محاضرات في مقياس الإحصاء التربوي1

السنة الأولى ماستر علم اجتماع التربية (السداسي الأول)

السنة الجامعية: 2023/2022

المحاضرة الأولى

ماهية علم الإحصاء

تمهيد:

ظهر علم الإحصاء منذ فجر التاريخ، غير أن الاستخدام كان يتمُّ في صُورَةٍ مُبسَّطة، ويتمثَّل ذلك في العَدِّ والحصر، فعلى سبيل المثال كانت هناك حاجة لحصر الجنود من أجل خوض المعارك الحربية، وكذا المحاصيل الزراعية التي يتم جمعها وتشوينها، وكذا حصر الثروة الحيوانية التي يمتلكها الأفراد، وغير ذلك من الموارد الاقتصادية، أو تعداد السكان، وفي القرن السادس عشر الميلادي ظهرت الحاجة للتوصيف، وتفسير العلاقة بين المتغيرات الاجتماعية، وفي القرن الثامن عشر ظهرت آليات إحصائية متطورة نتيجة اكتشاف نظرية الصدفة أو الاحتمال"LAW OF CHANCE" ، وتلك النظرية تقضي بحدوث أمر ما تحت ظرف مُعيَّن، ومن أوائل من قاموا بوضع أُسُس لهذه النظرية العالم "بليس باسكال"BLAIS PASCAL ، مع وجود دراسات بسيطة في الحقب الزمنية الماضية على يد العلم "جاليليو"، و"أرسطو"، وسنستعرض أنواع المتغيرات الإحصائية، وغيرها من الأطروحات المرتبطة بعلم الإحصاء.

1- مفهوم وطبيعة علم الإحصاء:

يُعرَّف مصطلح الإحصاء من الناحية اللغوية بأنه عدٌّ بشكل شامل، أما من الناحية الإجرائية فإن علم الإحصاء يُعرَّف بأنه أحد الفروع الأساسية لعلم الرياضيات، ويهدف إلى تجميع وتنظيم وتحليل البيانات، وبما يساعد على الوصول لاستنتاجات، ثم توصيات، ويتبع ذلك اتخاذ القرارات)، كما أن علماء الاجتماع قاموا بتعريف على الإحصاء بأنه: (مجموعة من الطرق التي يستخدمها الباحثون في تجميع البيانات والقيام بتنظيمها وإيجازها وتحليلها وعرضها في هيئة رسوم بيانية، بهدف تحديد العلاقات واتخاذ القرارات في ضوء ذلك).

2- أهداف علم الإحصاء بالنسبة للبحوث والرسائل العلمية:

– يُساعد علم الإحصاء في توصيف الظواهر بأسلوب منهجي دقيق، وهو إحدى الدعائم الرئيسية في استخدام المنهج الوصفي.

– التعرف على طبيعة العلاقات المُعقَّدة بين المتغيرات.

– التنبؤ بطبيعة الظواهر في فترات زمنية متتالية وبناءً على المعلومات الراهنة التي يتم جمعها من المفحوصين.

– التوصل إلى نتائج بهيئة مناسبة.

3– أهم الأدوات المستخدمة في جمع البيانات:

يحتاج الباحثون أدوات منهجية تساعد في تجميع البيانات، ومن أبرز هذه الأدوات ما يلي:

1) الاختبارات:

والاختبارات خيار مهم بالنسبة للباحثين، وتُستخدم في معرفة خصائص وصفات تتعلق بأفراد عيِّنة البحث، ويُصمِّم الباحث الأسئلة المتعلقة بالاختبارات حسب طبيعة موضوع البحث، كما تُستخدم تلك الأداة في كثير من مراكز الأبحاث العلمية.

2) الاستبيان:

يتألَّف الاستبيان من أسئلة تتعلَّق بموضوع الدراسة البحثية، وتُوضَّح في نموذج، ويتم إرسالها إلى المبحوثين من خلال البريد، وفي تلك الفترة شاع استخدام الاستبيان الإلكتروني، والذي يتم إرساله من خلال شبكة الإنترنت إلى المبحوثين، وبما يُوفِّر الوقت عليهم، وكذا ضمان جمع الاستمارات دون وجود بيانات مفقودة، ومن أشهر أنواع الاستبيانات:

الاستبيان المُغلق: وفيه يقوم الباحث بوضع أسئلة و خيارات يختار منها المفحوصين ما يرونه مُناسبًا من الإجابات.

الاستبيان المفتوح: ويتضمَّن الاستبيان المفتوح أسئلة دون تحديد للإجابات، وبما يساعد المفحوصين بالإدلاء بآرائهم بحرية كاملة.

الاستبيان المفتوح المُغلق: ويحتوي ذلك النوع من الاستبيانات على أسئلة مفتوحة ومُغلقة في الوقت نفسه.

3) الملاحظة:

تعتمد أداة الملاحظة على استخدام الحواس البشرية، مثل: السمع، والنظر، والشم، واللمس، وتتَّسم النتائج التي يتحصَّل عليها الباحث من استخدام تلك الأداة بالدقَّة، ومن خلال الملاحظة يُعبِّر الباحث عن الظاهرة بطريقة كيفية أو كمِّية، ومن أهم تصنيفات الملاحظة.

الملاحظة البسيطة: وهي تتم بشكل عفوي دون سابق تجهيز، ويحصل عن طريقها الباحث على معلومات أولية، يمكن من خلالها أن ينطلق في فهم وتفصيل الظاهرة محل البحث.

الملاحظة المُنتظمة: ويستخدم الباحث في الملاحظة المُنتظمة مجموعة من الأجهزة العلمية حسب الحاجة، وتتَّسم تلك الطريق بالمنهجية، والإعداد المُسبق، وتسجيل جميع السمات والخصائص المتعلقة بموضوع البحث.

الملاحظة بالمشاركة: وفي ذلك النوع يقوم الباحث بمشاركة المفحوصين فيما يقومون به من أعمال روتينية من غير أن يُشعرهم بهويَّته، كأن يدخل إلى السجن أو دور رعاية مُسنِّين، ويُسجِّل ما يراه من سمات وخصائص.

الملاحظة دون التداخل أو المشاركة: ويُشاهد الباحث في تلك الحالة المفحوصين من بعيد، بحيث لا يراه أحد منهم، ويسجل الملاحظات الخاصة به.

