذو طرف واحد هي:

z= 1.64 وهي أصغر من القيم المحسوبة z=3.22 بالتالي يرفض الفرض الصفري

ويقبل الفرض البديل أي أن معامل الارتباط دال احصائيا، بتعبير آخر هناك دليل كافي بان معامل الارتباط في العينة المبحوثة يشير الى ان علامات الطلاب تفوق القيمة 8.

ملاحظة:

ان قيمة معامل الارتباط غير الدالة احصائيا تعامل في تفسير النتائج على أنها لا توجد علاقة بين الظاهرتين موضوع الدراسة.

معامل الارتباط والسببية

تشير قيمة معامل الارتباط الى درجة العلاقة بين المتغير المستقل والمتغير التابع، ولا يشير هذا المعامل الى تفسير هذه العلاقة، بمعنى لا ينبغي للباحث استنتاج وجود علاقات سببية بين المتغيرين مهما كانت النتيجة المتوصل اليها من حساب معامل الارتباط، فوجود معامل ارتباط مرتفع بين متغيرين (x,y) لا يعني ان المتغير x سبب في وجود المتغير y، أو العكس نظرا لأن هذين المتغيرين ارتبطا ببعضهما لانهما معا نتيجة لمتغير ثالث، ومثال لذلك العلاقة بين تناول وجبة الفطور والسمنة :

حيث توصلت احدى الدراسات الى وجود علاقة بين أخذ فطور الصباح وعدم التعرض للسمنة، وخلال عرض النتائج أشار الى أن فطور الصباح يعد **سببا** في عدم التعرض للسمنة، وهذا لا يتوافق مع منطق البحث العلمي لان النتيجة هي عبارة عن ارتباط وليس دراسة الاسباب حيث تتداخل عدة عوامل أخرى بين التعرض السمنة وأخذ فطور الصباح، مثل النشاط البدني الذي يقوم به الفرد، مدة النوم، الاكل اللامتوازن، الفقر... وهي متغيرات ضمنية تتوسط العلاقة بين أخذ فطور الصباح والسمنة. فبالنسبة للنشاط البدني مثلا اذا زاد

النشاط البدني تزيد الحاجة الى فطور الصباح وهنا تفقد العلاقة الأولى أهميتها أي بين فطور الصباح وعدم التعرض للسمنة، لأنه اذا كان النشاط البدني ضعيفا وتم التشجيع على أخذ فطور الصباح فان هذا قد يؤدي الى السمنة. وهكذا بالنسبة لباقي المتغيرات المذكورة، ومنه فان معامل الارتباط يساعد على معرفة تبعية المتغيرات الى بعضها وليس لتحديد السبب والنتيجة.

(Khan, Corrélation et causalité, 2019)

10- معامل ارتباط الرتب "سبيرمان"

يرمز لهذا المعامل بالرمز $\boldsymbol{\rho}$ « » يقرأ (رو) ويستعمل في حالة كان مستوى القياس رتبي لكل من المتغيرين أو لاحدهما، او اذا كانت التكرارات ضعيفة (Grosjean, Dommergues, 2011).

كما يستخدم عي حالة العلاقات غير الخطية لذلك لجأ الاحصائيون الى العمل على الرتب في حسابه بدل القيم الاصلية للمتغيرين.

وحددت صيغته كالتالي :

صيغة رقم 1:

$$\boldsymbol{\rho} = \mathbf{1} - \frac{\mathbf{6}\sum \boldsymbol{D}^{\mathbf{2}}}{\boldsymbol{n}(\boldsymbol{n}^{\mathbf{2}} - \mathbf{1})}$$

حيث: D هي الفرق في الرتب لكل من المتغير(x) والمتغير (y)

مثال:

البيانات التالية تمثل العلاقة بين الطول والوزن لدى عينة من التلاميذ :

المطلوب: ايجاد نوع الارتباط وشدته .

x	y	رتب x	رتب y	D^2
1	4	1.5	1	0.25
1	8	1.5	2	0.25
2	10	3	3	0
3	14	4	4	0
4	15	5	5	0
5	16	6	6	0
6	17	7	7	0
7	18	8	8.5	0.25
8	18	9	8.5	0.25
$\sum$				$\sum D^2=1$

$$\rho = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

$$= 1 - \frac{6(1)}{9(9^2-1)}$$

$$\rho = 1 - \frac{6}{720} = 1 - 0.0083 = 0.992$$

اذن معامل ارتباط "سبيرمان" يساوي 0.99 ومنه فان العلاقة بين المتغيرين هي علاقة طردية قوية.

11- معامل الاقتران ومعامل التوافق

يستعمل هذان المعاملان في معرفة درجة العلاقة بين متغيرين من النوع الكيفي ويختلفان في الاستعمال حسب عدد الحالات الشرطية لكل جدول كما سيضح من صيغة وطريقة حساب كل منهما:

1-2- معامل الاقتران :

يستخدم في حالة متغير اسمي لكل منهما تفرعين او حالتين، مثل الجنس ذكور اناث، والموقف من قضية محددة موافق غير موافق. يرمز له بالرمز Ø فاي.

ويحسب بالصيغة التالية :

$$\emptyset = \frac{\boldsymbol{A * D - B * C}}{\sqrt{(\boldsymbol{A + B})(\boldsymbol{C + D})(\boldsymbol{A + C})(\boldsymbol{B + D})}}$$

مثال: الجدول التالي يمثل العلاقة بين الجنس والانتماء السياسي :

	ذكر		**أنثى**		**المجموع**
ديموقراطي	6	A	2	B	8
جمهوري	3	C	4	D	7
المجموع	9		6		15

بتطبيق الصيغة أعلاه نحصل على:

$$\emptyset = \frac{(6 * 4) - (2 * 3)}{\sqrt{(6 + 2)(3 + 4)(6 + 3)(2 + 4)}}$$

$$\frac{18}{\sqrt{8+7+9+6}}$$

$$\frac{18}{54.99}=0.327$$

*ومنه ان العلاقة ضعيفة.

مثال 2: توزيع العينة حسب الجنس والنجاح الدراسي :

الجنس النجاح	ذكر	أنثى	المجموع
راسب	4	4	8
ناجح	6	6	12
المجموع	10	10	20

$$\emptyset=\frac{(\mathbf{4}*\mathbf{6})-(\mathbf{4}*\mathbf{6})}{\sqrt{\mathbf{8}*\mathbf{12}*\mathbf{10}*\mathbf{10}}}$$

$$\frac{\mathbf{0}}{\mathbf{97.99}}=\mathbf{0}$$

*ومنه العلاقة بين المتغيرين معدومة.

13- معامل التوافق :

يستخدم معامل التوافق لحساب قيمة معامل الارتباط عندما يكون المتغيران المراد قياس الارتباط بينهم صفات أيضاً والجدول المزدوج الذي يمثل

العلاقة بينهم يزيد عدد خلاياه عن (4) خلايا دون خلايا المجموع ونستخدم القانون التالي لحساب قيمة معامل التوافق :

صيغة رقم 2: $cc = \sqrt{\frac{\beta-1}{\beta}}$

حيث β يساوي

$$\boldsymbol{\beta = \frac{(n_{11})^2}{n_{.1} * n_{1.}} + \frac{(n_{12})^2}{n_{.2} * n_{1.}} + \cdots + \frac{(n_{pq})^2}{n_{.p} * n_{q.}}}$$

مثال: الجدول التالي يمثل توزيع عينة من الاسر حسب المستوى المعيشي للأسرة وممارسة الزوجة لنشاط مهني :

الحل:

ممارسة نشاط / المستوى المعيشي	تعمل	لا تعمل	المجموع
جيد	50	42	92
متوسط	11	120	131
متدني	1	71	72
المجموع	62	233	295

$$\boldsymbol{\beta} = \frac{(50)^2}{62 * 92} + \frac{(42)^2}{233 * 92} + \frac{(11)^2}{62 * 131} + \frac{(120)^2}{233 * 131} + \frac{(1)^2}{62 * 72} + \frac{(71)^2}{233 * 72}$$

$$= \frac{2500}{5704} + \frac{1764}{21436} + \frac{121}{8122} + \frac{14400}{30523} + \frac{1}{4464} + \frac{5041}{16776}$$

$$= 0.46 + 0.08 + 0.01 + 0.47 + 0.00 + 0.3$$

$$= 1.32$$

ومنه :

$$CC = \sqrt{\frac{1.32 - 1}{1.32}}$$

$$= \sqrt{0.24} = 0.48$$

اذن معامل التوافق يساوي CC=0.48 ومنه فان العلاقة بين عمل الزوجة والمستوى المعيشي للأسرة هي علاقة متوسطة. (Grosjean, Dommergue ,2011).

معامل الارتباط النقطي

يعتبر معامل الارتباط النقطي حالة خصة من معامل الارتباط بيرسون، ويستخدم هذا المعامل لإيجاد العلاقة بين متغيرين أحدهما من النوع الكمي المتصل أو المستمر أي يقاس على الاقل على مقياس الفئوي او النسبي، مثل العلامة، الطول، الوزن ... والمتغير الثاني يكون متغير كيفي ثنائي التقسيم.

اولا: معامل الارتباط النقطي الطبيعي :

يستخدم هذا النوع في حالة كان المتغير الكيفي ذو تقسيم طبيعي لم يتدخل فيه الباحث، كالجنس: ذكور اناث، أو اليوم بين الليل والنهار... ويتم حسابه باستخدام المعادلة التالية:

$$r_p = \frac{\bar{x}_1 - \bar{x}_2}{\sigma}\sqrt{\frac{n_1 * n_2}{N(N-1)}}$$

مثال:

الجدول التالي يمثل العلامات المحصل عليها من طرف عينة الطلبة من الذكور والاناث في مادة الرياضات ويتضح من المثال ان الموضوع قائم على متغيرين المتغير الاول المستقل كيفي مقسم تقسيم طبيعي الى ذكور واناث. والمتغير التابع هو متغير كمي متصل يمثل العلامة.

رقم الطالب	1	2	3	4	5	6	7	8	9	10
x	1	1	0	0	1	0	1	1	1	0
y	8	7	10	9	4	10	6	3	4	9

حيث:

$\bar{x}_1$ المتوسط الحسابي للعلامة لدى الذكور $\bar{x}_1 = \frac{x_1+x_2+x_3+\cdots+x_n}{n_1}$

$\bar{x}_2$ المتوسط الحسابي للعلامة لدى الاناث $\bar{x}_2 = \frac{x_1+x_2+x_3+\cdots+x_n}{n_2}$

σ الانحراف المعياري لمشاهدات المتغير الكمي لكل العلامات المحصل عليها

$$\boldsymbol{\sigma} = \sqrt{\frac{\sum(\boldsymbol{x_i} - \bar{\boldsymbol{x}})^2}{\boldsymbol{n} - \mathbf{1}}}$$

n_1 عدد الذكور في العينة

n_2 عدد الاناث في العينة

N حجم العينة الكلي بما فيهم الذكور والاناث

الحل :

أ-حساب متوسط العلامة لدى الذكور

$$\bar{x}_1 = \frac{8+7+4+6+3+4}{6} = \frac{38}{6} = 5.33$$

ب-حساب متوسط العلامة لدى الاناث :

$$\bar{x}_2 = \frac{10+9+10+9}{4} = \frac{38}{4} = 9.5$$

ج-حساب المتوسط والانحراف المعياري لكل العينة (ذكور+ ناث):

قبل حساب هذا الانحراف المعياري لابد من حساب المتوسط الحسابي للعلامة في كل العينة **(ذكور+ ناث):

$$\bar{x}_N = \frac{8+7+10+9+4+10+6+3+4+9}{10} = \frac{70}{10} = 7$$

1- **حساب الانحراف المعياري**

$$\sigma_N = \sqrt{\frac{\sum 62^2}{10-1}} = \sqrt{6.89}$$

=2.624

2- $N = 10$ حجم العينة الكلي بما فيهم الذكور والاناث

3- $n_1 = 6$ عدد الذكور

4- $n_2 = 4$ عدد الاناث

د- حساب قيمة معامل الطبيعي تطبيق الصيغة :

$$r_p = \frac{\bar{x}_1 - \bar{x}_2}{\sigma} \sqrt{\frac{n_1 * n_2}{N(N-1)}}$$

$$r_p = \frac{5.33 - 9.5}{2.624} \sqrt{\frac{6 * 4}{10(10-1)}}$$

$$r_p = \frac{-4.17}{2.624}\sqrt{\frac{24}{90}}$$

=-1589*0.516

= -0.819

ومنه فان العلاقة بين جنس الطالب والعلامة في الرياضيات هي علاقة قوية في هذه العينة (-0.819)

ملاحظة: اشارة نتيجة معامل الارتباط النقطي في حالة المتغيرات الكيفية كما في هذه الحالة، ليس لها أي معنى ويتم الاكتفاء في تفسير النتيجة بناء على شدة العلاقة وليس اتجاهها.

