المادة: الإحصاء التربوي

المرحلة الثالثة / قسم رياض الأطفال / كلية التربية للبنات / جامعة ذي قار
العام الدراسي ٢٠١٨-٢٠١٩
مدرس المادة : أ. م. محمد عبد علي

مفردات الإحصاء التربوي

تعريف الإحصاء وانواعه

الاحصاء التربوي

الإحصاء الوصفي

الإحصاء الاستدلالي

المجتمع ،العينة

المتغير وانواعه

البيانات وانواعها

طرائق عرض البيانات

الأعمدة البيانية

المدرج التكراري

المضلع التكراري

الخط المنحني التكراري

تكوين الجداول التكرارية

مقاييس النزعة المركزية

الوسط والوسيط والمنوال

الوسط الوزني

العلاقة بين المقاييس الثلاثة

مقاييس التشتت وانواعها

المدى

الانحراف المتوسط

التباين

الانحراف القياسي

الدرجة المعيارية

معامل الاختلاف

معامل أرتباط بيرسون

معامل أرتباط سبيرمان

الانحدار .

أولا: تعريف الإحصاء :- هو العلم الذي يهتم بجمع البيانات وتبويبها وعرضها وتحليلها واستخراج النتائج والاستدلالات منها لغرض اتخاذ قرارات.

إذ يهتم الإحصاء بطرائق جمع وتمثيل وتفسير البيانات ، فجمع البيانات هي عملية الحصول على القياسات أو التعدادات أو قيم المشاهدات للتجارب التي يجريها الباحث أو الطالب ...، وكلما كان جمع البيانات دقيقا زادت ثقة الدارس بالاعتماد عليها ولا يكون هناك تحليل صحيح للبيانات إذا كان هناك أخطاء في جمع تلك البيانات.

أما تنظيم وعرض البيانات فهو عملية وضع البيانات في جداول منسقة وعرضها بطرائق مناسبة كالأشكال الهندسية والرسوم والبيانات والتوزيعات التكرارية التي سنتعرف عليها لاحقا.

أما تحليل البيانات فهي عبارة عن إيجاد قيم لمقاييس واقترانات معينة تتحدد فيها قيد الدراسة.

أما تفسير النتائج واتخاذ القرارات فهو من ابرز أهداف علم الإحصاء وأكثرها فائدة حيث يشمل معظم الدراسات والنظريات القائمة عليها والتطبيقات العلمية لها.

بمعنى آخر وباختصار تفسير النتائج تتألف من الاستنتاجات التي يتوصل إليها الباحث من تحليل بياناته وهي غالبا ما تكون على شكل تقديرات أو تنبؤات أو قرارات رفض أو قبول للفرضيات الإحصائية.

ومن خلال التعريف السابق للإحصاء يمكن أن نستنتج ما يأتي :-

١- إن الإحصاء لا يتناول بالدراسة مفردة بعينها، لكنه يشمل المجتمعات بالدراسة، والمجتمع يعني (أي تجمع لأفراد ما أو أشياء أو أدوات أو مفردات تجمع بينها صفة أو صفات مشتركة، لكن هذا لا يعني بالضرورة أن تجري دراسة المجتمعات على أساس من الشمول ولكن دراسة مجتمع ما قد تكون على أساس عينة تسحب من هذا المجتمع أو من هذه المجتمعات.

٢- إن الدراسة الراقية للمجتمعات تشمل البحث في أساليب جمع البيانات ووسائل عرضها وتحليلها بهدف الوصول نوع من المعرفة المبنية على أسس رقمية للمجتمعات في موضوع الدراسة.

٣- أن أسلوب الدراسة الإحصائية أسلوب رقمي، وهذا يعني إن كل مجتمع يمكن ان يقاس رقميا سواء اعتمد هذا القياس على أسس كمية أو ترتيبية فان ذلك المجتمع يمكن أن يخضع للدراسة الإحصائية.

٤- إن جوهر علم الإحصاء هو دراسة التغيرات ،فالمجتمعات كاملة التشابه أو التجانس يمكن دراستها من خلال خاصية واحدة منها ولا محل لقيام دراسة إحصائية وهذا ما دفع بعض الباحثين إلى تعريف علم الإحصاء بأنه علم دراسة المتغيرات.

ثانيا:العلاقة بين الإحصاء والقياس

من الممكن إن نوضح باختصار طبيعة العلاقة بين الإحصاء والقياس وهي أن القياس يعني أعطاء معنى كمي لسمات الفرد أو ما يحصل عليه الفرد من سمات وهذا المعنى للقياس يؤدي إلى ضرورة جمع البيانات التي هي قيم المشاهدات التي يلاحظها الباحث في بحوثه ويجمعها الباحث من أفراد عينته الدراسية.

أما اللغة التي تستخدم للتعامل مع هذه البيانات وإعطاء معنى واضح لها وهو الإحصاء.

مما سبق نستنتج أن العلاقة بين الإحصاء وكل من القياس والتقويم قوية جدا؛ ولذلك لان الإحصاء أداة جمع البيانات المتعلقة بمواضيع القياس والتقويم ومن ثم تبويب هذه البيانات وعرضها وتحليلها واستنباط النتائج واتخاذ القرارات بناءً على ذلك.

ثالثا:المجتمع والعينة

تعريف مجتمع البحث: ونعني به بأنه جميع المفردات الظاهرة التي يقوم بدراستها الباحث ، ولكن هل يستطيع الباحث أن يدرس جميع أفراد مجتمع البحث ؟ وهل يمتلك وقتا كافيا لدراسة أفراد هذا المجتمع.

في واقع الأمر أن دراسة مجتمع البحث الأصلي يتطلب وقت طويلا وجهدا شاقا وتكاليف مادية مرتفعة ، ويكفي أن يختار الباحث عينة ممثلة لمجتمع البحث بحيث تحقق أهدافه.

خطوات اختيار عينة البحث:-

1- **تحديد المجتمع الأصلي للدراسة.**

2- **تحديد أفراد المجتمع الأصلي للدراسة (إعداد قائمة بأسماء جميع الأفراد).**

3- **اختيار عينة مماثلة حسب مجتمع البحث.**

٤- اختيار عدد كافي من الأفراد في العينة، العينة تتناسب مع الحجم المناسب من خلال العوامل التالية:

أ- إذا كان المجتمع الأصلي متجانس سوف يسهل عملية اختيار العينة.

ب- إن أسلوب البحث المستخدم يؤثر في اختيار العينة وعليك أن تسأل نفسك عند اختيارك عينة البحث (هل تستخدم أسلوب البحث الوصفي، والأسلوب التجريبي وما نوع التصميم التجريبي المستخدم.

ت- درجة الدقة المطلوبة.

وقبل البدء باختيار عينة الدراسة لابد أن نتعرف على:-

١- ما هي المعلومات المطلوبة.

٢- ما هو الهدف منها.

٣- ما أهميتها.

٤- كيفية استخدامها.

من أنواع العينات:

١- العينة العشوائية:- ومنها العينة العشوائية (البسيطة ، العينة الطبقية، العينة المنتظمة).

٢- أسلوب العينة الغير عشوائية:- ومنها (عينة الصدفة ، العينة الحصصية).

رابعا: طرائق عرض البيانات

بعد ان يقوم الباحث بالحصول على البيانات باستعمال واحدة أو أكثر من وسائل جمع المعلومات (البيانات) كالملاحظة والمقابلة والاستبانة (الاستبيان) والاختبارات الخ... .

وبعد جمع هذه البيانات يصبح من الضروري عرضها بشكل يسهل استعمالها واستخلاص النتائج منها ، وتوجد أكثر من طريقة لعرض هذه البيانات ومنها ما يأتي:-

عرض البيانات إنشائيا :- وبها يصف الباحث بياناته بجمل إنشائية توضح النتائج التي استخلصها ، ومنها:-

أ- عرض البيانات في صورة جداول إحصائية.

ب- عرض البيانات في صورة خريطة او رسم بياني.

ت- عرض البيانات الإحصائية ملخصة في صورة رقم أو نسبة باستخدام مقياس أخر من المقاييس الإحصائية المعروفة في الوسط الحسابي أو الانحراف المعياري.