4) المقابلة:

هي وسيلة شائعة في جمع البيانات من عيِّنات البحث، وتتطلَّب خبرة كبيرة من الباحث، وقُدرة على محاورة المفحوصين، من خلال ما يُلقيه من أسئلة، ومن أهم أنواع المقابلة ما يلي:

المقابلة المفتوحة: ويستخدم الباحث في ذلك أسئلة لا تتقيَّد بإجابات، ويترك فيها العنان للمفحوصين للتعبير عن آرائهم.

المقابلة المغلقة: وتتضمن أسئلة قصيرة نوعًا ما، ويمنح فيها الباحث خيارات للمستجيبين، مثل: نعم، أو لا.

المقابلة المغلقة المفتوحة: ويشمل ذلك النوع من المقالات أسئلة مغلقة ومفتوحة.

4- تصنيفات علم الإحصاء:

الإحصاء الوصفي (الكيفي): ويساعد ذلك النوع في تجميع البيانات المتعلقة بالمشكلة أو الظاهرة محل الدراسة، ثم القيام بإجراءات منهجية تستهدف التوصل للنتائج التي توضح السمات المحورية، وينصبُّ ذلك النوع على دراسة مجموعة من المفردات، مثل تحديد الأجور، والعمر، ويتم العرض في جداول.

الإحصاء التحليلي (الإستنتاجي): ويستخدم الإحصاء التحليلي في دراسة الظواهر التربوية والاجتماعية، ولا يكتفي بالعرض فقط، بل يتعمق لتحليل الخصائص؛ من خلال مجموعة من الأدوات الإحصائية، وتُستخدم النتائج في إجراء التعميمات. ومثال على ذلك: (دراسة العلاقة بين التحصيل الدراسي وأسلوب المعاملة الوالدية).

5- أنواع البيانات في الإحصاء:

يمكن أن نُقسِّم أنواع البيانات الإحصائية إلى نوعين كبيرين، وهما البيانات الوصفية، والبيانات الكمية، وسنوضح نبذة عن كل منهما فيما يلي:

البيانات الوصفية (الكيفية):

وهي عبارة عن بيانات غير رقمية، ويمكن أن نُقسِّمها إلى شُعَب أخرى، وهي اسمية، وترتيبية:

البيانات الترتيبية: ويكون فيها نوع من الأفضلية بين قيمة وأخرى، وترتب بشكل تصاعدي أو تنازلي: ومثال على ذلك: تقديرات الجامعة (مقبول، جيد، جيد جدًا، ممتاز)، وكذلك (موافق، غير موافق، غير موافق مطلقًا).

البيانات غير الترتيبية (الاسمية): وهي بيانات لا يمكن أن يتم ترتيبها بشكل تنازلي أو تصاعدي، مثل: الحالة الاجتماعية للفرد: (أرمل، متزوج، أعزب)، والنوع (ذكر، أنثى)، ونرى أن تلك النوعية لا تُوجد أفضلية في قيم على أخرى، وليس هناك ضرورة لترتيبها بصورة تصاعدية أو تنازلية.

6- أنواع المتغيرات الإحصائية:

أنواع المتغيرات الإحصائية، وفقًا لطبيعة البيانات:

المتغيرات الوصفية:

وهي المتغيرات التي يمكن وصفها بطريقة لفظية، فعلى سبيل المثال:

. هل هناك فرق بين الذكور والإناث في استيعاب مادة الفيزياء؟

· هل يوجد تأثير للتدخين على التحصيل الدراسي؟

المتغيرات الكمية:

وهي متغيرات يمكن قياسها وتسجيلها بطريقة رقمية، مثل الوزن، والطول، والسن، ويمكن تقسيمها إلى:

متغيرات متصلة: وهي تقبل الكسور العشرية، مثل المسافة، والوزن.

متغيرات منفصلة : وهي لا تقبل الكسور العشرية، مثل عدد طلاب الجامعة.

أنواع المتغيرات الإحصائية حسب تأثير المتغير:

المتغير المستقل: وهو المتغير الذي يؤثر في غيره من المتغيرات بالزيادة أو النقصان.

المتغيرات الدخيلة: وهي متغيرات يُحتمل أن تُؤثِّر في الدراسة، ويقوم الباحث بتهيئتها والتعرف على مدى تأثيرها.

المتغير التابع: وهي تتأثر بالمتغير المستقل

المحاضرة الثانية

مصادر جمع البيانات الإحصائية

تهتم نظرية العينات بدراسة العلاقة بين المجتمع والعينات المسحوبة منه، بحيث هناك عدة طرق لأخذ العينات من المجتمع لاستخدامها في الاستدلال الإحصائي ومن أشهر هذه الطرق هي العينة العشوائية وهي العينة التي تكون لكل مفرده من مفردات المجتمع نفس فرصة الاختيار في العيينة. فمثلاً نستعين بعينه مسحوبة من المجتمع لتقدير معالم هذا المجتمع مثل متوسطه أو تباينة أو غير ذلك. أو أعطاء عينه من المرضى بارتفاع الضغط، مثلاً دواء معين ثم قياس ضغطهم قبل وبعد تناولهم لهذا الدواء لمعرفة ما إذا كان هذا الدواء مفيد في خفض الضغط أم لا.

أي مجموعات من المفردات تشترك في صفه أصفات وتكون موضوع دراسة أو بحث فإن هذه المجموعة يطلق عليها إحصائيا مجتمع الدراسة (أو اختصاراً المجتمع **Population**). والمجتمع قد يكون مجموعة ما من البشر أو أشجار أنواع معينه من الفاكهة أو الحيوانات الزراعية أو إنتاج دولة ما لسلع معينه خلال فترة زمنية محدده...الخ.

والمجتمع قد يكون محدوداً إذا كان يمكن حصر عدد أفراده مثل سكان مدينة ما أو طلاب مرحلة تعليمية معينة ، وقد يكون المجتمع غير محدود (لانهائي) إذا كان لا يمكن حصر عدد أفراده مثل النجوم والكواكب أو الكائنات الحية بمياه المحيطات والأنهار وعند دراسة صفة ما أو صفات معينه لمجتمع ما ، فإن البيانات الإحصائية عن تلك الصفة أو الصفات تجمع بأحد الأسلوبين هما:

أولاً: أسلوب الحصر الشامل (census): وفيه تجمع البيانات عن كل مفرده من مفردات المجتمع، وهذا الأسلوب يتطلب وفرة في الوقت والمال والمجهود الفني وتزداد هذه المتطلبات وتتضاعف كلما ازداد حجم المجتمع (عدد أفراد المجتمع). وهذا الأسلوب لا يتبع عادة إلا في حالة التعدادات التي تجريها الدول وتدعمها بإمكانيات ضخمه مثل تعدادات السكان والتعدادات الصناعية والتعدادات الزراعية.