ثانيا- معامل الارتباط النقطي الاصيل:

يعتبر معامل الارتباط النقطي حالة خصة من معامل الارتباط بيرسون، ويستخدم هذا المعامل لإيجاد العلاقة بين متغيرين أحدهما من النوع الكمي المتصل أو المستمر أي يقاس على الاقل على مقياس الفئوي او النسبي، مثل العلامة، السن... والمتغير الثاني يكون متغير كيفي ثنائي التقسيم لكن هذا التقسيم غير طبيعي إنما قام به الباحث، على شرط ان يكون ذلك التقسيم قد قام به الباحث قبل عملية القياس، كأن يحدد الاجابة على أحد فقرات المقياس بنعم او لا، أو يقوم بتقسيم متغير حضور المحاضرة مثلا الى حاضر وغائب.. ويتم حسابه باستخدام المعادلة التالية:

$$\boldsymbol{r_{pb} = \frac{\overline{x}_1 - \overline{x}_2}{\sigma}\sqrt{pq}}$$

حيث:

- $\bar{x}_1$: هو المتوسط الحسابي للمجموعة الاولى
- $\bar{x}_2$: هو المتوسط الحسابي للمجموعة الثانية
- σ: هو الانحراف المعياري لبيانات المجموعتين
- p: هي نسبة افراد المجموعة الاولى على العدد الكلي
- q: هي نسبة أفراد المجموعة الثانية على العدد الكلي

تفسير النتيجة: تفسر النتيجة المحصل عليها من هذا المعامل بنفس الطريقة التي تفسر بها نتيجة معامل بيرسون حيث تتراوح قيمتها بين +1،0،-1 حيث تضعف العلاقة كلما اقتربت قيمته من الصفر وتنعدم في حالة الصفر وتكون العلاقة قوية كلما اقتربت من الواحد وقد تكون موجبة او سلبية وفقا للإشارة.

مثال:

اراد أحد الباحثين أن يقيس العلاقة بين متغير قلق الامتحان والتحصيل في مادة الرياضيات وقام بتقسيم متغير قلق الامتحان الى حالتين قلق وغير قلق بناء العلامات التي يتحصل عليها الطالب من مقياس قلق الامتحان وجعل العلامة 50 هي الفاصل، أي يصنف الطالب الى قلق اذا تحصل على 50 علامة فأكثر على مقياس قلق الامتحان، ويصنف على انه غير قلق اذا تحصل على علامة أقل من 50 على نفس المقياس. في حين أجرى امتحان شامل في مادة الرياضات، فجاءت البيانات كالتالي:

x_i	علامة الطالب في	علامة الطالب على مقياس قلق	رقم

الطالب	الامتحان		امتحان الرياضيات	$\bar{x} -$
	العلامة	التصنيف		
1	82	قلق	8	
2	63	قلق	6	
3	27	غير قلق	7	
4	14	غير قلق	3	
5	93	قلق	10	
6	82	قلق	6	
7	45	غير قلق	5	
8	67	قلق	4	
9	82	قلق	3	
10	73	قلق	6	

الحل:

1- اعادة تصنيف الطلبة الى مجموعتين وفق المتغير المستقل وهو حالة الطالب وفق متغير قلق الامتحان بناء على المعلومات الميدانية:

تحصيل فئة الطلبة غير القلقين			تحصيل فئة الطلبة القلقين		
العلامة	حالته	رقم الطالب	العلامة	حالته	رقم الطالب
7	غير قلق	3	8	قلق	1
3	غير قلق	4	6	قلق	2
5	غير قلق	7	10	قلق	5
			6	قلق	6
			4	قلق	8
			3	قلق	9
			6	قلق	10

2- **حساب المتوسط الحسابي للتحصيل في مادة الرياضيات للطلبة الذين لديهم قلق الامتحان :**

$$\bar{x}_1 = \frac{\sum x_{i1}}{n_1} = \frac{8+6+10+6+4+3+6}{7} = \frac{43}{7} = 6.14$$

ومنه متوسط علامة امتحان الرياضيات بالنسبة للطلبة الذين لديهم قلق الامتحان هو 6.14

3- **حساب المتوسط الحسابي للتحصيل في مادة الرياضيات للطلبة الذين لديهم قلق الامتحان:**

$$\bar{x}_2 = \frac{\sum x_{i2}}{n_2} = \frac{7+3+5}{3} = \frac{14}{3} = 5$$

ومنه متوسط علامة امتحان الرياضيات بالنسبة للطلبة الذين ليس لديهم قلق الامتحان هو 4.67

4- **حساب الانحراف المعياري لمتغير التحصيل بالنسبة لكل الطلبة القلقين وغير القلقين:**

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{43.6}{9}}$$

=2.2

5- **استخراج « p » وهي نسبة مجموعة الطلبة القلقين من المجموع الكلي**

$$p = \frac{n_1}{n} = \frac{7}{10} = 0.7$$

6- **استخراج « q » وهي نسبة مجموعة الطلبة غير القلقين من المجموع الكلي**

$$q = \frac{n_2}{n} = \frac{3}{10} = 0.3$$

7- بعد الحصول على كل القيم نعوض في الصيغة الاولى لايجاد قيمة معامل الارتباط الثنائي كالتالي:

$$r_{pb} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma}\sqrt{pq}$$

$$r_{pb} = \frac{6.14 - 5}{2.2}\sqrt{0.7 * 0.3}$$

$$= \frac{1.14}{2.2}(0.46)$$

$$0.52 * 0.46$$

$$= 0.24$$

وهي علاقة ضعيفة بين قلق الامتحان ومستوى تحصيل الطالب في مادة الرياضيات.

كما يستخدم هذا المعامل في الصدق التمييزي للقياس عندما يكون الاجابة على فقرات المقياس بنعم أو نعم، (المنيزلي و غرايبة، 2006، صفحة 135)

فحص الدلالة الاحصائية لمعامل الارتباط النقطي:

ان الفرضية الصفرية التي تفحص في حالة معامل الارتباط النقطي تشير الى أن قيمة معامل الارتباط تساوي صفر: H_o p=0، اما الفرضية البديلة فتشير الى أن معامل الارتباط يتخلف عن الصفر: H_a p≠0.

و بما أن معامل الارتباط النقطي حالة خاصة من معامل الارتباط بيرسون فانه يمكن اتباع نفس الطريقة المتبعة لحساب الدلالة الاحصائية لهذا الاخير، حيث يستخدم اختبار « t » في حالة العينات الصغيرة الحجم باستخدام المعادلة التالية:

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

أو

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

وبالنسبة لدرجات الحرية فانها تساوي n-2، وباتباع خطوات اختبار الدلالة الاحصائية لمعامل الارتباط كما سبقت الاشارة اليه في معامل بيرسون وتطبيق هذه المعادلة على نتائج المثال السابق نحصل على مايلي:

1- الفرضية الصفرية والفرضية البديلة في اختبار ذو طرفين او حدين :

أ- الفرضية الصفرية $H_o=0$، لا توجد علاقة بين المتغيرين y , x.

ب- الفرضية البديل $H_1\neq 0$ توجد علاقة بين المتغيرين y , x.

2- تحديد مستوى الدلالة alpha وحدد هنا بـ α= 0,05

3- حساب القيمة الحقيقية لمعامل الارتباط وفق الصيغ « r ».، وتم تقديرها في المثال السابق بـ r = 0.24

4- ايجاد قيمة دالة الاختبار، هنا في حالة عينة أقل من 30 وحدة احصائية: والتي يمكن ايجادها بتطبيق اختبار « t » كما يلي:

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

$$t = \frac{0.24}{\sqrt{\frac{1 - (0.24^2)}{10 - 2}}}$$

=

$$t = \frac{0.24}{\sqrt{\frac{1 - 0.0576}{8}}}$$

$$t = \frac{0.24}{0.343}$$

= 0.699

5- حساب درجة الحرية α=n-2 = 10-2 = 8

6- استخراج القيمة المجدولة من الجدول "t" المقابلة لدرجة الحرية 8 ومستوى الدلالة 0.05 وهي: 2.306

7- مقارنة القيمة المحسوبة لـ "t" مع القيمة المجدولة نجد: المحسوبة 0.699 أصغر من القيمة المجدولة 2.306 وبالتالي نقبل الفرض الصفري ونرفض الفرض البديل بمعنى انه لا توجد علاقة ذات دلالة احصائية بين قلق الامتحان والعلامة المحصل عليها في امتحان الرياضيات.

ثالثا- مربع معامل الارتباط النقطي: هل يمكن تسميته معامل التحديد لمعامل بايسيريال ؟؟

يمكن ايجاد مربع هذا المعامل من أجل معرفة النسبة التي يؤثر بها المتغير المستقل في المتغير التابع باستخدام الصيغة التالية :

$$r_{pb}^2 = \frac{t^2}{t^2 + df}$$

أي ان:

$$r_{pb} = \sqrt{\frac{t^2}{t^2 + df}}$$

حيث:

- t^2 هو مربع قيمة اختبار t.test وقد في هذا المثال بـ 0.699
- df هي درجة الحرية (n-2) أي 10-2 في حالة المثال السابق.

وبالتعويض في الصيغة السابقة نجد مايلي:

$$r_{pb}^2 = \frac{t^2}{t^2 + df}$$

$$r_{pb}^2 = \frac{0.699^2}{0.699^2 + 8}$$

$$= \frac{0.4886}{0.4886 + 8}$$

$$= \frac{0.4886}{8.4886} = 0.0575$$

وهي قيمة مربع معامل بايسيريل النقطي: والتي من خلالها يمكن معرفة النسبة التي يفسرها المتغير المستقل (أي قلق الامتحان) من المتغير التابع (أي العلامة المحصل عليها في الرياضيات) بعد ضرب هذه النتيجة في مائة لتصبح:
%5.75= 100*0.0575

ومنه فان قلق الامتحان لا يفسر سوى 5.75% من العلامة المحصل عليها في امتحان الرياضيات في العينة المبحوثة.

كما انه بأخذ الجذر التربيعي لمربع هذا المعامل نحصل على القيمة الاصلية لمعمل الارتباط بايسريال كما تشير الصيغة أعلاه:

$$r_{pb} = \sqrt{\frac{t^2}{t^2 + df}}$$

$$r_{pb}^2 = \sqrt{\frac{t^2}{t^2 + df}}$$

$$= \sqrt{\frac{0.4886}{8.4886}} = \sqrt{0.0575} = 0.2397 \sim 24$$

وهي نفسها قيمة معامل الارتباط الاصلي. (المنيزلي و غرايبة، 2006، صفحة 137)

5- معامل لامبدا λ

يقيس معامل لامبدا الذي يشير رمزه الى الحرف الاغرقي Lambda λ، والمعروف أيضا بمعامل التنبؤ لـ "جوتمان" Guttmannقوة العلاقة بين متغيرين من النوع الاسمي وذلك بتخفيض نسبة الخطأ في التنبؤ عندما يستخدم المتغير المستقل للتنبؤ أو لتفسير المتغير التابع، او قيمة الخطأ الذي تم تخفيضه عند استخدام قيمة (س) للتنبؤ بقيمة (ص)، حيث يتوقف تمييز المتغير المستقل عن المتغير التابع بناء على مشكلة البحث. ويمكن ان يعطى في شكل نسبة مئوية، تساعد على معرفة الدرجة التي يمكن بها تقدير المتغير التابع من المتغير المستقل.

تنحصر قيمت هذا المعامل بين 0 – 1 وتشير القيمة 1 الى أن المتغير المفسِّر (المستقل) يتنبأ جيدا بحالات المتغير المفسَّر (التابع). واذا قدرت نسبته ب 0.5 فانه يشير الى ان خطأ التنبؤ بين المتغيرين خفّض الى 50% ويمكن للمتغير المستقل التنبؤ بالمتغير التابع على نحو متوسط، أما اذا قدرت قيمته بـ 0 فإنها تشير الى أن معرفة المتغير المستقل لن تساعد على التنبؤ بحالات المتغير التابع.

يفضل استخدام معامل لمدا المتماثل عوضا عن استخدام(χ^2، كا^2) في حالة وجود بعض تكرارات خلايا الجدول بها أصفار. (محمد شامل بهاء الدين، 2005، صفحة 569)

يحسب باستخدام الصيغة الاحصائية التالية :

$$\lambda = \frac{\varepsilon_1 - \varepsilon_1}{\varepsilon_1}$$

حيث:

- ε_1: تشير الى الخطأ الذي تم الوقوع فيه دون الاعتبار للمتغير المستقل ويمكن حسابه من الصيغة التالية :

$$\varepsilon_1 = N_{total} - N_{Mode}$$

- ε_2: تشير الى الخطأ الذي تم الوقوع فيه بالنسبة لكل حالة أو فئة من فئات المتغير المستقل، وهو عبار عن حاصل جمع الاخطاء الخاصة بكل فئة او حالة من حالات هذا المتغير ويمكن حسابها باستخدام الصيغة التالية:

$$\varepsilon_2 = (N_1 - N_{mode\ de\ 1}) + (N_2 - N_{mode\ de\ 2}) + \cdots + (N_k - N_{mode\ de\ k})$$

مثال: تمثل بيانات الجدول التالي المستوى المعيشي للأسرة والتحصيل العلمي للطالب:

المستوى التحصيل	جيد	ميسور	معوز	المجموع
عالي	189	53	4	246
متوسط	143	102	9	254
ضعيف	7	9	4	20
المجموع	339	164	17	520

لحساب معامل لامبد نتبع الخطوات التالية:

اولا: حساب الخطأ الاول:

$$\varepsilon_1 = N_{total} - N_{Mode}$$

$$\varepsilon_1 = 520 - 254 = 266$$

من هذه القيمة يمكن القول انه تم تصنيف 266 طالب حسب مستوى التحصيل بشكل خاطئ دون اعتبار المتغير المستقل.