ث- عرض البيانات باستخدام اكثر من طريقة واحدة كأن يستخدم الباحث الجداول الاحصائية او الرسوم البيانية. ويمكن عرض البيانات بالطرارق الاتية:

***عرض البيانات بالتوزيعات التكرارية (التوزيع التكراري)**

تعد طريقة التوزيع التكراري من طرائق عرض البيانات الوصفية والكمية وتهدف إلى تبسيط العمليات الإحصائية وذلك بتبويبها بصورة مناسبة يسهل أجرائها بسرعة ودقة . مثال على ذلك:-

إذا أردنا ان نحسب مرات تكرار كل عدد او درجة من الإعداد والدرجات التالية:-

٢,٤,٢,٦,٢,٤,٥,٧,٦,٢,٤,٤,٢,٧ فإننا نرى إن الدرجة (٢) تكررت (٥) مرات والدرجة (٤) قد تكررت (٤) مرات وهكذا لبقية الأرقام ، ويمكن تلخيص هذه البيانات على النحو التالي:-

الدرجة	العلامات التكرارية	عدد التكرارت	مجموع الدرجات
٢	~~////~~	٥	٢x٥=١٠
٤	////	٤	٤x٤=١٦

٥	/	١	٥=١x٥
٦	//	٢	١٢=٦x٢
٧	//	٢	١٤=٢x٧

* عرض البيانات عن طريق الفئات التكرارية (الفئة التكرارية)

١- إيجاد عدد الفئات:-

عند التفكير في إيجاد عدد الفئات أو تحديد أطوالها يجب أن نأخذ بعين الاعتبار الأمرين التاليين:-

- **إن جميع الإفراد الذين سيقعون في فئة معينة سيجري اعتبارهم متساويين بصرف النظر عن الفروق البسيطة التي ربما**

تكون بينهم وإنهم جميعا سيحظون القيمة العددية نفسها والتي تساوي قيمة متوسط تلك الفئة .

- من اجل الاختصار في حجم الجدول يفضل ان يكون عدد الفئات قليلا بشرط ان لا يتعارض ذلك مع ما ورد في النقطة السابقة.

٢- تعيين طول الفئة:-

عندما نفرغ من تحديد عدد الفئات التي سيقسم المدى العام تكون بذلك قد خطونا الخطوة الرئيسة في تحديد طول الفئة او اتساعها بعد ذلك نقسم طول المدى العام على عدد الفئات بالتساوي فينتج معنا طول الفئة الواحدة ، وفي حالة خارج القسمة عددا كسريا فمن المستحسن تقريبه الى اقرب عدد صحيح حتى لو ترتب على ذلك زيادة أو نقصان طفيفين في عدد الفئات، مثال على ذلك:-

لو كانت اعلى قيمة واصغرها هي ١٦,٩٢ على الترتيب فان المدى العام يساوي: ٩٢-١٦=٧٦ درجة.

ولو كانت عدد الفئات المطلوبة هي ١٥ فئة فان طول الفئة يساوي: ٧٦÷١٥=٥,٥ طول الفئة، وبالتقريب الى اقرب عدد صحيح يصبح طول الفئة = ٥ درجة.

٣- تعيين حدي الفئة:-

مثال/ لو أعطي امتحان موضوعي لمجموعة من الطلبة مؤلف من ٦٠ سؤال والإجابة تكون على كل سؤال أما صحيحة أو خاطئة فان (الإجابة الصحيحة= ١، الإجابة الخاطئة= صفر)، وبهذه الدرجات التي يمكن ان يأخذها أي طالب لأتخرج عن واحدة من الدرجات التالية {١,٢,٣، ... ، ٦٠} فاذا كانت الفئات الممثلة بدرجات مجموعة الطلبة على هذا الامتحان هي كما يلي:-

١- ٥ واقل من ١٠

٢- ١٠ واقل من ١٥

٣- ١٥ واقل من ٢٠

٤- ٢٠ واقل من ٢٥

٥- ٢٥ واقل من ٣٠ الخ...

فمن الممكن إعادة كتابة تها بشكل اكثر تحديد على النحو التالي:-

١- ٥ ـ ٩ مركز الفئة = ٥+٩\٢= ١٤\٢=
٧

٢- ١٠- ١٤ مركز الفئة = ١٠+١٤\٢=
٢٤\٢=١٢

٣- ١٥- ١٩ مركز الفئة =
١٥+١٩\٢=٣٤\٢ =١٧ الخ...

نستنتج مما سبق ان حدي الفئة الاولى (٥-٩) ومفرداتها هي (٩,٨,٧,٦,٥) وبذلك يكون مركزها (٧) والشيء نفسه يقال على بقية الفئة الثانية (١٠-١٤) ومفرداتها (١٤,١٣,١٢,١١,١٠) مما يجعل مركزها ١٢.

ملاحظة:- " جمع الحد الأعلى مع الحد الأدنى ÷٢= مركز الفئة ".

الاحصاء الوصفي : الإحصاء الذي يصف الظاهرة المقاسة في عينة من السلوك وصفاً بيانياً أو كمياً دون تعميم النتائج التي يتم التوصل إليها

الاحصاء الاستدلالي : الاحصاء الذي نستدل من خلاله على سمة المجتمع الذي سحبت منه عينة الدراسة

المجتمع : المجموعة الكلية التي تضم جميع المفردات المستهدفة بالبحث أو الدراسة والمشتركة في صفة أو صفات معينة.

العيِّنة : المجموعة الجزئية من المفردات التي اختيرت من مجتمع الدراسة

العينة العشوائية : العينة التي يكون لكل مفردة من المجتمع نفس فرصة الظهور ضمن أفرادها

المتغير : خاصية أو سمة يمكن أن تأخذ أكثر من قيمة مثل الذكاء ، التحصيل ، الطول ، الجنس

الثابت : الخاصية التي تأخذ قيمة ثابتة مثل درجة غليان الماء ، كثافة النحاس

المتغير المتصل: متغير يأخذ أي قيمة في مدى معين مثل الدرجات المحصورة بين (٢٠ - ٥٠) فإن أي درجة صحيحة أو كسرية تقع في هذا المدى يمكن أن تسند إلى ذلك المتغير

المتغير المنفصل: متغير يأخذ قيم صحيحة محددة في مدى معين مثل الجنس إما ١ أو ٢ فلا يوجد جنس يأخذ القيمة ١,٥مثلاً وكذلك القيم التي تعبر عن المؤهل أو الرتبة أو اللون

المتغير الكمي (الرقمي) : نوع من المتغيرات يأخذ قيم رقمية يمكن اجراء عمليات حسابية عليها مثل العمر ، الدرجة في الاختبار

المتغير الأسمي : متغير يصف أو يسمي عناصر في المجتمع لايمكن ترتيبها أو إجراء عمليات حسابية عليها مثل العنوان ، الجنس ، رقم الهاتف. رقم الشارع

المتغير الرتبي : متغير يصف عنصر في المجتمع يمكن وضع ترتيب منطقي لمفرداته ولا يمكن إجراء العمليات الحسابية عليه مثل : درجة المتسابق ، ترتيب الطالب في الاختبار" الأول ، الثاني ،الثالث،...............الخ"

البيانات : مجموعة القيم أو المفردات التي تسند إلى المتغيرات الداخلة في الدراسة

المعامل : قيمة عددية تصف نتيجة المعالجة التي جرت على البيانات التي تم جمعها

الإحصائي: قيمة عددية تصف السمة المقاسة من العينة المستهدفة بالدراسة.

المعلم : قيمة احصائية تصف المجتمع المسحوب منه عينة الدراسة

مقاييس السوسيوغرام: مقييس تصف العلاقات الاجتماعية بين الأفراد على شكل مخططات سهمية

مقاييس القدرات : مقاييس تحدد درجة ذكاء الأفراد وقدرتهم على التعلم

مقاييس الذكاء : مقاييس تحدد قدرة الفرد على التعلم والتفكير المنطقي وحل المشكلات والتكيف مع بيئته

مقاييس الاستعداد : مقاييس تحدد مدى قابلية الفرد للتعلم أو اكتساب سلوك أو مهارات معينة

بيانات مبوبة : وهي البيانات التي المرتبة والمنظمة في جدول تكراري.

بيانات غير مبوبة : البيانات في صورتها الخام قبل إجراء أي ترتيب أو تنظيم أو تصنيف لها.

طول الفئة: حاصل قسمة المدى للتوزيع على عدد الفئات لأقرب رقم صحيح

عدد الفئات : حاصل قسمة المدى للتوزيع على طول الفئة لأقرب رقم صحيح

مقياس اسمي :التعبيرعن السمات المقاسة بأعداد مجردة لا تحمل قيماً كمية وتستخدم للتسمية أو الترميز للدلالة على السمة المقاسة

مقياس ترتيبي : التعبير عن السمات المقاسة بأعداد تدل على ترتيب مفرداتها تصاعدياً أو تنازلياً

مقياس فئوي " فترات متساوية" : التعبير عن السمات المقاسة بأعداد تتساوى المسافات بينها بدءاً بصفر افتراضي لا تنعدم عنده السمة المقاسة

مقياس نسبي : التعبير عن السمات المقاسة بأعداد تتساوى المسافات بينها بدءاً بصفر مطلق " حقيقي"حيث تنعدم عنده الصفة المقاسة

مركز الفئة : حاصل قسمة الفرق بين الحد الأعلى والحد الأدنى لكل فئة على ٢ .

المعيار Norm: هو أساس الحكم على الظاهرة موضوع التقويم داخل الظاهرة، وليس من خارجها وليس من عينة أخرى أو أفراد آخرين وتأخذ طبيعة كمية فى أغلب الأحيان وتتحدد باستخدام بعض أساليب الإحصاء وفى ضوء الخصائص الواقعية للظاهرة.