الثاني: أسلوب المعاينة (Sampling method): وفيه يتم جمع البيانات عن جزء من مفردات المجتمع يختار بطريقة أو بأخرى ويطلق عليه عينه **(Sample)** ثم بعد ذلك يتم تعميم نتائج الدراسة على المجتمع بأكمله. أي أن إسلوب المعينة يقصد به دراسة خصائص المجتمع من خلال دراسة عينه مسحوبة منه ، ونجاح هذا الإسلوب يعتمد على أن تحمل العينة أقصى درجة من دقة التمثيل للمجتمع المسحوبة منه.

مزايا أسلوب المعاينة:

يتميز أسلوب المعاينة عن أسلوب الحصر الشامل بمزايا عديده منها:

1- يؤدي استخدام العينات العشوائية إلى خفض تكاليف الدراسات الميدانية بسبب صغر حجم العينة بالنسبة إلى حجم المجتمع وهو ما يؤدي إلى تخفيض الأعباء الإدارية والفنية التي تتطلبها أي دراسة ميدانية.

2- يتحقق وفر واضح في الوقت الذي ينفق في دراسة ميدانية على أساس عينة بدلاً من الحصر الشامل وتتضح أهمية الوقت عندما نقوم بدراسة ظاهرة تتغير بمرور الوقت.

3- في المجتمعات غير المحدودة (اللانهائية) مثل مجتمع الكائنات الحية في البحار والمحيطات لا يمكن أن تتم الدراسة على أساس الحصر الشامل ولكن لابد وأن تتم الدراسة بأسلوب المعاينة.

4- أيضاً هناك بعض الاختبارات لابد وأن تتم بأسلوب المعاينة لأن إجراء مثل هذه الاختبارات على أساس الحصر الشامل يؤدي إلى تلف المادة المختبرة أو هلاكها.. فاختبار صلاحية شحنه من المفرقعات مثلاً لابد وأن يتم على أساس العينة وبالمثل تحليل دم المرضى يتم على أساس عينه.

أنواع العينات:

تنقسم العينات عادة إلى قسمين رئيسين وهما عينات عشوائية وعينات غير عشوائية، وفيما يلي تفصيل لكل قسم منها.

1- العينات العشوائية

هي تلك العينات التي يتم اختيار مفرداتها حسب خطه إحصائية لا يكون فيها للباحث أو لمفردات العينة دخل في اختيار أي مفرده فيها ، حيث يتم الاختيار باستخدام أساليب معينة تلعب الصدفة خلالها الدور الأول في اختيار المفردة ولكن بشرط أن يتحقق لجميع المفردات احتمال ثابت ومحدد للاختيار. والعينات العشوائية إذا ما تم اختيارها بالطريقة العلمية السليمة والمناسبة يمكن أن تكفل درجه عالية من دقة التمثيل للمجتمعات المسحوبة منها لذلك فهي الوسيلة الأساسية في حالة البحوث العلمية الدقيقة .. من أهم أنواع العينات العشوائية مايلي.

(أ) العينة العشوائية البسيطة: **Simple random sample**

ويلجأ إليها الباحث في حالة ما إذا كان مجتمع الدراسة ليس كبيراً ويحمل قدراً من التجانس بين المفردات للصفة أو الصفات موضع الدراسة. والعينة العشوائية البسيطة تستغل فرص متكافئة لمفردات المجتمع للدخول في العينة ولكن المفردات التي تدخل في العينة تكون عن طريق الصدفة البحتة. والاختيار العشوائي يتم يدوياً عن طريق بطاقات متماثلة في الحجم واللون أو عن طريق جداول الأعداد العشوائية أو عن طريق الحاسب الآلي. ولكي يتحقق ذلك فإن الأمر يتطلب تحديد مفردات المجتمع تحديداً كاملاً ويكون هذا التحديد على شكل قائمة (أو خريطة) تضم كل مفردات المجتمع وهذه القائمة تسمى الإطار((Frame ولا يجوز الاختيار العشوائي إلا من المفردات التي يضمها الإطار.

(ب) العينة المنتظمة: **Systematic sample**

اختيار هذه العينة يتطلب وجود إطار للمجتمع كما في حالة العينة العشوائية البسيطة بحيث يعطى لكل مفرده من مفردات المجتمع رقماً متسلسلاً داخل الإطار ، ثم نختار مفردات العينة من الإطار بحيث يكون الرقم المتسلسل لكل مفردة يبعد بعداً ثابتاً منتظماً عن رقم المفردة السابقة لها وكذلك رقم المفردة اللاحقة لها. فمثلاً إذا كان لدينا مجتمعاً حجمه **2000** مفرده ونريد اختيار عينه منتظمة حجمها **100** مفرده فإننا نقسم الإطار إلى فترات منتظمة طول كل فترة = $\frac{2000}{100}=20$ مفردة ومن داخل مفردات الفترة الأولى **(1 – 20)** يختار مفرده واحدة عشوائياً ولتكن رقم **14** مثلاً وبناء على رقم تلك المفردة يتحدد باقي مفردات العينة المنتظمة فتكون هي المفردات ذات الأرقام **34، 54، ....، 1974، 1994.**

والعينة المنتظمة كثيرة الاستعمال في التطبيقات العملية لقلة تكاليفها وقلة الأخطاء التي ترتكب في اختيار مفردات العينة فضلاً عن سهولة إجرائها. ولكن أهم عيوب المعاينة المنتظمة هو عدم صلاحيتها إذا ما وجدت علاقة دورية مع ترتيب العناصر في القائمة وكان طول الفترة بين عناصر العينة مساوياً لطول الدورة أو إحدى مضاعفاتها.