ثانيا: حساب الخطأ الثاني :

$$\varepsilon_2 = (N_1 - N_{mode\ de\ 1}) + (N_2 - N_{mode\ de\ 2}) + \cdots + (N_k - N_{mode\ de\ k})$$

$$\varepsilon_2 = (339 - 189) + (164 - 102) + (17 - 9) = 220$$

من قيمة الخطأ الثاني يمكن القول أنه قد تم تصنيف 220 طالب بشكل خاطئ حسب المستوى المعيشي.

ثالثا: حساب قيم لامبدا باستخدام الصيغة الاحصائية الاولى:

$$\lambda = \frac{\varepsilon_1 - \varepsilon_1}{\varepsilon_1}$$

$$= \frac{266 - 220}{266} = 0.17$$

ومنه فان النسبة الخطأ المخفضة لتنبؤ متغير المستوى المعيشي للأسرة بمتغير التحصيل الدراسي هي 17% وهي نسبة ضعيفة تدل على ضعف العلاقة بين المتغيرين، أي يتنبا بدرجة ضعيفة .

اختبار الدلالة الاحصائية لمعامل لمبدا:

لمعرفة الدلالة الاحصائية لمعامل لمبدا يجب الرجوع الى معامل كا^2، فاذا كان هذا الاخير دال احصائيا فان ذلك يعني أن معامل لبدا له دلالة احصائية ايضا، والعكس صحيح. (محمد شامل بهاء الدين، 2005، صفحة 570) .

6- معامل ايتا

هو مقياس يستخدم لتفسير قوة العلاقة بين بن المتغير المستقل والمتغير التابع يعتمد على الكشف عن مقدار التباين في القيم المتغير التابع التي تعزى الى المتغير المستقل وهو ما يعرف بـ (**Ꞷ**^{2}) وتقرأ (Oméga تربيع) . وبعد اخذ الجذر التربيعي لقيمة هذا المعامل نحصل على ما يسمى نسبة الارتباط ويرمز لها بالرمز (**η**) وتقرأ (Éta) وهو يستخدم لقياس قوة الترابط بين متغيرين، بمعنى ما اذا كانت العلاقة قوية (أكبر من 0.60) او ضعيفة (أقل من 0.50) أم متوسطة (من 0.50 الى 0.60). (محمد شامل بهاء الدين، 2005، صفحة 576)

يستخدم معامل إيتا لمعرفة نسبة الارتباط في حالة العلاقة غير الخطية بين المتغيرين المدروسين، على عكس المعامل ارتباط بيرسون والمعاملات الاخرى المشتقة منه. اذ أن استعمال هذه المعاملات الأخيرة في تحليل بيانات بعلاقة غير الخطية والذي يمكن تأكيده من خلال رسم الشكل الانتشاري لسحابة النقاط بين المتغير المستقل والمتغير التابع، لان تلك المقاييس تصل الى نتائج تكون دائما تقديرا يقل بكثير من درجة الارتباط الحقيقية بين المتغيرين، والعكس اذا كان الارتباط الحقيقي بين المتغيرين قويا فان تقدير العلاقة بينهما باستخدام معاملات الارتباط الخطي قد يؤدي الى ارتباط صفري او قريب من الصفر، ويصبح المعامل الانسب

في هذه الحالة هو نسبة الارتباط أو معامل ايتا.(الاحصاء في علم النفس، ص262)

كما يستخدم في حالة كان المتغير التابع من النوع الكمي والمتغير المستقل من النوع الاسمي او الرتبي وكما سبقت الاشارة فان هذا المعامل يشير الى نسبة التباين الكلي في المتغير التابع والتي تعزى الى المتغير المستقل. (المنيزلي و غرايبة، 2006، صفحة 165). ويمكن حسابه باستخدام الصيغة التالية :

الصيغة رقم (1)	$\eta = \sqrt{\dfrac{SS_B}{SS_T}}$

حيث أن:

- SS_T هي مجموع مربع انحرافات كل قيمة من القيم الخاصة بالمتغير التابع عن الوسط الحسابي الكلي لهذا المتغير. ويمكن ايجادها كالتالي:

- **الصيغة رقم (02)**	$SS_T = \sum (x_i - \bar{\bar{x}})^2$

حيث :

- $\bar{\bar{x}}$ هو المتوسط الحسابي الكلي للمتغير التابع (يشمل كل المشاهدات الخاصة به).
- x_i تمثل كل مشاهدة أو قيمة من قيم المتغير التابع.

- SS_B هي الفرق بين SS_T وبين SS_W ويمكن ايجاده من الصيغة التالية:

الصيغة رقم (03)	$SS_B = SS_T - SS_W$

- حيث: SS_W هي مجموع مربع انحرافات قيم كل فئة catégorie عن الوسط الحسابي لنفس الفئة، ويمكن حسابه بالنسبة لكل فئة من فئات المتغير المستقل كالتالي بعد حساب المتوسط الحسابي لكل فئة على حدة:

الصيغة رقم (04)

$$\begin{aligned} SS_W = \sum(x_1-\bar{x}_{c1})^2+(x_2-\bar{x}_{c1})^2+\cdots(x_n-\bar{x}_{c1})^2 \\ +\sum(x_1-\bar{x}_{c2})^2+(x_2-\bar{x}_{c2})^2+\cdots(x_n-\bar{x}_{c1})^2 \\ +\cdots\sum(x_1-\bar{x}_{c3})^2+(x_2-\bar{x}_{c3})^2 \\ +\cdots(x_n-\bar{x}_{ck})^2 \end{aligned}$$

- حيث: c تمثل كل فئة من فئات المتغير المستقل والتي قد تكون فئتان فأكثر.

مثال: الجدول التالي يمثل عينة من الطلبة موزعين حسب المستوى الاقتصادي والتحصيل الدراسي :

			العمود1	العمود2	العمود 3
رقم الطالب	المستوى الاقتصادي	التحصيل	$\bar{x}_{ci}$	$(x_1 - \bar{x}_{ci})^2$	$(x_i - \bar{\bar{x}})^2$
1	ضعيف جدا	4	5	$(4-5)^2=1$	4-9.44=29.59
2	ضعيف جدا	6	5	$(6-5)^2=1$	6-9.44=11.83
3	ضعيف	8	10	$(8-10)^2=4$	8-9.44=2.07
4	ضعيف	12	10	$(12-10)^2=4$	12-9.44=6.55
5	متوسط	15	13.33	$(15-13.33)^2=2.78$	15-9.44=30.91
6	متوسط	15	13.33	$(15-13.33)^2=2.78$	15-9.44=30.91
7	متوسط	10	13.33	$(15-13.33)^2=11.09$	10-9.44=0.31
8	حسن	7	6.33	$(7-6.33)^2=0.45$	7-9.44=5.95
9	حسن	7	6.33	$(7-6.33)^2=0.45$	7-9.44=5.95
10	حسن	5	6.33	$(5-6.33)^2=1.77$	5-9.44=19.71
1	جيد	6	7.67	$(6-7.67)^2=2.78$	6-9.44=11.83
12	جيد	8	7.67	$(8-7.67)^2=2.11$	8-9.44=2.07
13	جيد	9	7.67	$(9-7.67)^2=1.77$	9-9.44=0.19
14	جيد جدا	12	13	$(12-13)^2=1$	12-9.44=6.55
15	جيد جدا	14	13	$(14-13)^2=1$	14-9.44=20.79
16	جيد جدا	13	13	$(13-13)^2=0$	13-9.44=12.67
المجموع		**151**		**35.98**	**197.88**

خطوات الحل :

1-حساب المتوسط الخاص بكل فئة كما يتضح في العمود الاول.

2- حساب SS_W كما يتضح من العمود الثاني وتطبيق الصيغة رقم (04) كما يتضح أعلاه.

3- حساب $_{T}$ بتطبيق الصيغة رقم (02) كما يتضح أعلاه.، كما يتضح من العمود الثالث، بعد حساب المتوسط الحسابي الكلي لمتغير التحصيل الدراسي كما يلي:

$$\bar{\bar{x}} = \frac{\sum x_i}{N} = \frac{151}{16} = 9.44$$

4- حساب $_{B}$ بتطبيق الصيغة رقم (03) كما يلي :

$$SS_B = SS_T - SS_W = 197.88 - 35.98$$

$$SS_B = 161.9$$

5- حساب قيمة معامل ايتا بتعويض القيم في الصيغة رقم (01) كما يلي :

$$\eta = \sqrt{\frac{SS_B}{SS_T}} = \sqrt{\frac{161.9}{197.88}}$$

$\sqrt{0.82}$=0.9055~ 0.91

ومنه فان العلاقة بين المستوى الاقتصادي لأسرة الطالب وتحصيله الدراسي هي علاقة قوية جدا. لا يستدعي هذا المعامل اختبار الدلالة الاحصائية لأنها لا تختلف عن الدلالة الاحصائية لقيمF التي تم الحصول عليها باختبار التباين.

الفصل التاسع: الانحدار الخطي البسيط

1- الانحدار الخطي البسيط

2- معامل التحديد

تمهيد

راينا في السابق مفهوم الارتباط والغاية من استخدامه في حالات مختلفة وقبل التطرق الى الانحدار لابد من التنويه الى الفرق بين دراسة الارتباط ودراسة الانحدار، حيث اذا كان هدف الباحث هو التنبؤ بالمتغير المتغير التابع (y) بناء على المعلومات المتوفرة في المتغير المستقل (x) نكون في هذه الحالة أمام ما يعرف بدراسة الانحدار Coefficient de régression ولكن اذا ان الباحث مهتم فقط بالحصول على تعبير احصائي عن درجة العلاقة بين المتغيرين محل الدراسة، فانه في هذه الحالة يكون امام مايعرف بدراسة الارتباط Coefficient de corrélation.

1-الانحدار الخطي البسيط

احصائيا يعود استخدام لفظ "الانحدار" الى عم 1885 عدما استخدمها "فرانسيس غالتون" « Galton »في مقاله حول نتائج دراسته للعلاقة بين أطوال الآباء وأطوال الابناء، والتي خلص فيها الى نتيجة مفادها أن قيم أطوال الابناء تنحدر نحو موضع ما يقع ما بين أطوال آبائهم والقيمة المتوسطة (المتوسط) للمجتمع الاصلي، ولقد استفاد بهذه النتيجة "كارل بيرسون" فيما بعد وطوره الى ما يسمى بمعامل الانحدار الخطي.

فاذا كان معامل الارتباط يقدم تلخيصا واضحا للعلاقة بين x, y كتغيرين فان معامل الانحدار يعبر عن التغير المتوقع في المتغير (y) كلما تغيرت قيم المتغير (x) المناظرة على أساس أنه متغير مستقل.

من ثمة يتحدد الهدف الاساسي لمعادلة خط الانحدار في قياس تأثير المتغير (x) على المتغير (y) ووضع العلاقة في شكل معادلة خطية بحيث تصاغ في شكل

معادلة من الدرجة الاولى على الصورة التالية (Grosjean, Dommergues, 2011):

4-2-معادلة خط الانحدار (x) عن (y) **x=c+dy**	**4-1-معادلة خط الانحدار (y) عن (x)** **y=a+bx**
نحتاج هنا لإيجاد قيمة كل من (c) و(d)	نحتاج هنا لإيجاد قيمة كل من (a) و(b)
1- يمكن الحصول على قيمة (d) من الصيغة التالية: $\boxed{\mathbf{d} = \mathbf{r}\left(\frac{\sigma_x}{\sigma_y}\right)}$ 2- أما قيمة (c) فيمكن الحصول عليها من الصيغة التالية: $\boxed{c = \bar{x} - d\bar{y}}$	1- يمكن الحصول على قيمة (b) من الصيغة التالية: $\boxed{\mathbf{b} = \mathbf{r}\left(\frac{\sigma_y}{\sigma_x}\right)}$ 2- أما قيمة (a) فيمكن الحصول عليها من الصيغة التالية: $\boxed{a = \bar{y} - b\bar{x}}$

الصيغة رقم 1: $\boxed{Y = a + bx}$

حيث (a) و(b) قيم يمكن حسابها من واقع البيانات التجريبية وذلك بتطبيق الصيغ التالية:

الصيغة رقم 2:

$$\mathbf{b} = \mathbf{r}\left(\frac{\boldsymbol{\sigma_y}}{\boldsymbol{\sigma_x}}\right)$$

أما قيمة (a) فيمكن الحصول عليها من المعادلة التالية:

الصيغة رقم 3:

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

مثال: البيانات التالية تمثل العلامات المحصل عليها من طرف عينة من الطلبة في كل من مادة الفلسفة ومادة الرياضيات المطلوب :

1- رسم سحابة النقاط وتحديد نوع العلاقة من خلالها.