تدريب :

تم تطبيق مقياس للمهارات الاجتماعية على ٦٠ تلميذ بالمدرسة فكانت درجاتهم كما هي موضحة بالجدول التالي:

الدرجة	٩	١٣	١٥	١٨	٢٠	٢٢	٢٥	٢٧	٢٨	٣٠
التكرار	٢	٣	٥	٨	٩	١٣	١٠	٦	٣	١

من بيانات الجدول أحسب :

- المتوسط الحسابي. ، الانحراف المعياري. ، بين ما إذا كان التوزيع اعتدالي أم لا .

- إذا تم اعتبار درجة القطع هي م - ١ع فكم طالب يمكن اعتبارهم يعانون من نقص في المهارات الاجتماعية؟

تدريب (٦) : الجدول التالي يوضح درجات مجموعة من التلاميذ في الخجل أكمل الجدول :

الفئات	التكرار	الحدود الحقيقة	منتصفات الفئات	التكرار المتجمع الصاعد	التكرار المتجمع النازل
٠ - ٤	٢				
٥ ــ ٩	٥				
١٠ ــ ١٤	٨				
١٥ ــ ١٩	١٠				
٢٠ ــ ٢٤	١٣				
٢٥ ــ ٢٩	١٢				
٣٠ ــ ٣٤	١٠				
٣٥ ــ ٣٩	٧				

٤٠ – ٤٤	٢				
٤٥ – ٥٠	١				

١- من بيانات الجدول السابق أحسب متوسط الدرجات، الانحراف المعياري ، الدرجة المنوالية.

٢- إذا أراد المرشد النفسي أن يخضع التلاميذ أصحاب أعلى ٢٥ % من الدرجات إلى برنامج إرشادي لخفض الخجل فكيف يمكن تحديد هؤلاء التلاميذ.

٣- كم تلميذ تقل درجتهم في الخجل عن ٢٥ ، وكم طالب تزيد درجتهم في الخجل عن ٣٤.

٤- ارسم المنحنى التكراري لدرجات التلاميذ في الخجل.

تدريب : تم قياس نسب الذكاء لدى مجموعة من تلاميذ الصف الرابع الابتدائي وتحصيلهم في اختبار للغة العربية فكانت درجاتهم كما هي موضحة بالجدول التالي:

م	الذكاء	التحصيل	م	الذكاء	التحصيل
١	١١٠	٩	١١	١٠٣	٧
٢	١١٥	١٤	١٢	١٠٩	٦
٣	١٠٥	١٢	١٣	١١٠	١٥
٤	٩٥	١١	١٤	١٢١	١٧
٥	٩٨	١٤	١٥	٩٠	١٣

٦	١٠٠	١٥	١٦	٩١	١١
٧	١٠٧	١٠	١٧	٩٤	١٣
٨	٩٣	١٣	١٨	١٠٤	١٧
٩	٩٩	١٢	١٩	١٠٠	١٣
١٠	١١٢	٨	٢٠	١٠١	٧

فإذا علمت أن التلميذ صاحب المستوى المتوسط في الذكاء أو أعلى منه وفي نفس الوقت تحصيله أقل من المتوسط في اللغة العربية لديه صعوبات في تعلم اللغة العربية فمن هم هؤلاء التلاميذ حتى يمكن إعداد جلسات علاجية لهم.

مقاييس النزعة المركزية

Central Tendency measurement

مقدمة

في كثير من النواحي التطبيقية يكون الباحث في حاجة إلى حساب بعض المؤشرات التي يمكن الاعتماد عليها في وصف الظاهرة من حيث القيمة التي تتوسط القيم أو تنزع إليها القيم ، ومن حيث التعرف على مدى تجانس القيم التي يأخذها المتغير، وأيضا ما إذا كان هناك قيم شاذة أم لا . والاعتماد على العرض البياني وحدة لا يكفى ، ولذا يتناول

هذا الفصل، والذي يليه عرض بعض المقاييس الإحصائية التي يمكن من خلالها التعرف على خصائص الظاهرة محل البحث، وكذلك إمكانية مقارنة ظاهرتين أو أكثر ، ومن أهم هذه المقاييس ، مقاييس النزعة المركزية والتشتت .

مقاييس النزعة المركزية

تسمى مقاييس النزعة المركزية بمقاييس الموضع أو المتوسطات ، وهى القيم التي تتركز القيم حولها ، ومن هذه المقاييس ، الوسط الحسابي ، والمنوال ، والوسيط ، والوسط التوافقي وفيما يلي عرض لأهم هذه المقاييس

١- الوسط الحسابي Arithmetic Mean

من أهم مقاييس النزعة المركزية ، وأكثرها استخداما في النواحي التطبيقية ، ويمكن حسابه للبيانات المبوبة وغير المبوبة ، كما يلي :

أولا: الوسط الحسابي للبيانات غير المبوبة

يعرف الوسط الحسابي بشكل عام على أنه مجموع القيم مقسوما على عددها . فإذا كان لدينا n من القيم ، ويرمز لها بالرمز : $x_1, x_2, \ldots, x_n$.

فإن الوسط الحسابي لهذه القيم ، ونرمز له بالرمز $\bar{x}$ يحسب بالمعادلة التالية :

$$\text{الوسط الحسابي} = \frac{\text{مجموع القيم}}{\text{عدد القيم}}$$

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

حيث يدل الرمز Σ على المجموع .

مثـال

فيما يلي درجات 8 طلاب في مادة الإحصاء .

40 36 40 35 37 42 32 34

والمطلوب إيجاد الوسط الحسابي لدرجة الطالب في الامتحان .

الحـل

لإيجاد الوسط الحسابي للدرجات تطبق المعادلة رقم (٣-١) كما يلي:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$
$$= \frac{34+32+42+37+35+40+36+40}{8} = \frac{296}{8} = 37$$

أي أن الوسط الحسابي لدرجة الطالب في اختبار مادة الإحصاء يساوي 37 درجة

ثانيا: الوسط الحسابي للبيانات المبوبة

من المعلوم أن القيم الأصلية ، لا يمكن معرفتها من جدول التوزيع التكراري ، حيث أن هذه القيم موضوعة في شكل فئات ، ولذا يتم التعبير عن كل قيمة من القيم التي تقع داخل حدود الفئة بمركز هذه الفئة ، ومن ثم يؤخذ في الاعتبار أن مركز الفئة هو القيمة التقديرية لكل مفردة تقع في هذه الفئة.

فإذا كانت k هي عدد الفئات ، وكانت $x_1, x_2, ..., x_k$ هي مراكز هذه الفئات، $f_1, f_2, ..., f_k$ هي التكرارات ، فإن الوسط الحسابي يحسب بالمعادلة التالية:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + ... + x_k f_k}{f_1 + f_2 + ... + f_k} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}$$

مثـال

الجدول التالي يعرض توزيع 40 تلميذ حسب أوزانهم .

فئات الوزن	32-34	34-36	36-38	38-40	40-42	42-44
عدد التلاميذ	4	7	13	10	5	1

والمطلوب إيجاد الوسط الحسابي.

الحــل

لحساب الوسط الحسابي باستخدام المعادلة رقم (٣-٢) يتم إتباع الخطوات التالية :

إيجاد مجموع التكرارات $\sum f$. **٢- حساب مراكز الفئات x .**

ضرب مركز الفئة في التكرار المناظر له $(x f)$، وحساب المجموع $\sum xf$

حساب الوسط الحسابي بتطبيق المعادلة رقم (٣-٢) .

فئات	التكرارات	مراكز	$x\ f$

	الفئات x	f	الوزن (C)
$4\times 33=132$	$(32+34)\div 2=33$	4	32-34
$7\times 36=252$	٣٦	7	35-37
$13\times 39=507$	٣٩	13	38-40
$10\times 42=420$	٤٢	10	41-43
$5\times 45=225$	٤٥	5	44-46
$1\times 48=48$	٤٨	1	47-49
1584		40	المجموع

إذا الوسط الحسابي لوزن التلميذ هو :

$$\overline{x}=\frac{\sum_{i=1}^{6} x_i f_i}{\sum_{i=1}^{6} f_i}=\frac{1584}{40}=39.6\ k.g$$

أي أن متوسط وزن التلميذ يساوي 39.6 k.g

٢- الوسيط Median

هو أحد مقاييس النزعة المركزية، والذي يأخذ في الاعتبار رتب القيم ، ويعرف الوسيط بأنه القيمة التي يقل عنها نصف عدد القيم $(n/2)$ **، ويزيد عنها النصف الآخر** $(n/2)$**، أي أن 50% من القيم أقل منه، 50% من القيم أعلى منه. وفيما يلي كيفية حساب الوسيط في حالة البيانات غير مبوبة ، والبيانات المبوبة.**

أولا: الوسيط للبيانات غير المبوبة

لبيان كيف يمكن حساب الوسيط للبيانات غير المبوبة ، نتبع الخطوات التالية:

ترتب القيم تصاعديا .