(جـ) العينة العشوائية الطبقية: **Stratified random sample**

ويلجأ إليها الباحث في حالة ما إذا كان مجتمع الدراسة واضحاً به فئات (طبقات) بحيث أن التجانس أو التقارب داخل كل طبقة من طبقات مجتمع الدراسة أكبر من التجانس داخل المجتمع ككل (أي أن التشتت داخل المجتمع ككل أكبر من التشتت داخل كل فئة من فئاته على حده). في هذه الحالة يجب على الباحث مراعاة أن الطبقة داخل العينة بنفس نسبة وجودها داخل المجتمع (وأحياناً يوضع في الاعتبار عناصر أخرى مثل التشتت داخل الطبقة أو عنصر التكلفة لجمع البيانات عن الطبقة). بعد أن يتم تحديد عدد المفردات التي يجب سحبها من كل طبقة للدخول في العينة فإن هذه المفردات يتم سحبها عشوائياً من داخل الطبقة ومجموع هذه المفردات تكون العينة الطبقية العشوائية.

(د) العينة متعددة المراحل أو العنقودية: **clustered sample**

يلجأ إليها الباحث عندما يكون مجتمع الدراسة كبير جداً ومتناثراً على مساحات شاسعة تكلف الكثير من الوقت والجهد في التنقل بينها عند جمع البيانات، أيضاً في حالة عدم وجود إطار يضم جميع مفردات المجتمع فيستحيل الاختيار العشوائي مباشر من المجتمع. لهذا يلجأ الباحث إلى أخذ العينة على مراحل متعددة متتالية. في المرحلة الأولى يتم تقسيم المجتمع إلى عدد محدد من وحدات المعاينة الكبيرة الحجم ومنها يختار بعضها عشوائياً ثم يتلو ذلك كمرحلة ثانية تقسيم الوحدات المختارة عشوائياً من المرحلة الأولى إلى وحدات أقل منها في الحجم ثم يختار بعضها عشوائياً.. وهكذا تتابع مراحل التقسيم والاختيار العشوائي ، وعدد هذه المراحل ليس ثابت

بل يتوقف على طبيعة مجتمع الدراسة وإمكانيات الباحث .. في المرحلة الأخيرة يصل الباحث إلى وحدات المعاينة التي سيجمع عنها بيانات البحث ويطلق عليها وحدات المعاينة الأولية.

2- العينات غير العشوائية:

هي تلك العينات التي لا تكفل لجميع مفردات المجتمع احتمال ثابت ومحدد للاختيار، وغالباً يتدخل الباحث في عملية الاختيار بصورة أو بأخرى ... ومن أهم أنواع العينات غير العشوائية:

(أ) العينة العمدية أو المقصودة: **Purposive sample**

يلجأ الباحث إلى هذه الطريقة فيما إذا كان مجتمع الدراسة كبير جداً وكانت إمكانياته لا تسمح له إلا بدراسة عينة حجمها صغير جداً بالنسبة لمجتمع الدراسة، في هذه الحالة يتعمد الباحث اختيار مفردات معينة كعينة لمجتمع الدراسة يرى بخبرته السابقة أن هذه العينة يمكن أن تعطي تمثيلاً مقبولاً لمجتمع الدراسة.

مثلاً إذا أراد باحث دراسة خصائص اقتصادية أو اجتماعية معينه عن ريف دولة ما ، وكانت إمكانياته المالية والإدارية لا تسمح له بعينة سوى سكان قرية واحدة ، فإنه في هذه الحالة إذا ما تم اختيار القرية عشوائياً من بين آلاف القرى بتلك الدولة فإن الصدفة قد تأتي بقرية بعيدة في خصائصها (من حيث الظاهرة موضوع الدراسة) عن خصائص معظم قرى تلك الدولة ... كأن تأتي بالصدفة قرية ساحلية معظم سكانها من الصيادين أو قرية قريبة من مشروع صناعي ضخم يستوعب في قواه العاملة معظم سكانها.. هذه القرية أو تلك قد يأخذ النمط المعيشي لسكناها طابعاً خاصاً – نابعاً عن ظروفها الخاصة – بعيداً عن النمط المعيشي المعتاد لبقية القرى، لذلك فأي منها لا يمكن أن يعطي تمثيلاً مقبولا لريف تلك الدولة. لهذا فإن الباحث وعلى ضوء خبراته السابقة يتعمد اختيار قرية معينة يرى أنها – من وجهة نظره الشخصية- يمكن أن تمثل الريف. وهذه الطريقة غير علميه وغالباً يتم اللجوء إليها في حالة البحوث التمهيدية.

(ب) العينة الحصصيه: **Quota sample**

وهي نوع خاص من العينات غير العشوائية وتستخدم كثيراً في معاينة الرأي العام (على سبيل المثال عمليات استطلاعية الرأي العام التي يقوم بها معهد جالوب قبل إجراء انتخابات الرئاسة في الولايات المتحدة الأمريكية)... في هذه الطريقة يقسم المجتمع موضوع الدراسة إلى طبقات بالنسبة إلى صفات أو خصائص معينة ويتم العمل على تمثيل كل طبقة منها في العينة بنسبة وجودها في المجتمع الأصلي (وعلى سبيل المثال في حالة دراسة الدخل لمنطقة ما ورؤى أن يكون حجم العينة المطلوبة 100 فرد مثلاً عندما يريد الباحث أن يقوم جامعوا البيانات بالحصول على البيانات من 20 موظفاً، 45 من العمال الحرفيين ، 35 من ذوي الأعمال الحرة .. وتترك الحرية لجامعي البيانات في اختيار الأفراد المطلوبة فيها حدود المواصفات الموضوعة لكل طبقة من الطبقات المذكورة.

واضح أنه رغماً من أن هذه الطريقة في ظاهرها مماثلة للعينة الطبقية العشوائية.. إلا أنه في الحالة الأخيرة (العينة الطبقية العشوائية) يكون اختيار المفردات عشوائياً من داخل كل طبقة ولا يترك لجامع البيانات حرية اختيار المفردات من كل طبقة والذي قد يترتب عليه تميزاً كبيراً.

عموماً.. يلجأ الباحث إلى العينة الحصصية إذا كان من المرغوب فيها اظهار النتائج في وقت قصير مع التغاضي عن توافر درجة دقة عالية بتلك النتائج.