2- المطلوب ايجاد معادلة خط الانحدار (y) عن (x).

3- ايجاد <u>وزن</u> الطالب المتوقع اذا كان طوله يساوي 150سم

4- ايجاد معادلة خط الانحدار (x) عن (y).

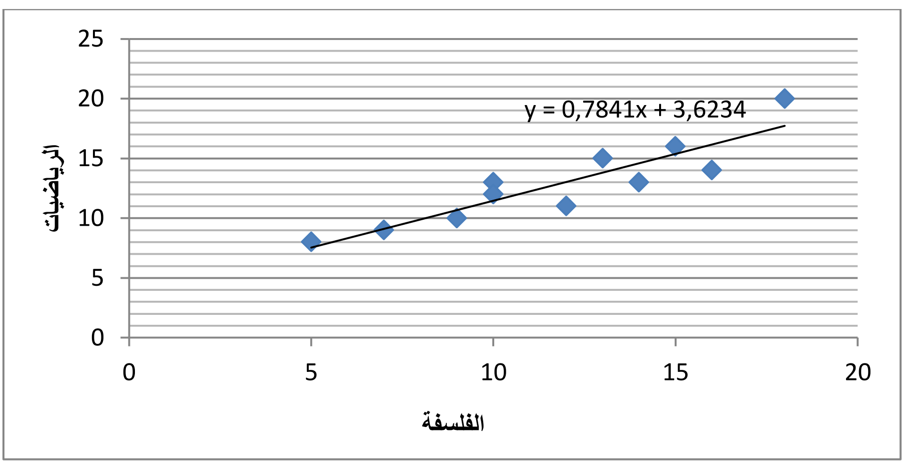
5- ايجاد <u>طول</u> الطالب المتوقع اذا كان وزنه 55كلغ.

			العمود (1)	**العمود (2)**	**العمود (3)**	**العمود (4)**	**العمود (5)**	**العمود (6)**
رقم الطالب	**لفلسفة**	**الرياضيات**	$\boldsymbol{x_i * y_i}$	$\boldsymbol{x_i - \overline{x}}$	$\boldsymbol{(y_i - \overline{y})}$	$\boldsymbol{(y_i - \overline{y})*(x_i - \overline{x})}$	$\boldsymbol{(x_i - \overline{x})^2}$	$\boldsymbol{(y_i - \overline{y})^2}$
1	15	16	240	3,27	3,18	10,40	10,69	10,11
2	10	13	130	-1,73	0,18	0,31-	2,99	0,03
3	12	11	132	0,27	-1,82	0,49-	0,07	3,31

0,67	2,99	1,42	-0,82	-1,73	120	12	10	**4**
1,39	18,23	5,04	1,18	4,27	224	14	16	**5**
7,95	7,45	7,70	-2,82	-2,73	90	10	9	**6**
0,03	5,15	0,41	0,18	2,27	182	13	14	**7**
14,59	22,37	18,07	-3,82	-4,73	63	9	7	**8**
23,23	45,29	32,44	-4,82	-6,73	40	8	5	**9**
4,75	1,61	2,77	2,18	1,27	195	15	13	**10**
51,55	39,31	45,02	7,18	6,27	360	20	18	**11**
117,61	156,15	122.47			1776	141	129	$\sum$
$\sigma_y = 3.42$	$\sigma_x = 3.95$					$\bar{y}$=12.82	$\bar{x}$ =11.73	الحسابي

الحل :

أولا: تحديد الشكل الانتشاري لسحابة النقاط (x,y)



يتبين من الشكل أن سحابة النقاط تأخذ مسارا خطيا متزايدا، وبناء على هذا المسار يمكن تقريب العلاقة بين المتغيرين (x)، (y) الى علاقة خطية طردية.

ثانيا :ايجاد معادلة خط الانحدار (y) عن (x): y=a+b(x)

1-2- حساب قيمة (b) من هذه القيم الثلاث يمكن حسابها كمايلي:

$$\mathbf{b} = \mathbf{r}\left(\frac{\boldsymbol{\sigma_y}}{\boldsymbol{\sigma_x}}\right)$$

$$b = 0.92\left(\frac{3.42}{3.92}\right)$$

$$= 0.79$$

ومنه فان: b=0.79

ولتحقيق متطلبات هذه الصيغة يجب حساب القيم التالية :

أ- المتوسط الحسابي لقيم (x) :

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

$$= \frac{129}{11}$$

=11.73

و منه فالمتوسط الحسابي لقيم (x) هو: 11.73

ب- المتوسط الحسابي لقيم (y) :

$$\bar{y} = \frac{\sum_{n}^{i=1} y_i}{n}$$

$$= \frac{141}{11}$$

=12.82

ومنه فالمتوسط الحسابي لقيم (y) هو 12.82

ج- الانحراف المعياري للمتغير(x) من العمود رقم05 في الجدول أعلاه :

$$(\sigma_x) = \sqrt{\frac{(\boldsymbol{x_i} - \boldsymbol{\overline{x}})^2}{n-1}} = \sqrt{\frac{156{,}18}{10}}$$

$\sqrt{15.61}$=3.95

و منه فان الانحراف المعياري لقيم (x) يساوي: $\boldsymbol{\sigma_x = 3.95}$

د- حسابالانحراف المعياري للمتغير(y) من العمود رقم06 في الجدول أعلاه :

$$\left(\sigma_y\right) = \sqrt{\frac{(\boldsymbol{y_i} - \boldsymbol{\overline{y}})^2}{\boldsymbol{n-1}}} = \sqrt{\frac{117}{10}}$$

$\sqrt{11.76}$=3.42

و منه فان الانحراف المعياري لقيم (x) يساوي: $\boldsymbol{\sigma_y = 3.42}$

هـ-حساب قيمة معامل الارتباط بيرسون:

$$\boldsymbol{r = \frac{Cov(x, y)}{(\sigma_x)\left(\sigma_y\right)}}$$

ه-1-حساب قيمة cov:

$$COV = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$

$$COV = \frac{122.45}{10} = 12.25$$

ه-2-حساب قيمة معامل بيرسون r من الصيغة أعلاه نجد:

$$r = \frac{12.25}{(3.95)(3.42)}$$

$$\frac{12.45}{13.51} = 0.90$$

ومنه فان العلاقة قوية جدا بين علامة الطالب فب مادة الفلسفة وعلامته في مادة الرياضيات.

2-2- حساب قيمة (b) من هذه القيم الثلاث يمكن حسابها كمايلي:

$$\mathbf{b} = \mathbf{r}\left(\frac{\boldsymbol{\sigma_y}}{\boldsymbol{\sigma_x}}\right)$$

$$b = 0.90\left(\frac{3.42}{3.92}\right)$$

$$= 0.87 * 0.90 = 0.0.78$$

ومنه فان: b=0.78

3-2- حساب قيمة (a) :

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

$$= 12.82 - (0.78)(11.73)$$

$$= 12.82 - 9.14$$

ومنه فان: a=3.68

3- تحديد معادلة خط الانحدار (y) عن (x) بناء على المعطيات السابقة:

y=a+b(x)

$$y_{(x)} = (3.68) + (0.78)(x)$$

ومن هذه المعادلة يمكن معرفة قيمة (y) اذا ما عرفنا (x) والاجابة على السؤال رقم 3.

- ما ماهي العلامة المتوقعة للطالب في مادة الرياضيات اذا تحصل في مادة الفلسفة على العلامة 10 ؟

من خلال التعويض في المعادلة السابقة y=(3.68)+(0.78)(x) نجد:

y=(3.68)+(0.78) (10)

=11.48

ومنه فان الطالب الذي تحصل على 10 في مادة الفلسفة يتوقع أن يتحصل على العلامة 11.48 في مادة الرياضيات.

ثالثا: ايجاد معادلة خط الانحدار (x) عن (y): x=c+dy

من العمليات الحسابية السابقة تم تحديد قيم المتوسط الحسابي والانحراف المعياري ومعامل الارتباط لكل من (x) و(y) كما يلي :

- المتوسط الحسابي (x) هو: $\overline{x} = \mathbf{11.73}$

- المتوسط الحسابي (y) هو: $\bar{y} = 12.82$
- الانحراف المعياري لقيم (x) يساوي: $\sigma_x = 3.95$
- الانحراف المعياري لقيم (x) يساوي: $\sigma_y = 3.42$
- قيمة معامل الارتباط: **r=0.90**

3-1- حساب قيمة (d) من هذه القيم الثلاث يمكن حسابها كمايلي:

$$\mathbf{d} = \mathbf{r}\left(\frac{\boldsymbol{\sigma_x}}{\boldsymbol{\sigma_y}}\right)$$

$$d = 0.90\left(\frac{3.95}{3.42}\right)$$

$$= 0.90(1.15) = 1.03$$

ومنه فان: d=1.03

3-2- حساب قيمة (c) :

$$\boldsymbol{c = \bar{y} - b\bar{x}}$$

$$= 3.42 - (1.03)(3.95)$$

$$= 3.42 - 4.07$$

$$= -0.65$$

3-3- تحديد معادلة خط الانحدار (x) عن (y) بناء على المعطيات السابقة:

$X_{(y)}$ =(-0.65)+(1.03)(y)

و منه يمكن الاجابة على السؤال الرابع :

** الاجابة على السؤال الرابع: ما هي علامة الطالب المتوقعة في مادة الفلسفة اذا تحصل في مادة الرياضيات على العلامة 14 ؟

بالتعويض في المعادلة الأخيرة نجد:

$$X_{(y)} =(-0.65)+(1.03)(y)$$

$$X_{(y)} =(-0.65)+(1.03)(\mathbf{14})$$

$$=(-0.65)+14.42$$

$$=13.77$$

ومنه فان الطالب الذي تحصل على العلامة 14 في الرياضيات يتوقع ان يحصل على العلامة13.77 في مادة الفلسفة.

2- معامل التحديد

يمكن الاستفادة من معامل الارتباط في التقدم خطوة إلى الأمام نحو تفسير العلاقة بين المتغيرين محل الدراسة (لكن ليس بمفهوم السببية). ومعامل التحديد هو مربع معامل الارتباط بارسون، وهو مقياس لمقدار التباين لمتغير ما الذي يفسره متغير آخر، مثل علاقة درجة الاهتمام بالمادة بالأداء في الامتحان.

اذ يتغير أداء الإمتحان من شخص إلى آخر لعدة أسباب (الاختلاف في القدرات، مستويات التحضير، درجة الاهتمام...) فإذا كانت العلاقة بين درجة الاهتمام بالمادة ومستوى الأداء فيه يقدر بـ($r= 0.441-$) إذن فمعامل التحديد هو $r^2=(-0.441)^2=0.194$ وتخبرنا هذه القيمة بمقدار التباين في مستوى الأداء في الامتحان الذي يمكن تفسيره في درجة الاهتمام بالمادة. وإذا حوّلنا هذه القيمة إلى نسبة مئوية بضربها في 100 فإنه يمكن القول أن درجة الاهتمام بالمادة يعتبر

مسؤولا عن 19.4% من التباين أو التغير في أداء الامتحان مقابل ذلك يمكن القول أن حوالي 80% المتبقية من التباين او التغاير تكون نتيجة متغيرات أخرى.

بمعنى أن التباين في أحد المتغيرات السابقة يمكن تفسيره من خلال تحديد التباين في المتغير الاخر والعكس صحيح، اما الجزء المتبقي من التباين فيطلق عليه اسم التباين العشوائي، ومثل هذا التباين لا يمكن تفسيره من خلال التباين في احد المتغيرات لأنه يعزى الى عوامل أخرى مثل عوامل الصدفة أو بسبب وجود متغيرات أخرى تؤثر في المتغيرات موضع البحث (الزغلول، 2005، صفحة 195).

وعلى الرغم من أهمية معامل التحديد كمقياس لأهمية أثر ما، فلا يمكن استخدامه في استنتاج العلاقات السببية، أي على الرغم من ان درجة الاهتمام بالمادة يعتبر مسؤولا عن 19.4% من التباين في نتائج الامتحان فليس من الضرور أن يكون سببا حتميا في هذا التباين.

تذكير:

رأينا عند التطرق للانحدار أنه يمكن ايجاد مجاهيل معادلة خط الانحدار البسيط كالتالي:

الصيغة رقم 1:

$$\boxed{Y= a+bx}$$

حيث (a) و(b) قيم يمكن حسابها من واقع البيانات التجريبية وذلك بتطبيق الصيغ التالية:

$$\boxed{\mathbf{b} = \mathbf{r}\left(\frac{\boldsymbol{\sigma}_y}{\boldsymbol{\sigma}_x}\right)}$$

الصيغة رقم 2:

ويتضح من الصيغة المتطلبات التي تقوم عليها وهي: معامل الارتباط والانحراف المعياري لكل من المتغيرين (x)و (y).