تحديد رتبة الوسيط، وهي : رتبة الوسيط = $\left(\frac{n+1}{2}\right)$

إذا كان عدد القيم (n) فردي فإن الوسيط هو:

$$\boxed{\text{الوسيط} = \text{القيمة رقم} \left(\frac{n+1}{2}\right)}$$

إذا كان عدد القيم (n) زوجي، فإن الوسيط يقع بين القيمة رقم $(n/2)$، والقيمة رقم $((n/2)+1)$، ومن ثم يحسب الوسيط بتطبيق المعادلة التالي:

$$\text{الوسيط} = \frac{\text{القيمة رقم}\left(\frac{n}{2}\right) + \text{القيمة رقم}\left(\frac{n}{2}+1\right)}{2}$$

مثـال

تم تقسيم قطعة أرض زراعية إلى 17 وحدة تجريبية متشابهة ، وتم زراعتها بمحصول القمح ، وتم استخدام نوعين من التسميد هما : النوع (a) وجرب على 7 وحدات تجريبية ، والنوع (b) وجرب على 10 وحدات تجريبية ، وبعد انتهاء الموسم الزراعي ، تم تسجيل إنتاجية الوحدة بالطن / هكتار ، وكانت على النحو التالي :

النوع (a)	1.2	2.75	3.25	2	3	2.3	1.5

النوع (b)	4.5	1.8	3.5	3.75	2	2.5	1.5	4	2.5	3

والمطلوب حساب وسيط الإنتاج لكل نوع من السماد المستخدم، ثم قارن بينها.

الحـل

أولا : حساب وسيط الإنتاج للنوع الأول (a)

ترتيب القيم تصاعديا :

قيمة الوسيط

الإنتاج	1.2	1.5	2	2.3	2.75	3	3.25
الرتبة	1	2	2	4	5	6	7

رتبة الوسيط

عدد القيم فردى $(n=7)$

إذا رتبة الوسيط هي: $((n+1)/2=(7+1)/2=4)$ **.**

ويكون الوسيط هو القيمة رقم 4 ، أي أن وسيط الإنتاج للنوع a هو:

$Me_a = 2.3$ **طن / هكتار**

ثانيا : حساب وسيط الإنتاج للنوع الثاني (b) :

ترتيب القيم تصاعديا .

قيمة الوسيط = $\frac{2.5+3}{2}$

2.75

الإنتاج	1.5	1.8	2	2.5	2.5	3	3.5	3.75	4	4.5
الرتبة	1	2	3	4	5	6	7	8	9	10

5.5

رتبة الوسيط

عدد القيم زوجي $(n=10)$ **إذا**

رتبة الوسيط هي : $((n+1)/2=(10+1)/2=5.5)$**.**

الوسيط = الوسط الحسابي للقيمتين الواقعتين في المنتصف (رقم 5 ، 6) .

$$Me_b=\frac{2.5+3}{2}=2.75$$ **طن / هكتار**

وبمقارنة النوعين من السماد ، نجد أن وسيط إنتاجية النوع (a) أقل من وسيط إنتاجية النوع (b) ، أي أن : $Me_b>Me_a$ **.**

ثانيا: الوسيط للبيانات المبوبة

لحساب الوسيط من بيانات مبوبة في جدول توزيع تكراري ، يتم إتباع الخطوات التالية .

تكوين الجدول التكراري المتجمع الصاعد .

تحديد رتبة الوسيط : $\left(\frac{n}{2}\right)=\left(\frac{\sum f}{2}\right)$

ويحسب الوسيط ، بتطبيق المعادلة .

$$Me=L_k+\frac{h_k}{f_k}\left[\frac{\sum f_i}{2}-f'_{k-1}\right]$$

حيث أن :

L_k : هي الحد الأدنى للفئة الوسيطية المقابلة لأكبر تكرار.

h_k : طول الفئة الوسيطية

f_k : تكرار الفئة الوسيطية.

f'_{k-1} : التكرار المتجمع الصاعد السابق للفئة الوسيطية

3-المنوال Mode

يعرف المنوال بأنه القيمة الأكثر شيوعا أو تكرارا ، ويكثر استخدامه في حالة البيانات الوصفية ، لمعرفة النمط (المستوى) الشائع، ويمكن حسابه للبيانات المبوبة وغير المبوبة كما يلي:

أولا: حساب المنوال في حالة البيانات غير المبوبة

$\boxed{\text{المنوال (Mod) = القيمة (المستوى) الأكثر تكرارا}}$ (٣-١٢)

ثانيا: حساب المنوال في حالة البيانات المبوبة.

$$Mod = L_k + \frac{(f_k - f_{k-1})}{(f_k - f_{k-1}) + (f_k - f_{k+1})} * h_k$$

حيث أن :

Lk : الحد الأدنى لفئة المنوال المقابلة لأكبر تكرار.

f_k **: تكرار الفئة المنوالية**

f_{k-1} **: التكرار السابق للفئة المنوالية**

f_{k+1} **: التكرار اللاحق للفئة المنوالية**

h_k **: طول فئة المنوال**

٤- الوسط التوافقي Harmonic mean

هو مقلوب الوسط الحسابي لمقلوب القيم و يتم حسابه وفق الصيغة التالية:

أولا: حساب الوسط التوافقي في حالة البيانات غير المبوبة

$$\overline{H} = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}$$

مثال: أوجد الوسط التوافقي للبيانات التالية:

2 , 5 , 3 , 4 , 7 , 8 , 8

$$\sum \frac{1}{X_i} = \frac{1}{2} + \frac{1}{5} + \frac{1}{3} + \frac{1}{4} + \frac{1}{7} + \frac{1}{8} + \frac{1}{8} = 1.68$$

$$\therefore \overline{H} = \frac{7}{1.68} \; 4.17$$

ثانيا: حساب الوسط التوافقي في حالة البيانات المبوبة

الوسط التوافقي لمراكز الفئات X1, X2,… ,Xn والمرجح بالتكرارات المناظرة

f1,f2,…,fn يكون

$$\overline{H} = \frac{\Sigma f_i}{\Sigma \frac{f_i}{X_i}}$$

مقاييس التشتت

Measures of Dispersion

تناولنا في محاضرات سابقة المتوسطات أو مقاييس النزعة المركزية والتي تعبر عن المستوى العام للظاهرة التربوية او السياسية أو الاجتماعية محل البحث. ولكن ترى هل هذا يعتبر كافياً لوصف البيانات وتحليلها كمياً لنصل بالتالي إلى فهم أكثر وضوحاً للظاهرة محل الدراسة ؟ للإجابة على هذا السؤال نسوق المثال التالي :

مثال ١ :

نفترض أن لدينا شريحتين من شرائح المجتمع تعيشان في منطقتين مختلفتين وكانت دخولهم الأسبوعية (بالدولار الأمريكي) هي كما يلي :

المجموعة الأولى A :

70	75	71	75	74	76	73	78

المجموعة الثانية B :

99	56	80	100	29	70	65	93

وبحساب الوسط الحسابي للشريحتين المذكورتين في كل من المجموعتين

– حسب ما أوضحناه في الفصل السابق – فإن الوسط الحسابي لدخول المجموعة الأولى :

$$\overline{X}_A = \frac{70+75+71+75+74+76+73+78}{8}$$

$$\overline{X}_A = \frac{592}{8} = 74 \text{ دولاراً}$$

أي أن الوسط الحسابي لدخل المجموعة A هو 74 دولاراً.

وكذلك بالنسبة للمجموعة B فإن الوسط الحسابي لدخولها هو :

$$\overline{X}_B = \frac{99+56+80+100+29+70+65+93}{8}$$

$$\overline{X}_B = \frac{592}{8} = 74 \quad \text{دولاراً}$$

والوسط الحسابي لدخل المجموعة B هو أيضاً 74 دولاراً.

ومعنى هذا أن المستوى العام لدخل الأفراد في المجموعتين واحد ويساوي 74 دولاراً. ولكن بإمعان النظر في دخول المجموعة الأولى نجد أنها متجانسة إلى حدٍ كبير أي أنها قريبة جداً من بعضها أو من الوسط الحسابي والذي يساوي 74 دولاراً. وهنا نقول أن تشتت الدخول قليل أو صغير. بينما دخول المجموعة الثانية غير متجانسة فهي بعيدة عن بعضها، أو عن الوسط الحسابي بشكل كبير. وهنا نقول أن تشتت دخول المجموعة الثانية كبير فهي أقل تجانساً (أو أكثر تشتتاً) من المجموعة الأولى.

(الفرق بين أكبر دخل وأصغر دخل في المجموعة الأولى: ٧٨-٧٠ =٨ دولارات فقط، بينما الفرق بين أكبر دخل وأصغر دخل في المجموعة الثانية: ١٠٠- ٢٩-٧١ دولاراً)

نستنتج من هذا المثال بأن المتوسط ليس كافياً لتوصيف البيانات أو تحليلها كمياً. فها هو المتوسط واحد في المجموعتين ورغم ذلك فإن البيانات تختلف تماماً في مدى تشتتها (أو تجانسها). وهنا تبرز الحاجة إلى مقاييس كمية أو إحصائية ليقاس مدى تشتت البيانات ولتعطي للباحث بالتالي صورة أكثر وضوحاً وصدقاً للظاهرة السياسية محل الدراسة. وفيما يلي نتناول بعض مقاييس التشتت التي تخدم الغرض وهي :

مقاييس التشتت

المدى، والتباين والانحراف المعياري، ومعامل اختلاف

٢,٥ المدى The Range

يعرّف المدى بأنه " الفرق بين أكبر قيمة وأصغر قيمة ".