المحاضرة الثالثة

العرض الجدولي والبياني

بعد جمع البيانات من مصادرها المباشرة وغير المباشرة يحتاج الباحث إلى تنظيم وعرض تلك البيانات الإحصائية كأول مرحلة للتحليل الإحصائي، وبتفريغها بشكل يمكن للباحث التعامل معها بكفاءة وفهمها وإجراء الحسابات عليها، ويعتمد نوع العرض على نوع البيانات إما الكمية أو الكيفية، ويكون العرض إما عن طريق **جداول تكرارية** أو **رسومات بيانية**.

1) الجداول التكرارية: بعدما يجمع الباحث البيانات المتوفرة حول مفردات العينة فإنه يحصل على سلسلة عشوائية من البيانات تسمة بيانات غير مبوبة حيث يتم اختيار طريقة تمثيلها حسب نوع المتغيرات وعددها منها جداول بسيطة ذات متغير واحد وجداول مركبة أو ذات متغيرين أو أكثر.

جدول تكراري في حالة متغير كيفي (نوعي): يمكن عرض البيانات في شكل جدول تكراري بسيط يتكون من عمودين الأول يحتوي على مستويات المتغير والثاني يمثل التكرارات الموافقة للفئة المناسبة لها يسمى تكرار مطلق.

مثال: فيما يلي بيانات المستوى التعليمي لعينة تتكون من عشرين عاملا بإحدى المؤسسات:

جامعي، متوسط، ثانوي ، متوسط، جامعي، ثانوي، متوسط، ، جامعي، ابتدائي، ثانوي، جامعي، جامعي، جامعي، ثانوي، جامعي، جامعي، ثانوي، جامعي ابتدائي، جامعي.

مثل هذه البيانات في جدول تكراري.

الحل:

المستوى التعليمي	العلامات	التكرارت **fi**
ابتدائي	//	2
متوسط	///	3

ثانوي	/////	5
جامعي	///// /////	10
المجموع		20

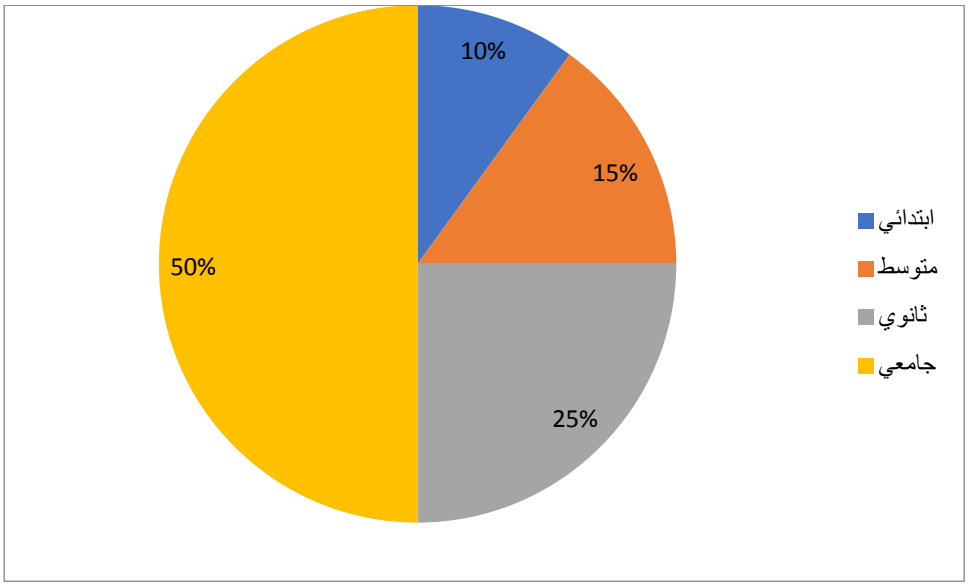
التمثيل البياني في حالة متغير كيفي:

1- **الدائرة النسبية** : تتمثل في دائرة بيانية مقسمة الى عدة اجزاء كل جزء يقابل زاوية مركزية تتناسب مع التكرار المقابل لكل خاصية من الخصائص المدروسى لتحقيق ذلك نقوم بحساب الزوايا المركزية المقابلة لكل تكرار بحيث: الزاوية المركزية $= \frac{\text{تكرار كل فئة}}{\text{مجموع تكرارات}} * 360$

لدينا الجدول التالي:

المستوى التعليمي	التكرارات Fi	التكرار النسبي fi	التكرار المئوي fi%	قياس الزاوية °
ابتدائي	2	0.10	10	°36
متوسط	3	0.15	15	°54
ثانوي	5	0.25	25	°90
جامعي	10	0.50	50	°180
المجموع	20	1	100	°360

تمثيل البيانات في شكل دائرة نسبية



المحاضرة الرابعة

العرض الجدولي (تابع)

جدول تكراري في حالة متغير كمي:

تمرين:

لنفرض أن الدراجات التي حصل عليها 70 طالب في أحد الاختبارات مرتبة ترتيبا تصاعديا وهي كما يلي:

40	44	46	46	47	49	50	51	52	52
53	54	55	55	56	57	58	58	59	60
60	60	60	61	61	61	62	62	63	63
63	63	64	64	65	65	65	66	66	66
66	67	67	67	68	68	68	69	69	69
69	69	71	72	73	73	73	74	74	75
76	78	79	79	81	81	82	84	84	84

المطلوب:

- أحسب طول الفئة (**Ki**) ؟
- كون الجدول التكراري.
- أحسب التكرار النسبي؟
- أحسب التكرار المئوي؟
- أحسب التكرار المتجمع الصاعد والنازل؟
- أرسم المدرج التكراري؟
- أرسم المضلع التكراري؟
- أرسم المنحنى التكراري؟

الحل:

حساب طول الفئة Ki

طول الفئة = المدى العام / عدد الفئات

هناك طريقتان لحساب طول الفئة Ki

1) عندما يكون N<100 $Ki = \frac{X_n - X_1}{\sqrt{N}}$

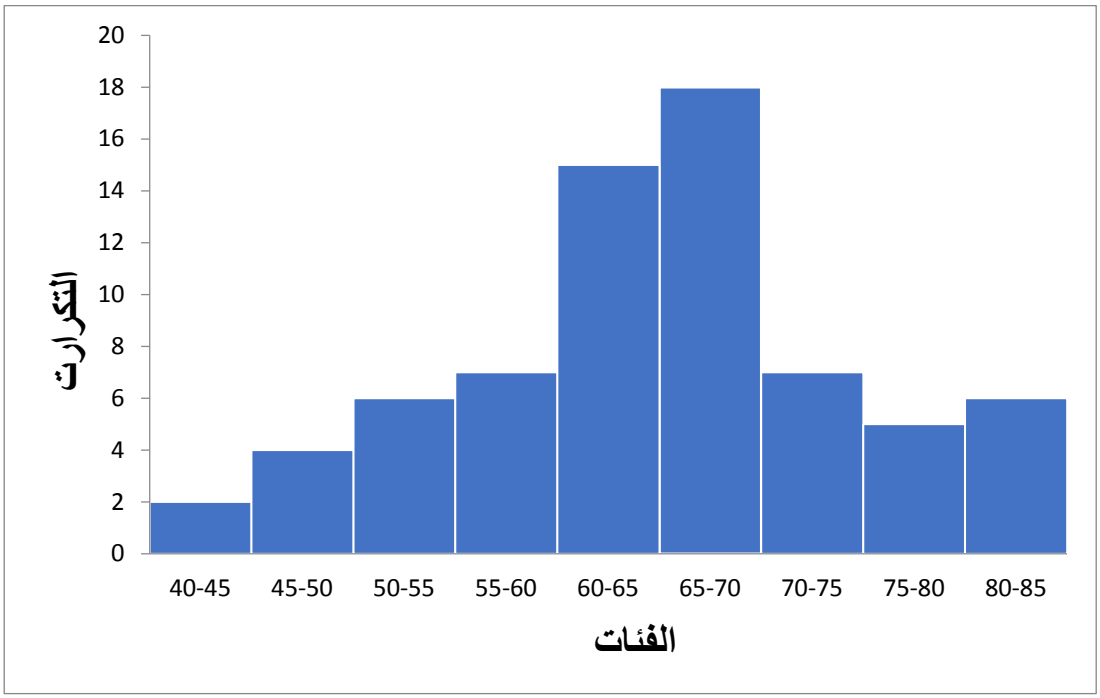
2) عندما يكون N≥ 100 $Ki = \frac{X_n - X_1}{1+3.32 \log N}$

$Ki = \frac{84-40}{\sqrt{70}} = \frac{44}{8.36} = 5.26 \approx 5$

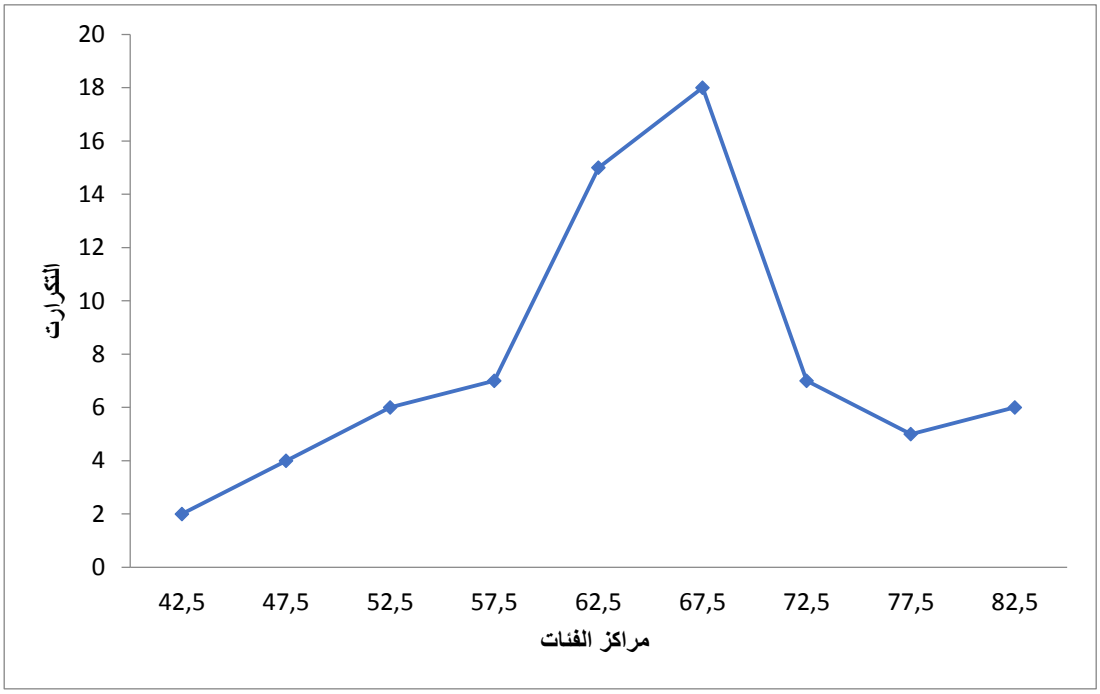
تكوين الجدول التكراري

الفئات x_i	العلامات	التكرار F_i	التكرار النسبي	التكرار المئوي	التكرار المتجمع الصاعد $F_i \nearrow$	التكرار المتجمع النازل $F_i \searrow$
45 – 40	//	2	0.03	3	2	70
50 – 45	////	4	0.06	6	6	68
55 – 50	/ /////	6	0.09	9	12	64
60 – 55	// /////	7	0.1	10	19	58
65 – 60	///// ///// /////	15	0.21	21	34	51
70 – 65	/// ///// ///// /////	18	0.25	25	52	36
75 – 70	// /////	7	0.1	10	59	18
80 – 75	/////	5	0.07	7	64	11
85 – 80	/ ////	6	0.09	9	70	6
المجموع		70	1	100		

رسم المدرج التكراري



رسم المضلع التكراري



المحاضرة الخامسة

مقاييس النزعة المركزية

المتوسط الحسابي

مقاييس النزعة المركزية:

تميل البيانات الإحصائية إلى التمركز حول قيمة معينة ،وكلما ابتعدنا عن هذه القيمة فإن عدد المعلومات يبدأ في التناقص ،ونسمي هذه الظاهرة بالنزعة المركزية ، بحيث يعبر قياس النزعة المركزية عن مركز التوزيع الاحصائي ،ولقياس هذه النزعة نستعمل عدة مقاييس من بينها:

1) المتوسط الحسابي ($\overline{X}$):

يعتبر المتوسط الحسابي من أهم مقاييس النزعة المركزية وأدقها في الحساب ، بحيث يمكن حسابه بمجموع القيم على عدد هذه القيم.