أما قيمة (a) في المعادلة السابقة فيمكن الحصول عليها من الصيغة الاحصائية التالية:

الصيغة 3:

$$\boxed{\boldsymbol{a = \bar{y} - b\bar{x}}}$$

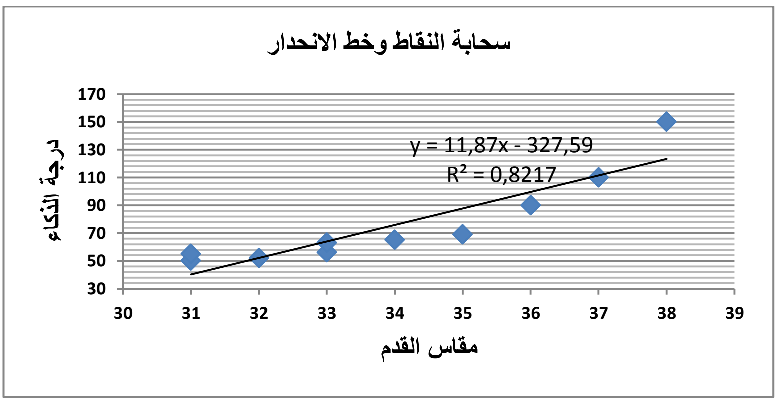
مثال :

الجدول التالي يمثل العلاقة بين مقاس القدم ودرجة الذكاء لدى عينة من التلاميذ :

العينة	مقاس القدم x_i	درجة الذكاء y_i	$x_i * y_i$	x^2		$x_i - \bar{x}$ f	$y_i - \bar{y}$ g	$x_i - \bar{x} * y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	31	50	1550	961		-3	-26	78	9	676
2	31	55	1705	961		-3	-21	63	9	441
3	32	52	1664	1024		-2	-24	48	4	576
4	33	56	1848	1089		-1	-20	20	1	400
5	33	63	2079	1089		-1	-13	13	1	169
6	34	65	2210	1156		0	-11	0	0	121

7	35	69	2415	1225		1	-7	-7	1	49
8	36	90	3240	1296		2	14	28	4	196
9	37	110	4070	1369		3	34	102	9	1156
10	38	150	5700	1444		4	74	296	16	5476
$\sum$	**340**	**760**	**26481**	**11614**		**0**	**0**	**641**	**54**	**9260**
قيم محسوبة	$\bar{x} = 34$	$\bar{y} = 76$								

سحابة النقاط ومعادلة خط الانحدار:



يتضح من الشكل وجود العلاقة الخطية بين المتغيرين محل الدراسة، وبالإضافة الى معادلة خط الانحدار (y) التي سنوضح طريقة الوصول اليها من خلال الخطوات التالية:

2-1-ايجاد معامل التحديد من تربيع قيمة معامل الارتباط بيرسون:

1-ايجاد معامل التحديد من معامل الارتباط :

1-1-حساب معامل الارتباط r :

أ- **حساب المتوسط الحسابي** لكل من المتغيرين (x)و (y):

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

$$= \frac{340}{10}$$

$\bar{x} = 34$

$$\bar{y} = \frac{\sum_{n}^{i=1} y_i}{n}$$

$$= \frac{760}{10}$$

$\bar{y} = 76$

ب- **حساب الانحراف المعياري** لكل من المتغيرين (x)و(y) :

$$(\sigma_x) = \sqrt{\frac{(\boldsymbol{x_i} - \overline{\boldsymbol{x}})^2}{n}} = \sqrt{\frac{54}{10}}$$

$\sqrt{6}$=2.45

$$(\sigma_y) = 2.32$$

$$\left(\sigma_y\right) = \sqrt{\frac{(y_i - \bar{y})^2}{10}} = \sqrt{\frac{9260}{10}}$$

$\sqrt{926}$=30.43

$$\left(\sigma_y\right) = 30.43$$

ج- حساب معامل لارتباط(r) :

ج-1- حساب قيمة التباين المشترك COV :

$$COV = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n}$$

$$COV = \frac{641}{10} = 64.1$$

ج-2-معامل الارتباط بيرسون r :

$$r = \frac{Cov(x,y)}{(\sigma_x)\left(\sigma_y\right)}$$

$$r = \frac{64.1}{(2.32)(30.46)}$$

$$\frac{64.1}{70.67} = 0.90$$

ومنه فان العلاقة قوية جدا بين مقاس قدم التلميذ ودرجة ذكائه في العينة المبحوثة.

د- ايجاد قيمة معامل التحديد R^2 :

$$R^2 = r^2$$

$$= 0.90^2$$

=0.82

وهي تشير الى أن مقاس القدم يفسر نسبة 82% من التحول المسجل في درجة الذكاء بينما تقى نسبة 18% المتبقية تعود الى عوامل أخرى.

2- حساب معامل التحديد بناء على معادلة خط الانحدار: وذلك باستخدام الصيغة الاحصائية التالية :

$$r^2 = 1 - \frac{SDE_D}{SCE_{\bar{y}}}$$

حيث :

D: هي الخط الذي يمر بأقرب ما يكون من كل نقطة من نقاط السحابة.

$\bar{y}$: هي المتوسط الحسابي لقيم y.

SCE_D: هي مجموع مربع انحرافات قيم المتغير التابع عن القيم التي تقابلها على خط الانحدارD.

$SCE_{\bar{y}}$: هي مجموع مربع انحرافات قيم (y) عن وسطها الحسابي $(\bar{y})$.

1-2-يمكن ايجاد قيمة SCE_D باستعمال الصيغة التالية:

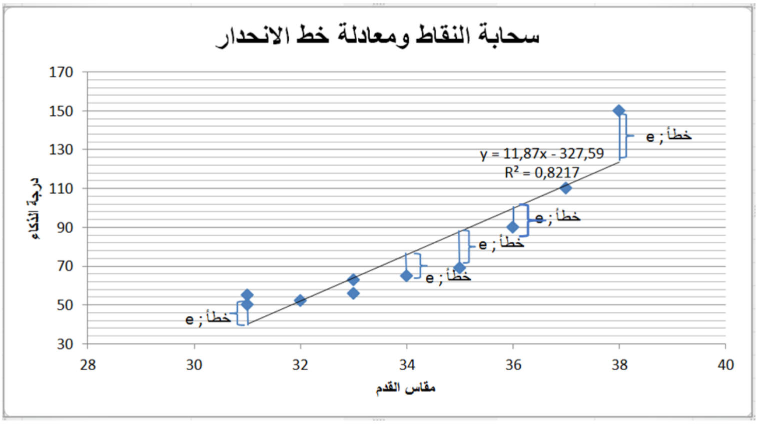
$$SCE_D = e_1^2 + e_2^2 + \cdots e_n^2$$

حيث: *(e)* هو قيمة الخطأ الذي ينتج من تعويض النقطة (x_1,y_1)بالنقطة $(\hat{y}, x1)$ والتي ينتج عنها خطأ المسافة التي تبتعد بها هذه النقطة عن خط الانحدار كما يتضح من خلال الشكل البياني.

ويمكن ايجاد قيمها (e) بالصيغة التالية :

$$e = \left(y_1 - (bx_1 + a)\right)^2 + \left(y_2 - (bx_2 + a)\right)^2 + \cdots \left(y_n - (bx_n + a)\right)^2$$

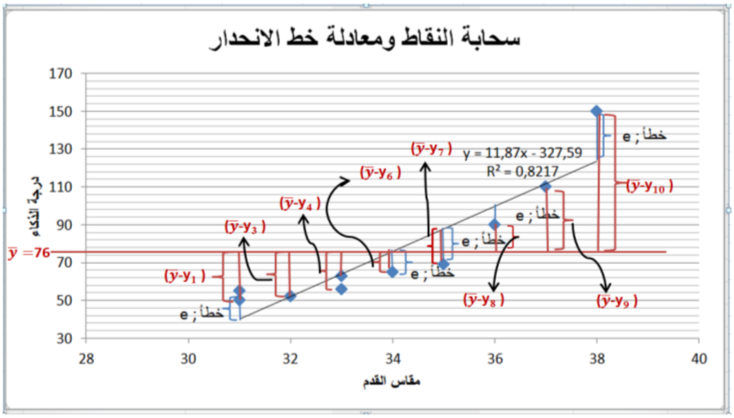
ويتضح المقصود بهذه الاخطاء على الشكل التالي:



2-2-يمكن ايجاد قيمة $SCE_{\bar{y}}$ من الصيغة التالية :

$$SCE_{\bar{y}} = (y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \cdots (y_n - \bar{y})^2$$

ويتضح المقصود بهذه الفروق على الشكل التالي على سحابة النقاط :



و نظرا لاعتماد كل من $SCE_{\bar{y}}$و SCE_D على معلومات معادلة خط الانحدار لابد أولا ايجاد مجاهيل هذه المعدلة المتمثلة في القيم الامبريقة لكل من b وa كما يلي:

أ-ايجاد قيمة المجهول (b) في معادلة خط الانحدار:

$$\mathbf{b = r}\left(\frac{\sigma_y}{\sigma_x}\right)$$

$$b = 0.90\left(\frac{30.43}{2.32}\right)$$

$$= 0.90 * 13.12 = 11.81$$

ومنه فان: b=11.81

ب- ايجاد قيمة المجهول (a) في معادلة خط الانحدار:

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

$$= 76 - (11.81)(34)$$

$$= 76 - 401.52$$

ومنه فان: a=-325.54

ج- تحديد معادلة خط الانحدار:

y=a+b(x)

$$y_{(x)} = (-325.54) + (11.81)(x)$$

3-2- ايجاد معامل التحديد بالاعتماد على بيانات معادلة خط الانحدار من خلال حساب قيم كل من $SCE_{\bar{y}}$و SCE_D بتطبيق الصيغ الخاصة بها والمذكورة أعلاه كما يتضح في الجدول التالي:

طريقة الحساب		$\left(y_1 - (bx_1 + a)\right)^2$		
$(y_i - \bar{y})^2$	$\left(y_i - (11.81x_i + (-324.52))\right)^2$			
$SCE_{\bar{y}}$	SCE_D	درجة الذكاء y	مقاس القدم x	القيم
676	88,92	50	31	**1**
441	208,22	55	31	**2**
576	0,14	52	32	**3**
400	67,08	56	33	**4**
169	1,42	63	33	**5**
121	121,00	65	34	**6**
49	353,82	69	35	**7**
196	92,54	90	36	**8**
1156	2,04	110	37	**9**
5476	716,10	150	38	**10**
9260	1651,29	760	340	$\sum$

ومنه :

$$r^2 = 1 - \frac{SDE_D}{SCE_{\bar{y}}}$$

$= 1 - \frac{1651.29}{9260}$= 1-0.178 =0.82

وهي نفس القيمة التي تم التوصل اليها مباشرة من تربيع قيمة معامل الارتباط بيرسون.

والتي تشير الى ان التغير في مقاس القدم يفسر 82% من التغير الموجود في درجة الذكاء، مع وجود ما يقارب 18% من التغير في درجة الذكاء يعود اما الى عوامل اخرى او الى عوامل الصدفة.

(Khan, R-carré ou le coefficient de détermination, 2019)

الفصل العاشر اختبار الفرضيات

1- مفاهيم أساسية في اختبار الفرضيات
2- اختبار T test
3- اختبار كاي تربيع χ^2

اختبار الفرضيات

أولا: مفاهيم اساسية في اختبار الفرضيات

1- مفهوم اختبار الفرضيات: هو أحد أساليب الاحصاء الاستدلالي، الذ تستخدم فيه بيانات العينة المسحوبة من مجتمع الدراسة لاتخاذ قرارات او اصدار أحكام حول قيمة أو معلمة أو أكثر من معلمات المجتمع . ان الفكرة أساسية في اختبار الفرضيات هي قسمة قيمة المعلمة التي نفترضها للمجتمع (في الفرضية الصفرية) والقيمة المقابلة لها في العينة (الاحصائية)وننسب هذا الفرق الى الخطأ المعياري للإحصائية، والمعيار الذي يستطيع من خلاله الباحث الحكم على هذا الفرق تتم من خلال قسمة هذا الفرق على الخطأ المعياري، ثم مقارنة خارج القسمة بالقيمة الجدولية أو ما يسمى بحدود منطقتي القبول او الرفض.

مع العلم أن المنهج الاساسي لاختبار الفرضيات بالمعنى الدقيق هو المنهج التجريبي. في حين يمكن لأي منهج بحثي آخر أن تصاغ له فرضيات ويتم اختبارها بالطرق الملائمة.

2- أنواع الفرضيات في الاختبار الاحصائي :

الفرض الصفري: يرمز له بالرمز: H_o يمتضمن الهدف من الاختبار، وقبوله يعني عدم رفض نتائج العينة.