فمدى المجموعة A في المثال رقم (١) هو : 70 – 78 = 8

ومدى المجموعة B في نفس المثال هو : = 29 – 100 71

وواضح تماماً أن مدى المجموعة الثانية أكبر بكثير جداً من مدى المجموعة الأولى مما يعني أن المجموعة الثانية أكثر تشتتاً من المجموعة الأولى.

ومن تعريف المدى يتضح أنه مقياس بسيط جداً وأن كل المطلوب معرفته هو أكبر قيمة وأصغر قيمة فقط. ومن هذا التعريف نخرج بالملاحظات التالية :

١ – أن المدى يعتمد في حسابه على قيمتين فقط (أكبر قيمة وأصغر قيمة) وأنه يهمل بالتالي باقي القيم.

٢ – أنه لا يقيس تشتت البيانات عن متوسطها، فهو لا يشير – من قريب أو من بعيد – إلى متوسط البيانات (أو مركزها).

٣ – أنه حساس جداً لأي قيمة شاذة أو متطرفة والمثال التالي يوضح هذه النقطة بشيء من التفصيل.

مثال (٢) :

إذا كانت أعمار أعضاء السلطة التشريعية في بلد ما هي :

72 78 20 74 73 75 70

فإن المدى في هذه الحالة هو : 58 = 20 – 78

وهنا نلاحظ وجود قيمة شاذة بالنسبة لباقي القيم وهي 20 وهي أصغر قيمة. وإذا أهملت هذه القيمة (أو لم تكن موجودة أصلاً) لكان المدى :

78 – 70 = 8

وهذا يعني أن وجود قيمة شاذة (20) رفعت قيمة المدى من 8 سنوات إلى 58 سنة. وهذا يوضح مدى حساسية هذا المقياس للقيم الشاذة (أو المتطرفة).

ولكل هذه الأسباب – أو الملاحظات – فإن كثيراً من الإحصائين والباحثين لا يعتمدون كثيراً على المدى كمقياس للتشتت. ويستخدم فقط إذا كان المطلوب فكرة سريعة أو عامة (وليست دقيقة) عن مدى تشتت البيانات.

التباين The Variance :

يعتبر التباين أحد مقاييس التشتت المهمة لأنه من ناحية يأخذ جميع القيم في الاعتبار عند حسابه، ومن ناحية أخرى لأنه يقيس التشتت عن الوسط الحسابي للقيم، هذا بالإضافة إلى أنه تسهل معالجته رياضياً، وأنه يدخل في تكوين عدد من المقاييس والاختبارات الإحصائية المهمة.

والفكرة الأساسية للتباين هي حساب إنحرافات جميع القيم عن وسطها الحسابي (أي حساب الفرق بين كل قيمة والوسط الحسابي)، وسوف نجد أن بعض القيم أكبر من الوسط فتكون الفروق (أو الإنحرافات) بالموجب، والبعض الأخر أصغر من الوسط فتكون الفروق (أو الإنحرافات) بالسالب. ودائماً يكون مجموع هذه الإنحرافات مساوياً للصفر. ويكون الحل هنا إما إهمال الإشارات السالبة أو تربيع هذه الإنحرافات. وإهمال الإشارات السالبة ليس له مبرر رياضي، فيكون الحل هو تربيع هذه الإنحرافات، ثم نحسب متوسط الانحرافات المربعة فنحصل على التباين. أي أن التباين يعرف كما يلي:

التباين : هو متوسط مربعات إنحرافات القيم عن وسطها الحسابي. والمثال التالي يوضح كيفية حساب التباين.

مثال (٣) :

أحسب تباين دخول الشريحية الاجتماعية التالية : مثال رقم (١) :

78 76 76 74 75 71 75 70

الحل :

١ – نحسب أولاً الوسط الحسابي للدخول كما يلي :

$$\overline{X} = \frac{592}{8} = 74$$

٢ – نحسب إنحرافات القيم عن الوسط الحسابي (أي الفروق بينها وبين الوسط الحسابي) كما يلي :

4 + , 1 - , 2 + , 0 , 1 + , 3 - , 1 + , 4 -

(لاحظ أن مجموع الانحرافات يساوي صفر)

٣ – نربع هذه الانحرافات كما يلي :

16 1 4 0 1 9 1 16

٤ – التباين يساوي متوسط هذه المربعات، أي يساوي :

$$\frac{16+1+9+1+0+4+1+16}{8}=\frac{48}{8}=6$$

٥,٤ الإنحراف المعياري : Standar Deviation

نلاحظ أن تمييز التباين سيكون " دخول مربعة " وبصفة عامة " وحدات مربعة " لأنه يتم تربيع الانحرافات (أو الفروق). لذلك فإنه يؤخذ الجذر التربيعي الموجب للتباين فنحصل على ما يسمى " الإنحراف المعياري " والذي سيكون تمييزه هو نفس الوحدات. ففي هذا المثال يكون الانحراف المعياري للدخول هو :

دولاراً $\sqrt{6}=2.449$

والانحراف المعياري – عادة – يستخدم بدلاً من التباين كأهم مقياس للتشتت. فالتباين ما هو إلا مربع الانحراف المعياري، والانحراف المعياري ما هو إلا الجذر التربيعي الموجب للتباين.

تعريف الانحراف المعياري : الانحراف المعياري هو. الجذر التربيعي الموجب للوسط الحسابي لمربعات انحرافات القيم عن وسطها السحابي. أي هو الجذر التربيعي الموجب للتباين. وعادة يرمز للانحراف المعياري للعينة بالرمز S ولتباين العينة بالرمز S2 وللانحراف المعياري للمجتمع بالرمز اللاتيني σ (سيجما) ولتباين المجتمع بالرمز σ^2.

ويمكن التعبير عن كل من التابين والانحراف المعياري بالرموز كما يلي :

إذا فرضنا أن قيم العينة هي :

X1 , X2 , ... , Xn (أي عددها n) فإن الخطوات تكون كما يلي :

١ ــ حساب الوسط الحسابي أي حساب $\overline{X}$ حيث :

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

٢ ــ حساب إنحرافات القيم عن الوسط الحسابي، أي :

$$X_1 - \overline{X}, X_2 - \overline{X}, ..., X_n - \overline{X}$$

٣ ــ تربيع هذه الانحرافات، أي :

$$(X_1 - \overline{X})^2, (X_2 - \overline{X})^2, ..., (X_n - \overline{X})^2$$

٤ – التباين هو الوسط الحسابي لهذه المربعات، أي :

$$S^2 = \frac{(X_1 - \overline{X})^2 +, (X_2 - \overline{X})^2 ,+ ... +, (X_n - \overline{X})^2}{n}$$

والتي يمكن كتابته (بإختصار البسط) كما يلي :

تباين العينة

$$\boxed{S^2 = \frac{\sum_{i=1}^{n} (X_I - \overline{X})^2}{n}}$$

٥ – الانحراف المعياري هو الجذر التربيعي الموجب للتباين، أي :

الانحراف المعياري للعينة

$$S = \sqrt{\frac{\sum (X_I - \overline{X})^2}{n}}$$

وبالطريقة نفسها إذا كانت لدينا قيم المجتمع الإحصائي. فإذا كانت قيم المجتمع هي : X1 , X2 , ... , XN أي عددها N فإن قانوني التباين والانحراف المعياري للمجتمع – بافتراض أن الوسط الحسابي للمجتمع هو μ (ميو) – يمكن كتابتها كما يلي:

تباين المجتمع

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

الانحراف المعياري للمجتمع

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

ملاحظة مهمة :

يمكن حساب كل من التباين والانحراف المعياري بطريقة مختصرة – وتعطي النتائج نفسها بطبيعة الحال – كما يلي :

*** تباين العينة بالطريقة المختصرة :**

$$\boxed{S^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

*** الانحراف المعياري للعينة بالطريقة المختصرة :**

$$\boxed{S = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}}$$

وفي الطريقة المختصرة فإن كل المطلوب معرفته لحساب التباين أو الإنحراف المعياري هو $\sum x$ (أي مجموع القيم)، $\sum x^2$ (أي مجموع مربعات القيم) ثم التعويض في القانون.

مثال (٤) :

أحسب التباين والإنحراف المعياري بالطريقة المختصرة لبيانات المثال رقم (٣).