1.1 حساب المتوسط الحسابي في القيم غير مبوبة:

يمكن حساب المتوسط الحسابي في حالة البيانات غير مبوبة بالقانون التالي :

$$\overline{X} = (\sum x_i)/_N$$

مثال:

احسب المتوسط الحسابي لأجور اليومية بالدولار للعينة من العمال وهي كما يلي:

60,90,80,70,50

$$\overline{X} = \frac{\sum Xi}{\mathrm{N}} = \frac{60+90+80+70+50}{5} = 70\ \$$$

2.1 حساب المتوسط في حالة القيم المبوبة:

يمكن حساب المتوسط الحسابي في حالة البيانات المبوبة بالقانون التالي:

$$\overline{X} = (\sum \mathrm{fi}.\, x_i)/_N$$

مثال:

يمثل الجدول درجات 45 طالب في إحدى الامتحانات في مادة الرياضيات وهو كما يلي:

الفئات x_i	fi		fi . x_{ci}
2 – 0	3	1	3
4 – 2	6	3	18
6 – 4	9	5	45
8 – 6	10	7	70
10 – 8	8	9	72
12 –10	7	11	77
14 –12	2	13	26
المجموع	45		311

المطلوب: حساب المتوسط الحسابي

$$\overline{X} = \frac{\sum \text{fi}.x_i}{N} = \frac{311}{45} = \boxed{6.91}$$

مزايا وعيوب المتوسط الحسابي:

من مزايا المتوسط الحسابي أنه سهل الحساب وأهم مقياس من مقاييس النزعة المركزية، كما يأخذ بعين الاعتبار كل قيم التوزيع.

أما فيما يخص عيوب المتوسط الحسابي فتتمثل في تأثره بالقيم الشاذة والمتطرفة وكذلك يصعب حسابه في البيانات الاسمية والترتيبية، بالإضافة إلى ذلك يصعب حسابه في حالة الجداول التكرارية ذات الفئات المفتوحة.

المحاضرة السادسة

الوسيط

الوسيط (Med):

يعتبر الوسيط من أهم مقاييس النزعة المركزية ،بحيث هو تلك القيمة التي تقسم السلسلة الاحصائية إلى قسمين متساويين بعد ترتيب هذه القيم ترتيبا تصاعديا أو تنازليا.

1.2 حساب الوسيط في حالة البيانات غير مبوبة:

يتم تحديد الوسيط في حالة البيانات غير مبوبة حسب الخطوات التالية:

– ترتيب قيم السلسلة الإحصائية ترتيبا تصاعديا أو تنازليا.

– نحدد ترتيب الوسيط وهنا لابد أن نميز بين حالتين:

عندما يكون عدد القيم N فردي رتبة الوسيط هي: $\frac{N+1}{2}$

أما عندما يكون عدد القيم N زوجي هنالك رتبتين هما : $\frac{N}{2}$ و $\frac{N}{2}+1$ على التوالي.

مثال : في حالة N فردي تبين السلسلة الإحصائية التالية علامات 9 طلبة في مقياس الإحصاء

13 ، 14 ، 7 ، 17 ، 15 ، 9 ، 8 ، 11 ، 10.

أولا: ترتيب العلامات ترتبا تصاعديا 7 ، 8 ، 9 ، 10 ، $\boxed{11}$ ، 13 ، 14 ، 15 ، 17.

حساب رتبة الوسيط $\frac{9+1}{2}=5$ إذن: Med = 11

مثال : في حالة N زوجي تبين السلسلة الإحصائية التالية علامات 10 طلبة في مقياس الرياضيات

6 ، 7 ، 9 ، 10 ، 11 ، 12 ، 12 ، 13 ، 14 ، 15.

حساب رتبي الوسيط هما: $\frac{10}{2}=5$ و $\frac{10}{2}+1=6$

إذن : قيمة الوسيط هي: $\text{Med}=\frac{11+12}{2}=11.5$

2.2 حساب الوسيط في حالة البيانات المبوبة:

تختلف طريقة حساب الوسيط قي هذه الحالة عن الحلة الأولى وهذا حسب الخطوات التالية:

– تحديد قيم التكرار المتجمع الصاعد أو النازل.

– تحديد ترتيب الوسيط وهو عبارة عن منتصف مجموع التكرارات أي: $\frac{N}{2}$

– تحديد الفئة الوسيطية التي يقع فيها الوسيط وهي الفئة التي تقابل التكرار المتجمع الصاعد (النازل) الذي يساوي ترتيب الوسيط الأكبر أو الأصغر منه مباشرة.

– تحديد وحساب الوسيط بتطبيق العلاقة الإحصائية التالية:

$$Med = L_1 + \frac{\frac{N}{2} - F_1}{f_m}.ki$$

تحديد المجاهيل في القانون:

L_1: الحد الأول للفئة الوسيطية 25

$\frac{N}{2}$: منتصف مجموع القيم 50

F_1: التكرار المتجمع الصاعد القبلي 23

f_m : التكرار المقابل للفئة الوسيطية 28

Ki: طول الفئة الوسيطية 5

تمرين: ينثل الجدول توزيع 100 طفل بالنسبة للنساء المتزوجات من سن 15 إلى 50 سنة وهذا السن تكون فيه المرأة قادرة على الحمل (خصوبة المرأة).

الفئات	fi	$F_i\nearrow$	$F_i\searrow$
20 −15	8	8	100
25 −20	15	23	92
30 −25	28	51	77
35 −30	20	71	49
40 −35	15	86	29
45 −40	8	94	14
50 −45	6	100	6
المجموع	100		

المطلوب : حساب الوسيط

$$Med = 25 + \frac{50-23}{28} * 5 = 25 + 4.82 = 29.82 \text{ ans.} \quad \boxed{Med = 29.82\ ans}$$

مزايا وعيوب الوسيط:

من مزاياه أنه سهل الحساب ولا يتأثر بالقيم الشاذة والمتطرفة.

ومن عيوبه أنه لا يأخذ عند حسابه كل القيم بعين الاعتبار ، فهو يعتمد على قيمة أو قيمتين فقط.

كما أنه يصعب حسابه في حالة البيانات الوصفية المقاسة بمعيار اسمي.