الفرض البديل: هو نفسه الفرض البحثي، وتكمن أهميته في انه يحدد قيمة الدرجة الحرجة التي تستخدم للتحقق من الفرضية احصائيا، فاذا كان الفرض البديل عديم الاتجاه فان القيم المحسوبة كنتائج للبحث تتم مقارنتها مع التوزيع النظري المعروف بذو الحدين Bilatéral، أما اذا كان الفرض البديل ذو اتجاه

محدد فنقارن النتائج مع التوزيع النظري المعروف باختبار ذو الحد الواحد Unilatéral. (الاحصاء بلا معاناة)

بهذا يتحدد الفرض البديل (او الفرض البحثي) أولا بفرضية البحث، ثم يحدد بعد ذلك بناء على الفرض الصفري كنقيض للفرض البديل، وتصاغ الفروض (الصفرية والبديلة) باحدى الصيغ التالية (مثل في حالة اختبار المتوسطات) (نفس المرجع الاحصاء بلا معاناة)

مثال:

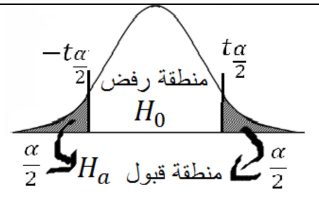
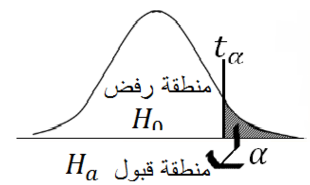
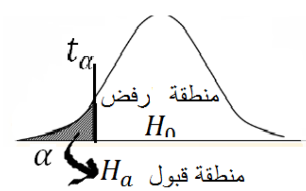
طريقة جديدة في التدريس ولتكن استعمال الوسائط الحديثة في التدريس، تصاغ الفرضيات بالشكل التالي:

الفرضية الصفرية:

طريقة الوسائط الحديثة في التدريس لا تختلف عن الطريقة الكلاسيكية في تحصيل الطلبة.

الفرضية البديلة:

يساعد استخدام الوسائط الحديثة في التدريس على تحصيل أفضل لدى الطالبة. بعد اجراء التجربة وجمع البيانات التي قد تكون متوسطات التحصيل لدى الطلبة تصاغ الفرضيات حسب أهداف البحث بشكل من الاشكال التالية :

نوعية الفرضية	الفرض الصفري	الفرض البديل	الاختبار المستخدم	منطقة القبول والرفض (المنطقة الحرجة)
غير موجهة	$h_o: \overline{x}_1 = \overline{x}_2$	$h_a: \overline{x}_1 \neq \overline{x}_2$	اختبار ذو حدين Bilatéral	
موجهة	$h_o: \overline{x}_1 \leq \overline{x}_2$	$h_a: \overline{x}_1 > \overline{x}_2$	اختبار ذو حد واحد ناحية اليمين Unilatéral	
موجهة	$h_o: \overline{x}_1 \geq \overline{x}_2$	$h_a: \overline{x}_1 < \overline{x}_2$	اختبار ذو حد واحد ناحية اليسار Unilatéral	

3- مستوى المعنوية الاسمي α:

يمثل الحد الأقصى المقبول لاحتمال الخطأ من النوع الاول، أو نسبة الخطأ المحتملة في اتخاذ قرار رفض أو قبول الفرضية الصفرية، ويحدده الباحث قبل جمع البيانات، وفي مجال البحوث الاجتماعية فان أقصى مستوى مقبول للدلالة الاحصائية هو α=0.05 أي 5% ويمكن ان ينخفض الى 1% والمقصود بهذه النسبة في البحث تعني أنه لو تم تكرار التجربة 100 مرة فان احتمال رفض الفرض الصفري وهو صحيح ستكون خمس مرات أي أن نسبة الشك فيما تم التوصل اليه هي 5%، في المقابل فان الاستنتاج المتوصل اليه يكون صحيحا بنسبة 95%.

4-مستوى المعنوية الحقيقي: P. Value يسمى أيضا بالقيمة الاحتمالية، وهي قيمة احتمال الخطأ المحسوب فعليا من بيانات العينة، وهذه القيمة تقدمها مختلف البرمجيات عند حساب الدلالة الاحصائية لأي معامل مثل الدلالة الاحصائية لمعاملات الارتباط، أو عند اختبار الفرضيات عموما، حيث يرفض الفرض الصفري اذا ما جاءت قيمة P.value على من مستوى الدلالة $\alpha=0.05$ والعكس، كما تعد هذه القيمة أفضل مؤشر على مدى مصداقية الفرض محل الاختبار. (نفس المرجع الاحصاء بلا معاناة)

5-قرارات اختبار الفرضية الصفرية:

ان اختبار الفرضية بأسلوب احصائي يؤدي الى اتخاذ قرار اذا ما كانت الفرضية مقبولة ام مرفوضة:

فرفض الفرضية الصفرية لا يعني بالضرورة أن تكون خاطئة في المقابل فان قبول الفرضية الصغرية لا يعني بالضرورة أن تكون صحيحة، وبناء على ذلك حدد المتخصصون نوعين من الخطأ الذي قد يحدث في اختبار الفرضيات الاحصائية يمكن اختصارها في الجدول التالي:

القرار	**الفرضية الصفرية**	
	صحيحة	**خاطئة**
قبول الفرضية الصفرية H_0	صحيحة	خطأ من النوع الثاني (β) $1-\beta$
رفض الفرضية الصفرية H_0	خطأ من النوع الاول (α) $1-\alpha$	صحيحة

- يمكن تخفيض الخطأ من النوع الاول بزيادة مستوى الدلالة، الذي يزيد من فرص الوقوع في الخطأ الثاني(أي قبول الفرض الصفري بينما هو خاطئ) .
- والتقليل من الخطأ الثاني يأتي بزيادة حجم العينة بهدف الحصول على قوة اختبارية عالية.
- ان الخطأ الاول أكثر خطورة من الخطأ الثاني(...)
- ان تقليل أحد الخطأين يؤدي لزيادة الخطأ الآخر (توجد علاقة عكسية بينهم).

ثانيا: اختبار T.test

1- **تعريفه:** هو أحد الاختبارات البارامتيرية التي يتطلب تطبيقها تحقق جملة من الشروط هي :

- ان تكون ان تكون المعاينة عشوائية: استخدام الاسلوب العشوائي في اختيار العينات.
- أن يكون المتغير التابع في الدراسة كميا (من مستوى القياس الفئوي أو النسبي)
- ان يكون توزيع متغير الدراسة توزيعا اعتداليا
- استقلالية القياس أو المشاهدات.
- تجانس التباين: أي تماثل تشتت درجات المجموعات.

ملاحظة: بالإضافة الى الشروط السابقة، اذا كان تباين المتغير في المجتمع معلوما فانه يتم استخدام اختبار Z بدلا م اختبار t بغض النظر عن حجم العينة،

كما يفضل استخدام اختبار Z اذا كان تباين المتغير في المجتمع غير معلوم، لكن بشرط أن يكون حجم العينة 30 على الاقل، وهذا ينطبق على كل اختبارات t الاخرى. وفي حالة عدم تتحقق تلك الشروط فان الاختبار الذي

يستخدم وخاصة عندما يكون حجم العينة أقل من (30) وحدة احصائية فان اختبار (T.Student) يكون الانسب. أما بالنسبة للحالات التي يستخدم فيها هذا الاختبار هي:

1- اختبار (t) لعينة واحدة:

يستخدم اختبار (t) لعينة واحدة من أجل فحص فرضية حول معلمة المجتمع، أي مقارنة متوسط العينة بمتوسط المجتمع، مثل ادعاء أحد مدراء المؤسسات التربوية أن متوسط التحصيل في مؤسسته أعلى من المتوسط التحصيل العام في باقي المؤسسات المقدر بـ العلامة 11.5.

لاختبار هذا الفرضية نتبع الخطوات التالية:

1- صياغة الفرضيات الاحصائية:

$H_{o:}\ \mu = x$: لا يوجد اختلاف معنوي بين متوسط المجتمع ومتوسط العينة.

$H_{a:}\ \mu \neq x$: يوجد اختلاف معنوي بين متوسط المجتمع ومتوسط العينة..

2- تحديد اتجاه الفرض البديل ان كان ذو حدين أو ذو حد واحد (العودة الى المفاهيم الاساسية لاختبار الفرضية)

$Ho: \mu = x$: لا يوجد اختلاف معنوي بين متوسط المجتمع ومتوسط العينة.

$Ha: \mu \geq x$: يوجد اختلاف معنوي بين متوسط المجتمع ومتوسط العينة.

أو

$Ha: \mu < x$: يوجد اختلاف معنوي بين متوسط المجتمع ومتوسط العينة.

3- تحديد مستوى المعنوية: والذي غالبا ما يكون في العلوم الاجتماعية **α=0.05**

4- حساب درجة الحرية لهذا الاختبار كما يلي: (n-1)في حالة العينة الواحدة .

ملاحظة: عندما تكون درجة الحرية في هذا الاختبار ما لا نهاية)∞(فان توزيع (t) يصبح نفسه توزيع (Z).

5- اجراء الاختبار الاحصائي المناسب: والذي يختلف باختلاف عدد العينات (واحدة، اثنين)

6- اتخاذ القرار إما برفض أو قبول الفرض الصفري بدرجة معينة من الثقة.

اما بالنسبة للصيغة الرياضية لهذا الاختبار فهي كالتالي :

الصيغة رقم ..

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

مثال: على فرض ان المتوسط العام لطلبة كلية العلوم الانسانية والاجتماعية في مادة الاحصاء يساوي 10 وارتأى احد الأساتذة أن متوسط طلبة قسمه يختلف عن المتوسط العام لطلبة الكلية، وليتأكد من ذلك اختار عينة من 25 طالب فوجد ان متوسط العلامة في هذه العينة 13 بانحراف معياري يساوي 8، المطلوب فحص الفرضية الصفرية عند مستوى الدلالة **(α=0.05)** .

الحل:

1- وضع الفرضيات:

الفرضية الصفرية: **H_0 μ :10 =**

الفرضية البديلة :**H_a 10 ≠ μ :**

2- تحديد قيمة ألفا **(α=0.05) أو ما يعرف بأعلى نسبة خطأ يسمح بها الباحث.**

3-	حساب درجة الحرية (df): في اختبار (t) لعينة واحدة =(n-1 = 25-1= 24)

4-	بناء على القيمتين السابقتين يمكن استخراج قيمة الاختبار الجدولية بعد تحديد نوع التوزيع لهذا الاختبار ان كان ذو حد واحد أو ذو حدين هنا الاختبار ذو حدين (Bilatéral) ومنه فان القيمة الجدولية لاختبار (t) عند **(α=0.05)** و**(df=24)** هي: **(±2.0639)**

5-	**حساب قيمة الاختبار: باستخدام معادلة اختبار (t) المذكورة اعلاه نحصل على القيمة التالية:**

$$\boldsymbol{t = \frac{13 - 10}{\frac{8}{\sqrt{25}}} = \frac{3}{1.6} = 1.875}$$

6-	المقارنة بين القيمتين المحسوبة والجدولية للاختبار واتخاذ القرار:

بما ان القيمة المحسوبة للاختبار=1.875t أصغر من القيمة الجدولية $t_{0.05}$=2.063 نقبل الفرض الصفري ونرفض الفرض البديل، ما يدل على انه لا توجد فروق ذات دلالة احصائية بين متوسط العينة ومتوسط المجتمع، وأن الفروق المسجلة تعود الى محض الصدفة، وذلك مع احتمال وجود خطأ يقدر بـ 5%.

مثال 2: نفس المثال مع افتراض أن المتوسط المسجل في العينة المختارة قدر بـ 15 نقطة، وانحراف معياري يساوي 10 نتبع نفس خطوات الحل السابقة كما يلي:

الحل:

1- وضع الفرضيات:

الفرضية الصفرية: H_o μ :10 =

الفرضية البديلة: H_a $\mu \neq 10$:

2- تحديد قيمة ألفا **(α=0.05)**

3- حساب درجة الحرية (df): في اختبار (t) لعينة واحدة =n-1 = 25-1= (24)

4- بناء على القيمتين السابقتين يمكن استخراج قيمة الاختبار الجدولية بعد تحديد نوع التوزيع لهذا الاختبار ان كان ذو حد واحد أو ذو حدين هنا الاختبار ذو حدين (Bilatéral) ومنه فان القيمة الجدولية لاختبار (t) عند **(α=0.05)** و**(df=24)** هي: **(±2.0639)**

5- حساب قيمة الاختبار: باستخدام معادلة اختبار (t) المذكورة اعلاه نحصل على القيمة التالية:

$$\boldsymbol{t = \frac{15 - 10}{\frac{10}{\sqrt{25}}} = \frac{5}{2} = 2.5}$$

6- المقارنة بين القيمتين المحسوبة والجدولية للاختبار واتخاذ القرار:

بما ان القيمة المحسوبة للاختبار t=2.5 أصغر من القيمة الجدولية $t_{0.05}$=2.063 نرفض الفرض الصفري ونقبل الفرض البديل، ما يدل على انه توجد فروق ذات دلالة احصائية بين متوسط العينة ومتوسط المجتمع، وأن الفروق المسجلة تؤكد أن مستوى التحصيل في العينة أعلى من مستوى التحصيل في المجتمع، وذلك مع احتمال وجود خطأ يقدر بـ 5%.

2- اختبار t لمتوسطي عينتين مستقلتين في حالة $n_1 \neq n_2$:

شروط استخدامه:

أن يكون المتغير المستقل متغيرا تصنيفيا ذا مستويين أثنين (ذكر.أنثى أو متعلم-غير متعلم)

السحب العشوائي للعينتين من مجتمعاتها

توزيع المتغير التابع اعتدالي.