الحل :

يمكن تنظيم الحل في الجدول التالي :

X	X2
70	4900
75	5625
71	5041
75	5625
74	5476
76	5776
73	5329
78	6084
$\sum x = 592$	$\sum x^2 = 43856$

ويكون التباين (حيث n = 8) هو :

$$S^2 = \frac{\sum x^2}{n} = (\frac{\sum x}{n})^2$$
$$= \frac{43856}{8} = \left(\frac{592}{8}\right)^2$$
$$= 5482 - 5476$$
$$S^2 = 6$$

وبالتالي فإن الإنحراف المعياري هو : $S = \sqrt{6} = 2.449$ وهي النتائج نفسها التي حصلنا عليها سابقاً ولكن بطريقة أبسط.

بعض خصائص الإنحراف المعياري :

١ – قيمة الانحراف المعياري دائماً موجبة أو أكبر من أو تساوي صفر. فأقل قيمة تساوي الصفر (وذلك عندما تكون جميع القيم متساوية، وفي هذه الحالة لا توجد فروق أو إنحرافات بينها وبين الوسط الحسابي وبالتالي لا يوجد أي تشتت بين القيم، وبالتالي فإن قيمة الإنحراف المعياري في حالة تساوي جميع القيم تساوي الصفر).

٢ – كلما كان التشتت كبيراً حول الوسط كلما كان الانحراف المعياري كبيراً، والعكس صحيح.

٣ – إذا أضفنا وطرحنا مقداراً ثابتاً من كل القيم فإن قيمة الانحراف المعياري (أو التباين) لا تتغير (أي لا تتأثر قيمة الانحراف المعياري بالطرح أو الجمع). ولتوضيح هذه الخاصية نأخذ المثال التالي :

مثال (٥) :

حل المثال السابق رقم ٤ بعد طرح 70 (على سبيل المثال من كل القيم لاحظ أن هذا سوف يسهل الحسابات).

الحـل :

المربعات (بعد الطرح)	القيمة بعد طرح (70)
X2	X
0	0
25	5
1	1
25	5
16	4
36	6
9	3
64	8

ΣX = 176	ΣX 2 = 176

ويكون التباين

$$S^2 = \frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2$$

$$= \frac{176}{8} - \left(\frac{32}{8}\right)^2$$

= 22-16

= 6

والانحراف المعياري $S = \sqrt{6} = 2.449$ وهي النتائج نفسها مع ملاحظة أن العمليات الحسابية أسهل في هذه الحالة. وتجدر الإشارة إلى أنه لو طرحنا أي قيمة أخرى سنحصل على النتائج نفسها.

٤- إذا ضربنا كل قيمة في مقدار ثابت ثم حسبنا الانحراف المعياري للقيم الجديدة فإنه يجب القسمة على هذا المقدار الثابت. وإذا قسمنا كل قيمة على مقدار ثابت ثم حسبنا الانحراف المعياري للقيم الجديدة فإنه يجب الضرب في هذا المقدار الثابت.

٥,٥ معامل الاختلاف : The Coefficien of Variation

لمقارنة تشتت مجموعتين (أو أكثر) من البيانات وكانت البيانات تختلف في مستواها العام (أي في أوساطها

الحسابية) و / أو تختلف في وحدات القياس (مثلاً مقارنة بيانات الدخل حيث تقاس بالريال ببيانات العمر حيث تقاس بالسنوات) فإن المقارنة لا تتم مباشرة بمقارنة الانحراف المعياري لكل منهما بل تتم من خلال مقياس آخر هو "معامل الاختلاف" أو ما يسمى أحياناً مقياس التشتت النسبي حيث ينسب الانحراف المعياري لكل مجموعة إلى وسطها الحسابي والضرب في 100 فنحصل على مقياس نسبي أو مئوي (وبدون تمييز) أي تتم المقارنة بحساب معامل الاختلاف لكل منهما، والمجموعة التي لها معامل اختلاف أكبر تكون أكبر تشتتاً والعكس صحيح أي أن :

معامل الاختلاف = ______ × 100

وإذا رمزنا لمعامل الاختلاف بالرمز C. V فإن :

$$C.V = \frac{S}{\overline{X}} \times 100$$

مثال (٦) إذا كان الوسط الحسابي والانحراف المعياري لدخول عينة من الناخبين ¯ بالريالات هو :

$X_1 = 1500, S_1 = 152$

وكان الوسط الحسابي والانحراف المعياري لأعمارهم (بالسنوات) هو :

X2 = 42, S2 = 9̄.2

فأيهما أكثر تشتتاً الدخل أم العمر ؟

الحل :

لمقارنة التشتت نحسب معامل الاختلاف لكل من الدخل والعمر كما يلي :

١- معامل اختلاف الدخل :

$$c.v_1 = \frac{s_1}{\bar{x}_1} \times 100 = \frac{152}{1500} \times 100 = 10.13\%$$

٢- معامل اختلاف العمر :

$$c.v_2 = \frac{s_2}{\bar{x}_2} \times 100 = \frac{9.2}{42} \times 100 = 21.9\%$$

٣- بما أن معامل اختلاف العمر أكبر من معامل اختلاف الدخل فإن بيانات العمر تكون أكثر تشتتا من بيانات الدخل.

حساب التباين والانحراف المعياري للبيانات المبوبة :

إذا كانت البيانات مبوبة على شكل قيم وتكرارات، فإن التباين والانحراف المعياري بالطريقة المختصرة يمكن حسابهما كما يلي :

التباين في حالة البيانات المبوبة :

$$S^2 = \frac{\sum x^2 f}{\sum f} - \left(\frac{\sum xf}{\sum f}\right)^2$$

الانحراف المعياري للبيانات المبوبة

$$S = \sqrt{\frac{\sum x^2 f}{\sum f} - \left(\frac{\sum xf}{\sum f}\right)^2}$$

حيث X تمثل القيم، f تمثل التكرارات. وإذا كانت البيانات على شكل فئات تحسب مراكز هذه الفئات ويرمز لها بالرمز X، ويتم حساب $\sum Xf$ ، $\sum X^2 f$

مثال (٧) : الجدول التالي يمثل توزيع مجموعة من الناخبين حسب أعمارهم وفي وقت معين، والمطلوب حساب الانحراف المعياري لأعمار الناخبين ؟

عدد الناخبين F	الأعمار X
3	20
4	25
5	30
2	35
14	المجموع

الحل :

ينظم الحل كما في الجدول التالي حيث يتم حساب كل من Xf، X2f للحصول على $\Sum Xf$، $\Sum X^2 f$

حاصل ضرب X2f	حاصل ضرب Xf	أعداد الناخبين f	الأعمار X
1200	60	3	20

2500	100	4	25
4500	150	5	30
2450	70	2	35
Σx2f = 10650	Σxf = 380	Σf = 14	المجموع

ويحسب الانحراف المعياري باستخدام الصيغة التالية :

$$S = \sqrt{\frac{\sum x^2 f}{\sum f} - \left(\frac{\sum xf}{\sum f}\right)^2}$$

$$= \sqrt{\frac{10650}{14} - \left(\frac{380}{14}\right)^2}$$

$$= \sqrt{760.7 - 736.6}$$

$$= \sqrt{24.1} = 4.9$$

أي أن الانحراف المعياري لأعمار الناخبين يساوي ٤,٩ سنة.

مثال :

البيانات في الجدول الخاص بالمثال الشامل رقم (٦) على المتوسطات التي تمثل أهم الحروب التي شهدها العالم خلال الفترة من عام 1945، وحتى عام 1980م. احسب المدى، الانحراف المعياري، ومعامل الاختلاف لعدد الحروب خلال الفترة ؟

الحل :

يمكن تلخيص بيانات الجدول رقم (١) في الجدول التالي :

التكرارات F	الحروب X
4	0
8	1
7	2
9	3

6	4
2	6
36	المجموع

والجدول يوضح أن هناك أربع سنوات لم تحدث فيها حروب، وثمان سنوات حدثت في كل منها حرب واحدة، وسبع سنوات حدثت في كل منها حربان، وهكذا....

أولاً: المدى

المدى هو الفرق بين أكبر قيمة وأصغر قيمة. وأكبر قيمة هنا تساوي 6 حروب وأصغر قيمة تساوي صفر (لا توجد حروب)، وبالتالي فالمدى هو :

(6 = 0 - 6) أي أن المدى يساوي 6 حروب.

ثانياً: الانحراف المعياري :

يمكن تلخيص الحسابات كما في الجدول التالي :

الحروب X	التكرارات f	حاصل ضرب X. f	حاصل ضرب X2f
0	4	0	0
1	8	8	8
2	7	14	28
3	9	27	81
4	6	24	96

6	2	12	72
المجموع	Σf =36	Σxf = 85	Σx2f = 285

$$S = \sqrt{\frac{\sum x^2 f}{\sum f} - \left(\frac{\sum xf}{\sum f}\right)^2}$$

$$= \sqrt{\frac{285}{36} - \left(\frac{85}{36}\right)^2}$$

$$= \sqrt{7.917 - 5.569}$$

$$= \sqrt{2.348} = 1.53$$

أي أن الانحـراف المعيـاري لعـدد الحـروب خـلال تلـك الفترة يساوي 1.53 حرباً

(لاحـظ أن التبـاين يسـاوي 2.348 فهـو القيمـة قبـل استخراج الجذر)

ثالثاً: معامل الاختلاف :

معامل الاختلاف يساوي خارج قسمة الانحراف المعياري على الوسط الحسابي مضروباً في (100) أي أن :

$$CV = \frac{s}{\bar{X}} \times 100$$

١- الوسط الحسابي (كما حسبناه سابقا) :

$$\bar{X} = \frac{\sum xf}{\sum f} = \frac{85}{36} = 2.36$$

٢- الانحراف المعياري (كما حسبناه أيضا) S = 1.53

٣- معامل الاختلاف هو:

$$CV = \frac{1.53}{2.36} \times 100$$

= 64.8 %

معامل الارتباط

Auto Correlation

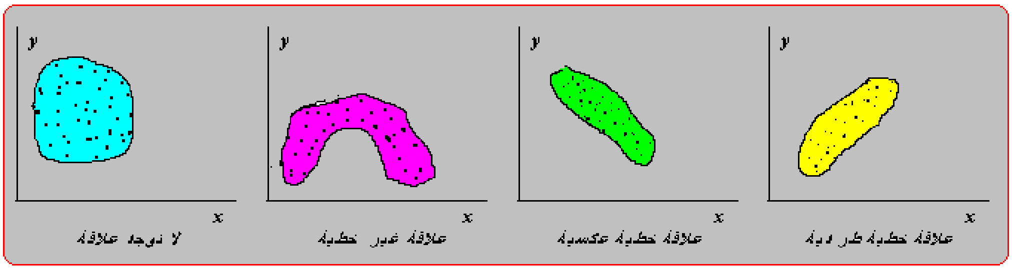
لاحظنا في الفصول السابقة الطرق و الأساليب المختلفة في جمع البيانات و تصنيف و تبويب البيانات ، كذلك عملية استخراج بعض المقاييس التي تعطي فكرة أكثر و وضوحا عن تلك البيانات كالمتوسطات و مقاييس التشتت ، إن هذه الطرق و الأساليب استندت على البيانات المجمعة عن متغير واحد فقط سواء كانت هذه البيانات مبوبة في توزيع تكراري أم غير ذلك . وفي أحوال كثير يواجه الباحث حالات تتطلب دراسة متغيرين أو أكثر في آن واحد لبيان طبيعة و نوع العلاقة التي تربط بها هذه المتغيرات ، عليه فان هذا الفصل سوف يخصص لدراسة مقاييس أخرى تحدد درجة و نوع وشكل العلاقة بين متغيرين أو أكثر.

الارتباط الخطي Linear Correlation

أن مفهوم الارتباط الخطي يقترن بحالة وجود متغيرين أو أكثر ترتبط مع البعض بعلاقات خطية معينة على سبيل المثال العلاقة بين طول الشخص و وزنه ، العلاقة بين تحصيل الطالب

المتخرج من الكلية و معدل درجاته في الثانوية و مستوى ألمعاشي لأسرته ، العلاقة بين نسبة الشفاء من مرض معين و كمية الجرعة من الدواء المخصص للمريض و عمر المريض. فإذا كان التغير في إحدى المتغيرات يؤثر في تغير متغير أخر أو مجموعة متغيرات أخرى عندئذ يقال أن هذه المتغيرات مرتبطة فيما بينها و إذا كان المتغيرين المرتبطين (أو مجموعة من المتغيرات المترابطة) يتغيران (تتغير) بنفس الاتجاه أي زيادة أو نقصان في إحداهما تؤدي إلى الزيادة أو النقصان في الأخر (الأخرى) عندئذ يقال أن الارتباط ما بينهما هو ارتباط موجب ، على سبيل المثال زيادة طول الشخص يتوقع أن يقابلها زيادة في وزنه. انخفاض في دخل الفرد يتوقع عنه انخفاض في إنفاقه على بعض السلع. أما إذا كان المتغيرين المرتبطين (أ، مجموعة من المتغيرات المرتبطة) يتغيران (تتغير) باتجاه معاكس أي زيادة أ، نقصان في إحداهما تؤدي إلى نقصان (زيادة) في الأخر (الأخرى)، عندئذ يقال أن الارتباط ما بينهما هو ارتباط سالب على سبيل المثال زيادة سعر الوحدة من سلعة معينة يتوقع إن يؤدي إلى انخفاض في الطلب على تلك السلعة ، انخفاض في درجات الحرارة يتوقع أن ينجم عنه زيادة الطلب على الوقود.

و يقال أن الارتباط بين المتغيرين أو أكثر هو ارتباط تام **perfect** إذا كان التغير في إحداهما متناسب مع التغير في الأخر على سبيل المثال إن الارتباط بين درجة الحرارة المئوية و درجة الحرارة الفهرنهايتية هو ارتباط تام باعتبار إن التغير في الأول متناسب مع التغير في الثاني. $F=(9/5)C+32$



الارتباط الخطى البسيط Simple Correlation

إذا كان الغرض من التحليل هو تحديد نوع وقوة العلاقة بين متغيرين ، يستخدم تحليل الارتباط ، وأما

إذا كان الغرض هو دراسة وتحليل أثر أحد المتغيرين على الآخر ، يستخدم تحليل الانحدار، وفي هذا الفصل يتم عرض أسلوب تحليل الارتباط الخطي البسيط، أي في حالة افتراض أن العلاقة بين المتغيرين تأخذ الشكل الخطي .

الغرض من تحليل الارتباط الخطى البسيط

الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين، ويرمز له في حالة المجتمع بالرمز ρ (رو)، وفي حالة العينة بالرمز r، وحيث أننا في كثير من النواحي التطبيقية نتعامل مع بيانات عينة مسحوبة من المجتمع، سوف نهتم بحساب معامل الارتباط في العينة r كتقدير لمعامل الارتباط في المجتمع، ومن التحديد السابق للغرض من معامل الارتباط، نجد أنه يركز على نقطتين هما:

نوع العلاقة:ـ وتأخذ ثلاث أنواع حسب إشارة معامل الارتباط كما يلي:

إذا كانت إشارة معامل الارتباط سالبة ($r < 0$) توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة أحد

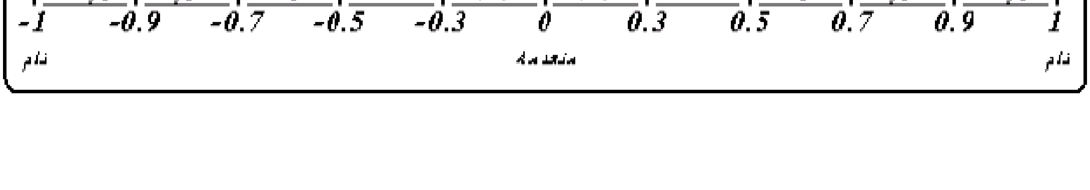
المتغيرين يصاحبه انخفاض في المتغير الثاني، والعكس.

إذا كانت إشارة معامل الارتباط موجبة ($r > 0$)توجد علاقة طردية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه زيادة في المتغير الثاني، والعكس .

إذا كان معامل الارتباط قيمته صفرا ($r = 0$) دل ذلك على انعدام العلاقة بين المتغيرين.

قوة العلاقة:- ويمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن (±1)، حيث أن قيمة معامل الارتباط تقع في المدى($-1 < r < 1$)، وقد صنف بعض الإحصائيين درجات لقوة العلاقة يمكن تمثيلها على الشكل التالي

درجات قوة معامل الارتباط



معامل الارتباط الخطي البسيط Simple Linear Correlation Coefficient

يعرف معامل الارتباط الخطي البسيط بأنه القيمة العددية للعلاقة الخطية بين متغيرين و تحسب من القانون الأتي:

$$r_{xy}=\frac{n\sum_{i=1}^{n} X_i Y_i-(\sum X_i \sum Y_i)}{\sqrt{\left\{(n\sum X_i^2-(\sum X_i)^2)(n\sum Y_i^2-(\sum Y_i)^2)\right\}}}$$

معامل الارتباط الجزئي Partial Correlation Coefficient

يلاحظ في بعض الأحيان و جود ارتباط بين متغيرين يعزي جزئيا إلى ارتباط متغير ثالث مرتبط مع كلا المتغيرين . لنفرض إن المتغيرين X1,X2 مرتبطين و هناك متغير أخر X3 مرتبط مع كلا المتغيرين و إننا نرغب في قياس درجة الارتباط بين X1,X2 باستبعاد اثر المتغير الثالث X3 على كلا المتغيرين.إن المعامل الذي يقيس درجة ارتباط متغيرين باستبعاد اثر الثالث