استقلالية البيانات في العينتين عن بعضهما

أن تكون تباينات المتغير التابع للمجموعات متجانس(يمكن اختباره)

الخطوات الاساسية للاختبار:

1- التأكد من تجانس تباين المجموعتين قبل الاختبار.

2- التأكد من اعتدالية توزيع العينتين.

3- صياغة الفرض الصفري والفرض البديل.

4- تحديد أعلى نسبة خطأ يسمح بها الباحث (مستوى الدلالة(α)

5- تحديد الاختبار المناسب لاختبار الفرضية الصفرية، في هذه الحالة يتمثل في الصيغة التالية :

الصيغة رقم ..	$\boldsymbol{t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\dfrac{n_1\sigma^2{}_1 + n_2\sigma^2{}_2}{n_1 + n_2 - 2}\right)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}}$

6- جمع المعلومات واجراء الاختبار

7- اتخاذ القرار

مثال توضيحي:

البيانات التالية تمثل نتائج تطبيق اختبار التحصيل في مادة الاحصاء على عينتين الأولى من الذكور والثانية من الاناث في كلية العلوم الاجتماعية لإحدى الجامعات، فكانت البيانات كالتالي:

$\mathbf{n_1} = 101$	$\mathbf{n_2} = 81$
$\boldsymbol{\bar{x}_1} = 55.02$	$\boldsymbol{\bar{x}_2} = 53.20$
$\boldsymbol{M_{e1}} = 54$	$\boldsymbol{M_{e2}} = 56.40$
$\boldsymbol{\sigma_1} = 16.33$	$\boldsymbol{\sigma_2} = 14.67$

المطلوب:

هل يمكن تطبيق اختبار (t)

هل هناك فروق ذات دلالة بين متوسط الذكور ومتوسط الاناث في تحصيل هذه المادة؟

1- **التأكد من اعتدالية التوزيع:** نحسب الالتواء في العينتين بتطبيق الصيغة التالية :

الصيغة رقم :	$\boldsymbol{\gamma = \frac{3(\bar{x} - M_e)}{\sigma}}$

العينة 1:

$$\gamma_1 = \frac{3(55.02 - 54)}{16.33} = 0.18$$

العينة 2:

$$\gamma_1 = \frac{3(53.20 - 56.40)}{14.67} = 0.65$$

النتيجة: توضح قيمتي الالتواء أن التوزيعين اعتداليين.

2- **التأكد من تجانس العينتين**: باستخدام اختبار (f) الذي تتماثل خطوات اجرائه مع خطوان أي اختبار احصائي للفرضيات وذلك باتباع الخطوات التالية:

أ- صياغة الفرضية الصفرية والفرضية البديلة :

$H_o: \sigma_1 = \sigma_2$

$H_a: \sigma_1 \neq \sigma_2$

ب- استخدام اختبار(f) بتطبيق المعادلة التالية :

$$f = \frac{\sigma_1^2}{\sigma_2^2} = \frac{(16.33)^2}{(14.67)^2} = 1.23$$

يقوم اختبار (f)بقسمة التباين الاكبر على التباين الاصغر كما يتضح اعلاه، لنحصل على قيمة *(f)* المحسوبة ومنه : *f=1.23*

- تحديد مستوى الدلالة **α=0.05**
- حساب درجة الحرية لهذا الاختبار كما يلي: df = (101+81)-2=180 $=(n_1 + n_2)-2$
- استخراج قيمة (f) الجدولية عند ***α=0.05* و *df=180***

$F_{0.05}=1.45$

- مقارنة القيمتين واتخاذ القرار: $f=1.23<f_{0.05}=1.45$

بما ان القيمة المحسوبة اصغر من القيمة الجدولية نقبل الفرض الصفري القائل أنه لا يوجد اختلاف ذو دلالة احصائية بين تباين العينتين، وبالتالي فالعينتين متجانستين.

**بهذا نكون قد تحققنا من الشرط الثاني وهو تجانس العينتين.

3- صياغة الفرضيات الاحصائية:

$H_o\ \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

وهو فرض غير موجهة، أي ذو حدين.

4- **مستوى الدلالة** أو أعلى نسبة خطأ يسمح بها الباحث **α=0.05**

5- **درجة الحرية: (1-81)+(-101) =(n-1)+(n-1) 180=**

6- **تطبيق الاختبار:**

$$t = \frac{55.02 - 53.20}{\sqrt{\left(\frac{101(16.33)^2 + 81(4.67)^2}{(101+81)-2}\right)\left(\frac{1}{101}\right) + \left(\frac{1}{81}\right)}}$$

$$= \frac{1.82}{\sqrt{246.5}} \quad =0.77$$

ومنه فان قيمة (t) المحسوبة هي 0.77.

أما قيمة (t) المجدولة المقابلة لدرجة الحرية 180 ومستوى الدلالة 0.05 هي:

$t_{0.05}$=1.97

7- **القرار:** وبما ان (t) المحسوبة اصغر من ($t_{0.05}$) المجدولة، نقبل الفرض الصفري القائل ان المتوسطين متساويين، أي لا توجد فروق ذات دلالة احصائية بين المتوسطين.

ملاحظة: (الدلالة توجه الى المتوسط الاكبر في حالة المقارنة بين مجموعتين)

3- **اختبار t لمتوسطي عينتين غير مستقلتين :**

في حالة تساوي حجم العينتين أي (n_1=n_2) وكثيرا ما نجد هذه الحالة عند اتباع المنهج التجريبي أين يتم اختيار عينة تجمع منها البيانات قبل وبعد الاختبار لنحصل على مجموعتين من البيانات من نفس العينة لكن في ظروف مختلفة. ولمعرفة ما اذا كان للعامل التجريبي الذي اختلف حضوره بين الجمع الاول للبيانات والجمع الثاني نختبر الفرق بين متوسط البيانات التي تم جمعها

قبل ادخال المتغير التجريبي ومتوسط البيانات التي جمعت بعد ادخال المتغير التجريبي لمعرفة ان كان لهذا الاخير أثر في النتائج، وفي هذه الحالة نكون امام ما يعرف باختبار الفرق بين متوسطين مرتبطين بعينة واحدة ولتحقيق ذلك نتبع نفس الخطوات السابقة لاختبار (t) في حالة عينتين مستقلتين مع استخدام الصيغة الاحصائية التالية:

صيغة رقم: ..

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{n-1}}}$$

مثال:

العينة 1	**العينة 2**
$n_1 = 15$	$n_2 = 15$
$\bar{x}_1 = 40$	$\bar{x}_2 = 35$
$\sigma_1 = 5$	$\sigma_2 = 7$

1- **صياغة الفرضيات الاحصائية:**

$H_o\ \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

وهو فرض غير موجهة، أي ذو حدين.

2- **مستوى الدلالة** أو أعلى نسبة خطأ يسمح بها الباحث **α=0.05** ؛ **5**%.

3- **درجة الحرية:** 2-(n*2) =28 2-(2*15)=

4- **تطبيق الاختبار:**

$$t = \frac{40 - 35}{\sqrt{\frac{25 + 49}{14}}}$$

$$= \frac{5}{2.29} \quad = \mathbf{2.18}$$

ومنه فان قيمة (t) المحسوبة هي **2.18**.

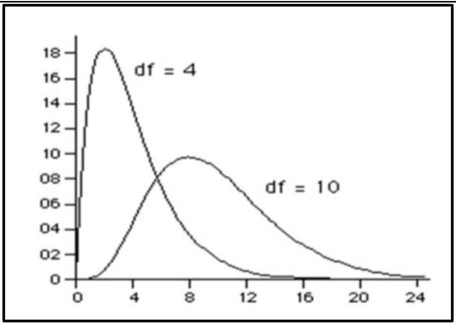
أما قيمة (t) المجدولة المقابلة لدرجة الحرية 28 ومستوى الدلالة 0.05 هي:

$t_{0.05}=2.04$

القرار: بما ان (t) المحسوبة اكبر من ($t_{0.05}$) المجدولة، نرفض الفرض الصفري القائل ان المتوسطين متساويين، ونقبل فرضية البحث، أي توجد فروق ذات دلالة احصائية بين المتوسطين.

ثالثا: اختبار كاي تربيع χ^2

يعتبر من التوزيعات الاحتمالية الشائعة الاستخدام حيث توجد له تطبيقات عديدة بدرجة يمكن معها القول أنه يأتي في المرتبة الثانية بعد التوزيع المعتدل (الطبيعي) من حيث كثرة تطبيقاته، وهو من الاختبارات لا معلمية، ويعتمد توزيع χ^2 اعتمادا كاملا على درجات الحرية كما هو الحال بالنسبة لتوزيع (t)، وعلى الرغم من ذلك يوجد اختلافات رئيسية بين التوزيعين حيث أن توزيع (t) متماثل حول وسطه الحسابي $\mu=0$ بينما يعتبر توزيع χ^2 توزيعا ملتويا جهة اليمين (التواء موجب) وخاصة عندما تكون درجات الحرية (df) صغيرة، في المقابل كلما زادت درجات الحرية كلما قل إلتواء التوزيع واقترب الى التماثل كما يتضح من الشكل البياني التالي:



شكل رقم: .. توزيع χ^2 حسب درجات الحرية

حالات استخدام اختبار χ^2 :

- اذا كان التوزيع غير اعتدالي.
- عندما يكون حجم العينة قليل مقارنة مع حجم المجتمع الذي سحبت منه.
- نوع العينة غير احتمالية
- في حالة طبيعة المتغيرات نوعية وليست كمية.
- يساعد هذا الاختبار الباحثين على معرفة حقيقة الفروق المسجلة بين المبحوثين، ان كانت فروقا جوهرية أم ناتجة عن الصدفة، خصوصا فيما يتعلق بمواقف هؤلاء المبحوثين واتجاهاتهم نحو قضية اجتماعية معينة، وذلك من خلال مقارنة التكرارات المشاهدة بالتكرارات المتوقعة، مما يستدعي استخدام اختبار χ^2 (بن محمد الصغير، 2001، صفحة 146).
- يستخدم لهدفين لاختبار جودة المطابقة كما يستخدم لاختبار الاستقلالية بين متغيرين.

I - اختبار χ^2 لجودة التطابق:

يستخدم اختبار الكاي تربيع لجودة التطابق لمعرفة امكانية تطابق التكرارات الملاحظة لكل فئة من فئات المتغير الفئوي مع ما هو متوقع لها.

- **شروط استخدامه:**

أن تنتمي البيانات لفئات او تقسيمات ولا يهم ان كانت هذه الفئات اسمية مثل الحالة الاجتماعية (أعزب، مطلق، متزوج...) أو رتبية مثل المستوى التعليمي (ابتدائي، متوسط، ثانوي ...)

- أن تكون الفئات أو التقسيمات شاملة وغير متداخلة
- أن يكون حجم العينة كبير.

الخطوات الاساسية للاختبار:

1. **صياغة الفرضيات:**

- تبين الفرضية الصفرية لاختبار جودة المطابقة بأن االنموذج الاحتمالي المفروض يصف بدقة توزيع البيانات في المجتمع ويضاغ كالتالي:
- H_o: يوجد تطابق بين القيم المشاهدة والقيم النظرية
- بينما الفرضية البديلة تحتوي على رفض الادعاء السابق، وتصاغ كما يلي:
- H_a: لا يوجد تطابق بين القيم المشاهدة والقيم النظرية.

2. تحديد أعلى نسبة خطأ يسمح بها الباحث (مستوى الدلالة(α)
3. حساب درجة الحرية df= n-1
4. استخراج القيمة الجدولية للكاي تربيع.
5. ايجاد القيمة χ^2 المحسوبة لاختبار الفرضية الصفرية، باستخدام الصيغة التالية :

الصيغة رقم ..	$\chi 2=\frac{(O_1-E_1)^2}{E_1}\frac{(O_2-E_2)^2}{E_2}\ldots.+\frac{(O_6-E_6)^2}{E_6}$

حيث **O** هو التكرار المشاهد و**E** هو التكرار المتوقع

و يمكن ايجاد قيمة التكرار المتوقع او النظري لكل تكرار مشاهد بتطبيق الصيغة التالية :

صيغة رقم ..	$\boldsymbol{E = \frac{n_i}{k}}$

حيث :

- n_i: هو تكرار الفئة.
- *K: هو عدد الفئات.*

6. اتخاذ القرار

مثال:

البيانات التالية مرتبطة بالحالة الاجتماعية لعينة من 50 عامل توزعت كما يلي:

الحالة الاجتماعية	التكرار المشاهد	التكرار المتوقع
أعزب	16	52/4=13
متزوج	28	13
مطلق	5	13
أرمل	3	13
المجموع	52	

1- صياغة الفرضيات:

Ho: يوجد تطابق بين القيم المشاهدة والقيم النظرية

Ha: لا يوجد تطابق بين القيم المشاهدة والقيم النظرية

2- حساب التكرارات المتوقعة بقسمة مجموع التكرارات (52) على عدد الفئات المقدرة بـ (5).

$$E = \frac{52}{4} = 13$$

3- ايجاد القيمة الحسابية للكاي تربيع كما يلي:

$$\boldsymbol{\chi 2} = \frac{\mathbf{(16-13)^2}}{\mathbf{13}} + \frac{\mathbf{(28-13)^2}}{\mathbf{13}} + \frac{\mathbf{(5-13)^2}}{\mathbf{13}} + \frac{\mathbf{(3-13)^2}}{\mathbf{13}}$$
$$= 0.692 + 17.307 + 4.923 + 7.692 = \mathbf{30.614}$$

ومنه فان $\boldsymbol{\chi 2}$ المحسوبة = **30.614**

4- ايجاد قيمة الكاي تربيع الجدولية عند مستوى الدلالة: **α=0.05**، **ودرجة الحرية (df=(n-1= 4-1 =3**

فان $\boldsymbol{\chi 2}$ من جدول التوزيع الكاي تربيع هي $\boldsymbol{\chi 2 = 7.8174}$

5- المقارنة واتخاذ القرار: يتبين من القيم أعلاه أن القيمة حسابية للاختبار اكبر من القيمة الجدولية وبناء عليه يرفض الفرض الصفري القائل بوجود تطابق بين القيم المشاهدة في هذا التوزيع والقيم النظرية ويقبل الفرض البديل القائل بعدم وجود تطابق بين القيم المشاهدة والقيم النظرية.