يدعى بمعامل الارتباط الجزئي . على سبيل المثال الارتباط بين طول الفرد X1 و وزنه X2 يتأثر بارتباط عمر الفرد X3 مع كل من X1,X2 ، كذلك الارتباط بين دخل الأسرة الشهري X1 و إنفاقها الشهري X2 يتأثر بارتباط عدد أفراد الأسرة X3 مع كل من X1,X2 ، على سبيل المثال عدد السكائر المدخنة X1 يوميا و الإصابة بنوع معين من أمراض الرئة X2 يتأثر بارتباط عمر المدخن X3 و عدد سنوات التدخين X4 مع كل من X1,X2. و يمكن إيجاد قيمة معامل الارتباط الجزئي وفق ما يلي: افرض إن X1,X2,X3 تمثل ثلاثة متغيرات عشوائية بحيث أن المتغيرين X1,X2 مرتبطين و أن X3 مرتبط مع كل من X1,X2 و افرض أن على أساس عينة من المفردات قوامها n تم الحصول على القياسات المتناظرة لهذه المتغيرات

((X1i , X2i , X3i), i=1,2,3………n) و أن معامل الارتباط البسيط بين متغيرين منهما هو r12 , r13 , r23 عندئذ فان معامل الارتباط الجزئي بين X1,X2 باستبعاد اثر الثالث X3 هو

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} \quad \ldots\ldots\ldots\ldots 1$$

و في حالة وجود أربعة متغيرات و نرغب في حساب الارتباط الجزئي بين الأول و الثاني باستبعاد اثر الثالث و الرابع فان هذا المعامل هو

$$r_{12.34} = \frac{r_{12.3} - r_{14.3}\, r_{24.3}}{\sqrt{(1-r_{14.3}^2)(1-r_{24.3}^2)}} \ldots\ldots\ldots\ldots 2$$

معامل الارتباط الرتب لسبيرمان Spear mans rank correlation Coefficient

افرض إن X,Y متغيرين من النوع الوصفي و افرض إن البيانات المستحصل عليها من X,y على أساس عينة عشوائية من المفردات قوامها n هي بهيئة صفات غير قابلة للقياس الكمي .

افرض إن (Xi , Yi , i=1,2,3,.....n) ممكنة الترتيب تصاعديا أو تنازليا وفق معيار معين يمتاز به كل متغير (مثلا تقديرات درجات مجموعة من الطلبة يمكن ترتيبها تصاعديا على أساس معيار الأقل إلى اعلي درجة أو العكس) عندئذ استنادا لهذا الترتيب و لكل متغير يمكن تخصيص قيم سلسلة الإعداد الطبيعية (1,2,3,......,n) لصفات الترتيب بحيث إن كل صفة يخصص لها احد إعداد هذه السلسة ، في حالة عدم تكرار أيه صفة منها (وقد يكون

التخصيص تنازلي) و سوف نوضح حالة التكرار بعض الصفات و أسلوب علاج ذلك في أمثلة. و القانون المستخدم لإيجاد الارتباط لمثل هذه الحالات:

$$r_{xy} = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \qquad , \quad d_i = X_i - Y_i$$

الانحدار الخطي البسيط Simple Linear regression

يعرف الانحدار الخطي البسيط بأنه عملية تقدير العلاقة الخطية بين متغيرين أحداهما مستقل و الأخر تابع . إن مفهوم الانحدار الخطي البسيط يقترن بمفهوم الارتباط الخطي البسيط . و يهدف الانحدار الخطي البسيط إلى تقدير قيم عددية لمعالم النموذج ، أي تقدير قيم عددية لكل من (a,b) و معادلة خط الانحدار البسيط هي كالأتي:

$$\hat{y} = \hat{a} - \hat{b}X$$

$$\hat{b} = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2}$$

$$\hat{a} = \bar{Y} - b\bar{X}$$

ومن أعلاه نلاحظ بان $\hat{b}$،$\hat{a}$ هي دالة بدلالة قياسات مفردات العينة (Xi,Yi) إن $\hat{a}$ تمثل في الحقيقة مقطع خط الانحدار مع المحور الصادي Y إي ما نعنيه قيمة Y التقديرية عندما X=0 و إذا كانت $\hat{a}$=٠ عندئذ يقال إن خط الانحدار يمر من نقطة الأصل (٠,٠) وان $\hat{b}$ تسمى معامل الانحدار أي مقياس يوضح مقدار تغير Y إذا ما تغيرت X بوحدة واحدة و يلاحظ إن $\hat{b}$ قد تكون موجبة أو سالبة تبعا لقيمة الانحراف المعياري لــX,Y (Sxy) إذا كانت موجبة أم سالبة فقيمة $\hat{b}$ الموجبة تعني إن العلاقة بين X,Y موجبة و القيمة السالبة لها تعني إن العلاقة سالبة (عكسية) و غالبا ما يقال لما سبق أنة تم تقدير معادلة الانحدار (Y / X) أي إيجاد معادلة الانحدار فيها المتغير التابع هو Y و المستقل X و العكس ممكن.

*تحليل الإرتباط

**دراسةالإرتباط بين متغيرين كلاهما كمي

@الهدف الأساسي من التحليل:

+ هل هناك علاقة (إرتباط) بين المتغيرين؟

+ هل العلاقة خطية أم غير خطية؟ نوع

+ هل العلاقة طردية (موجبة) أم عكسية (سالبة)؟ إتجاه

+ هل العلاقة قوية أم ضعيفة؟ درجة

تنبيه هام جدا -----(((((الإرتباط لا يعني السببية)))))------

+ أدوات إحصائية مساعدة:

(١) شكل الإنتشار (الشكل الإنتشاري ، لوحة الإنتشار، شكل التبعثر)

+ رسم بياني

+ نوع(خطية، غير خطية) ، إتجاه ، درجه تقديريه

+ لا يعطي قياس كمي لقوة (درجة العلاقة)

(٢) معامل التغاير

+ نوع(خطية) ، إتجاه ، درجة

+ يصعب الحكم على درجة (قوة) العلاقة

(٣) معامل الإرتباط الخطي البسيط (معامل إرتباط بيرسون)

+ نوع(خطية) ، إتجاه ، درجة

+ حد أدنى (-١) وحد أعلى (+١).

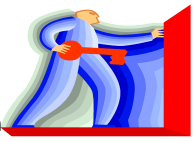
+ صفر يعني عدم وجود علاقة(خطية)

+ فكر فيه كنسخة مطورةقيمة معيرة أو معيارية من معامل التغاير.

تحليل الإرتباط

الإرتباط بين متغيرين كميين

الإرتباط الخطي



معامل إرتباط بيرسون - --(معامل الإرتباط الخطي البسيط)---(معامل إرتباط العزوم)

الصيغ :

معامل الإرتباط للمجتمع (معلمة) :

$$\rho = \frac{Cov(x,y)}{\sigma_X \sigma_Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

معامل الإرتباط للعينة (إحصائية عينة) :

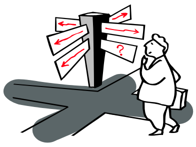
التغاير

معامل الإرتباط

$$\tilde{\rho} = \frac{\operatorname{cov}(x, y)}{s_x s_y} = \frac{s_{xy}}{s_x s_y}$$

الإنحراف المعياري

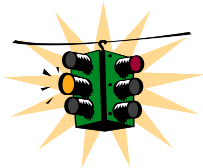
الإنحراف المعياري للمتغير X

بالتعويض  **نحصل على الصيغة :**

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{[\sum_{i=1}^{n} (x_i - \bar{x})^2][\sum_{i=1}^{n} (y_i - \bar{y})^2]}}$$

 **بعد إجراء بعض العمليات الجبرية يمكن الحصول على الصيغة الحسابية المبسطة:**

$$r = \frac{[(n\sum_{i=1}^{n} x_i y_i) - [(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)]]}{\sqrt{[n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2][n(\sum_{i=1}^{n} y_i^2) - (\sum_{i=1}^{n} y_i)^2]}}$$



معامل إرتباط بيرسون ملائم لقياس الإرتباط الخطي بين متغيرين كميين.

تتراوح قيمة معامل إرتباط بيرسون بين –١ و +١ أي أن : $+1 \leq r \leq -1$.

يتم الحكم على قوة العلاقة الخطية بالنظر للقيمة المطلقة لمعامل إرتباط بيرسون $|r|$ فكلما إقتربت القيمة المطلقة لمعامل إرتباط بيرسون من الواحد الصحيح دل ذلك على وجود علاقة خطية قوية.

يتم الحكم على إتجاه العلاقة من خلال إشارة معامل إرتباط بيرسون. فالإشارة الموجبة لمعامل إرتباط بيرسون تعني علاقة طردية (موجبة) والإشارة السالبة تعني علاقة عكسية (سالبة).

عندما يكون معامل إرتباط بيرسون مساويا للصفر (أو قريبا منه) فإن ذلك يعني عدم وجود علاقة خطية (أو أن العلاقة الخطية ضعيفة جدا) ولكن لا يمكننا القول بعدم وجود علاقة بشكل مطلق فقد يكون هناك علاقة بين المتغيرين ولكنها علاقة غير خطية وفي تلك الحالة لن يكون معامل إرتباط بيرسون الأداة المناسبة للكشف عن مثل هذا النوع من العلاقة.