II - اختبار χ^2 للاستقلالية:

هو اختبار يقوم به الباحث لمعرفة ما إذا كان هناك استقلالية بين متغيرين في الدراسة أم أنهما تابعين لبعضهما في التغير. تتبع نفس الخطوات السابقة مع وجود اختلاف في حساب قيمة التكرارات المتوقعة كما سيتضح:

أ- خطواته :

1-صياغة الفرضيات :

الفرضية الصفرية: لا توجد تبعية بن المتغيرين محل الدراسة، ويرمز لها بالرمز H_0 والذي يتم اختبار صحته عند القيام بالاختبار.

الفرضية البديلة: توجد تبعية بين المتغيرين محل الدراسة، ويرمز لها بالرمز H_A.

2-حساب قيمة اختبار كاي تربيع: يتم ذلك بتطبيق الصيغة التالية:

صيغة رقم 01	$$\chi 2=\frac{(O_1-E_1)^2}{E_1}+\frac{(O_2-E_2)^2}{E_2}\dots.+\frac{(O_6-E_6)^2}{E_6}$$

حيث **O** هو التكرار المشاهد و**E** هو التكرار المتوقع

و يمكن ايجاد قيمة التكرار المتوقع او النظري لكل تكرار مشاهد بتطبيق الصيغة التالية:

صيغة رقم ..	$$E=\frac{R.total * C.total}{G.total}$$

حيث أن:

R.total: هو مجموع الصف الخاص بالخانة

C.total: هو مجموع العمود الخاص بالخانة

G.total: هو المجموع الكلي للجدول

بهذا يمكن ايجاد قيمة التكرار المتوقع أو النظري للخانة الأولى في الجدول كما يلي:

الموقف / النوع		موافق	محايد	غير موافق	R.total
ذكور	O	93	12	70	175
	E	**(105.00)**	**(10.50)**	**(59.50)**	
اناث	O	87	6	32	125
	E	**(75.00)**	**(7.50)**	**(42.50)**	
C.total		180	18	102	**G.total** 300

بعد حساب التكرارات النظرية يمكن حساب قيمة كاي تربيع بتبيق الصيغة **رقم:** .. كما يلي :

$$\chi 2=\frac{(93-105)^2}{105}+\frac{(87-75)^2}{75}+\frac{(12-10.50)^2}{10.50}+\frac{(6-7.50)^2}{7.50}+\frac{(70-59.50)^2}{59.50}+\frac{(32-42.50)^2}{42.50}$$

$$=1.371+1.920+0.214+0.300+1.853+2.594=\mathbf{8.252}$$

3- تحديد درجة الحرية الخاصة باختبار الكاي: يمكن تطبيق الصيغة التالية لتحديد درجة الحرية:

درجة الحرية	
صيغة رقم...	$df = (r-1)(c-1)$

حيث: c = عدد الاعمدة في الجدول

r= عدد الأسطر في الجدول.

ومنه فان df = (3-1)*(2-1) =2

4-استخراج قيمة χ^2 الجدولية: بالاعتماد على قيمة درجة الحريةdf =2 ومستوى الدلالة α=0.05 فان كاي المجدولة هي $\chi^2_{0.05}$=5.991 كما يتضح من الجدول التالي :

ν \ α	0,990	0,975	0,950	0,900	0,100	0,050	0,025	0,010	0,001
1	0,0002	0,0010	0,0039	0,0158	2,7055	3,8415	5,0239	6,6349	10,8276
2	0,0201	0,0506	0,1026	0,2107	4,6052	5,9915	7,3778	9,2103	13,8155
3	0,1148	0,2158	0,3518	0,5844	6,2514	7,8147	9,3484	11,3449	16,2662
4	0,2971	0,4844	0,7107	1,0636	7,7794	9,4877	11,1433	13,2767	18,4668
5	0,5543	0,8312	1,1455	1,6103	9,2364	11,0705	12,8325	15,0863	20,5150
6	0,8721	1,2373	1,6354	2,2041	10,6446	12,5916	14,4494	16,8119	22,4577
7	1,2390	1,6899	2,1673	2,8331	12,0170	14,0671	16,0128	18,4753	24,3219
8	1,6465	2,1797	2,7326	3,4895	13,3616	15,5073	17,5345	20,0902	26,1245
9	2,0879	2,7004	3,3251	4,1682	14,6837	16,9190	19,0228	21,6660	27,8772
10	2,5582	3,2470	3,9403	4,8652	15,9872	18,3070	20,4832	23,2093	29,5883

جدول رقم: .. التوزيع الاحتمالي χ^2

5- المقارنة واتخاذ القرار: نقارن قيمة كاي المحسوبةχ^2=8.252 مع قيمة كاي المجدولة $\chi^2_{0.05}$ =5.991

أين يتبين أن القيمة المحسوبة أكبر من القيمة المجدولة: $\chi^2 > \chi^2_{0.05}$

عند درجة الحرية **df =2**، ومستوى الدلالة **α=0.05**، بمعنى آخر فهي تقع داخل منطقة الرفض.

- **القرار:** لا يمكن قبول الفرضية الصفرية عند مستوى الدلالة $\boldsymbol{\alpha=0.05}$، بينما تقبل الفرضية البديلة.

- **الاستنتاج:** نستنتج انه توجد تبعية بين متغير الجنس ومتغير الموقف في هذه الدراسة.

قائمة المراجع

1. أبو النيل محمود السيد، **الاحصاء النفسي والاجتماعي والتربوي**، بيروت: دار النهضة العربية، 1987.
2. الصياد جلال، محمد ربيع عبد الحميد، **مبادئ الطرق الاحصائية**، ط1، جدة: تهامة، 1983.
3. بن محمد الصغير صالح، **مقدمة في الاحصاء الاجتماعي** ،الرياض :النشر العلمي والمطابع، 2002.
4. بوحفص عبد الكريم، **الحصاء المطبق في العلوم الاجتماعية والانسانية**، ط3،الجزائر: ديوان المطبوعات الجامعية، 2011.
5. بونوارة خزار محمد، مبادئ الاحصاء، الجزائر: مركز منشورات باتنة، 1996.
6. جلاطو جيلالي، **الاحصاء**، الجزائر، ديوان المطبوعات الجامعية، 2002.
7. حليمي عبد القادر، **مدخل الى الاحصاء**، ط2، الجزائر ديوان المطبوعات الجامعية، 1993.
8. خلف عبد الجواد مصطفى، **الاحصاء الاجتماعي**، المبادئ والتطبيقات، ط2، عمان: دار المسيرة للنشر والتوزيع، 2013.
9. خليل شرف الدين، **الاحصاء الوصفي**، شبكة الابحاث والدراسات الاقتصادية، www.rr4ee.net.
10. شاكر حمودي سعدي، **مبادئ الاحصاء وتطبيقاته**، ط1، عمان: دار الثقافة للنشر والتوزيع، 2000.
11. عبد السميع طبية أحمد، **مبادئ الاحصاء**، ط1، عمان: دار البداية، 2007.

12. 10علام اعتماد، رسلان يسرى، **اساسيات الاحصاء الاجتماعي**، عمان، دار الثقافة للنشر والتوزيع، 1987

13. ر. شبيجل موارى، **الاحصاء**، ط8، سلسلة شوم، مصد: الدار الدولية للاستثمارات الثقافية2008.

14. محمد موسى أماني، **التحليل الاحصائي للبيانات**، القاهرة: معهد الدراسات والبحوث الاحصائية، 2007.

15. Boukella-bouzouane malika, **Statistique descriptive**, Algerie: Casbah édition, 2001.

16. Grosjean François, Dommergue Jean-Yves, **La statistique en clair**, Paris :2011,

17. Py Bernard, **La statistique sans formule mathématique**, comprendre la logique et maitriser les outils, Paris: Pearson, 2010.

18. Py Bernard, **Statistique descriptive**, 4e éd, Paris :Economica, 2001.

الملاحق

ملحق (01) جدول توزيع t.test

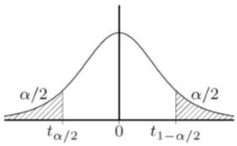
ملحق (02) جدول توزيع χ^2

A.3. Lois de Student

Si T est une variable aléatoire suivant la loi de Student à ν degrés de liberté, la table donne, pour α fixé, la valeur $t_{1-\alpha/2}$ telle que

$$\mathbb{P}\{|T| \geqslant t_{1-\alpha/2}\} = \alpha.$$

Ainsi, $t_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi de Student à ν degrés de liberté.



ν \ α	0,900	0,500	0,300	0,200	0,100	0,050	0,020	0,010	0,001
1	0,1584	1,0000	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567	636,6193
2	0,1421	0,8165	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248	31,5991
3	0,1366	0,7649	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409	12,9240
4	0,1338	0,7407	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041	8,6103
5	0,1322	0,7267	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321	6,8688
6	0,1311	0,7176	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074	5,9588
7	0,1303	0,7111	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995	5,4079
8	0,1297	0,7064	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554	5,0413
9	0,1293	0,7027	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498	4,7809
10	0,1289	0,6998	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693	4,5869
11	0,1286	0,6974	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058	4,4370
12	0,1283	0,6955	1,0832	1,3562	1,7823	2,1788	2,6810	3,0545	4,3178
13	0,1281	0,6938	1,0795	1,3502	1,7709	2,1604	2,6503	3,0123	4,2208
14	0,1280	0,6924	1,0763	1,3450	1,7613	2,1448	2,6245	2,9768	4,1405
15	0,1278	0,6912	1,0735	1,3406	1,7531	2,1314	2,6025	2,9467	4,0728
16	0,1277	0,6901	1,0711	1,3368	1,7459	2,1199	2,5835	2,9208	4,0150
17	0,1276	0,6892	1,0690	1,3334	1,7396	2,1098	2,5669	2,8982	3,9651
18	0,1274	0,6884	1,0672	1,3304	1,7341	2,1009	2,5524	2,8784	3,9216
19	0,1274	0,6876	1,0655	1,3277	1,7291	2,0930	2,5395	2,8609	3,8834
20	0,1273	0,6870	1,0640	1,3253	1,7247	2,0860	2,5280	2,8453	3,8495
21	0,1272	0,6864	1,0627	1,3232	1,7207	2,0796	2,5176	2,8314	3,8193
22	0,1271	0,6858	1,0614	1,3212	1,7171	2,0739	2,5083	2,8188	3,7921
23	0,1271	0,6853	1,0603	1,3195	1,7139	2,0687	2,4999	2,8073	3,7676
24	0,1270	0,6848	1,0593	1,3178	1,7109	2,0639	2,4922	2,7969	3,7454
25	0,1269	0,6844	1,0584	1,3163	1,7081	2,0595	2,4851	2,7874	3,7251
26	0,1269	0,6840	1,0575	1,3150	1,7056	2,0555	2,4786	2,7787	3,7066
27	0,1268	0,6837	1,0567	1,3137	1,7033	2,0518	2,4727	2,7707	3,6896
28	0,1268	0,6834	1,0560	1,3125	1,7011	2,0484	2,4671	2,7633	3,6739
29	0,1268	0,6830	1,0553	1,3114	1,6991	2,0452	2,4620	2,7564	3,6594
30	0,1267	0,6828	1,0547	1,3104	1,6973	2,0423	2,4573	2,7500	3,6460
40	0,1265	0,6807	1,0500	1,3031	1,6839	2,0211	2,4233	2,7045	3,5510
60	0,1262	0,6786	1,0455	1,2958	1,6706	2,0003	2,3901	2,6603	3,4602
80	0,1261	0,6776	1,0432	1,2922	1,6641	1,9901	2,3739	2,6387	3,4163
120	0,1259	0,6765	1,0409	1,2886	1,6577	1,9799	2,3578	2,6174	3,3735
∞	0,1257	0,6745	1,0364	1,2816	1,6449	1,9600	2,3263	2,5758	3,2905

Lorsque $\nu = \infty$, $t_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi normale $\mathcal{N}(0,1)$.

EDITION EL MOTANABY

دار المتنبي للطباعة والنشر