المحاضرة السابعة

المنوال

المنوال (Mo):

يعتبر المنوال من أهم مقاييس النزعة المركزية ،بحيث هو القيمة الأثر انتشارا أو تكرارا ويستخدم للمتغيرات الاسمية.

أنواع المنوال:

- وحيدات المنوال والتي لها منوال واحد.
- ثنائية المنوال والتي لها منوالين.
- متعددة المنوال والتي لها أكثر من منوالين.
- عديمات المنوال والتي لا يوجد لها منوال.

المنوال في حالة البيانات غير مبوبة:

مثال: لدينا مجموعة من البيانات في ثلاثة أمثلة مختلفة وهي كالتالي:

أ) 6 ، 10 ، 6 ، 4 ، 9 ، 6 ، 2. $\boxed{Mo = 6}$

ب) 10 ، 5 ، 7 ، 9 ، 8 ، 7 ، 5 ، 7 ، 5. $\boxed{Mo_2 = 7}$ $\boxed{Mo_1 = 5}$

ج) 15 ، 7 ، 11 ، 12 ، 8 ، 9 ، 4. لا يوجد منوال

حساب المنوال في حالة البيانات المبوبة:

ان طريقة حساب المنوال في حالة البيانات المبوبة لا بد من اتباع الخطوات التالية:

1) تحديد الفئة المنوالية وهي الفئة التي تقابل أكبر تكرار عندما يكون طول الفئة ثابت ،أو الفئة التي تقابل أكبر تكرار معدل عندما تكون أطوال الفئات غير متساوية.

2) وضع المدرج التكراري قصد تحديد علاقة المنوال ،والعلاقة الأكثر انتشارا هي علاقة الفروق أو علاقة بيرسون.

تمرين:

لدينا توزيع 30 أسرة حسب الانفاق الاستهلاكي الشهري لها بالألف وهو موضح في الجدول التالي:

فئات الانفاق x_i	20 -15	25 -20	30 -25	35 -30	40 -35
عدد الأسر fi	4	7	10	5	4

المطلوب:

حساب المنوال (Mo)

$$Mo = L_1 + \frac{\Delta_1}{\Delta_1+\Delta_2}.K_i$$

L_1: الحد الأول للفئة المنوالية هو 25

Δ_1: الفرق بين أكبر تكرار والتكرار القبلي (*10-7*)

Δ_2: الفرق بين أكير تكرار والتكرار البعدي (10-5)

k_i: طول الفئة المنوالية هو5.

بالتعويض في المعادلة نجد:

$$\mathbf{Mo}=25+\frac{(10-7)}{(10-7)+(10-5)}.5$$

$$\boxed{\boldsymbol{Mo = 26.87}} \quad \boxed{Mo = 26870\ DA}$$

المحاضرة الثامنة

مقاييس التشتت

المدى والانحراف المتوسط

لقد ذكرنا في الفصل السابق بعضًا من مقاييس النزعة المركزية والتي تمثل مقاييس عددية لموضع أو مكان تركز البيانات لظاهرة ما. وقد ذكرنا بأن هذه المقاييس تستخدم لمقارنة مجموعات البيانات المختلفة. وفي الحقيقة فإن مقاييس النزعة المركزية غير كافية لإيجاد مقارنة شاملة بين مجموعات البيانات المختلفة. فقد تكون هناك مجموعات من البيانات لها نفس مقاييس النزعة المركزية (لها نفس الموضع) ولكنها تختلف في بعض الصفات الأخرى. فمثلاً المثال التالي يبين لنا مجموعتين من البيانات لهما نفس المتوسط ولكنهما مختلفتان في طبيعة تشتت البيانات.

مثال (1):

الجدول التالي يوضح مجموعتين لهما نفس المتوسط $\bar{x} = 60$ ونفس الوسيط $Med = 60$ ونفس المنوال $Mod = 60$ ولكنهما مختلفتان في طبيعة التشتت.

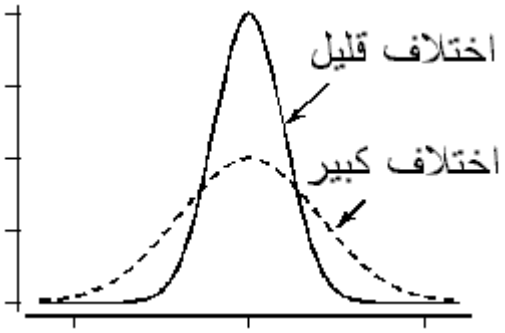
المجموعة	البيانات	شكل انتشار البيانات
الأولى	59, 61, 62, 60, 58, 60	• – ● – ● – ●– ● – ● – 60
الثانية	50, 60, 66, 54, 60, 70	• – ● – – ● – – – – ●– – – – ● – – ● – 60

بالرغم من تساوي مقاييس النزعة المركزية للمجموعتين إلا أن التشتت (أو الاختلاف) بين القيم في كل مجموعة غير متساو. فمن الواضح أن بيانات المجموعة الأولى أكثر تقاربًا فيما بينها (أقل تشتتًا وتباعدًا فيما بينها) من بيانات المجموعة الثانية. لذلك دعت الحاجة لإيجاد مقاييس تقيس طبيعة تشتت (أو تفرق أو اختلاف أو تباعد) البيانات فيما بينها. هذه المقاييس تسمى مقاييس التشتت أو الاختلاف.

مثال (2):

الشكل التالي يوضح المضلعين التكراريين لمجموعتين من البيانات لهما نفس مقاييس النزعة المركزية ولكنهما

مختلفتان في طبيعة التشتت.



شكل (1) المضلعان التكراريان لتوزيعين لهما نفس مقاييس
النزعة المركزية ولكنهما مختلفين في التشتت

مقاييس التشتت هي مقاييس عددية تستخدم لقياس اختلاف أو تشتت البيانات. والاختلاف أو التشتت لمجموعة من البيانات هو مقدار تفرق أو تباعد أو انتشار البيانات فيما بينها. فتشتت البيانات يكون صغيرًا إذا كانت البيانات متقاربة فيما بينها والعكس بالعكس. وأما البيانات المتساوية فلا اختلاف ولا تشتت فيها. ومقاييس التشتت تستخدم لوصف مجموعة البيانات وكذلك لمقارنة مجموعات البيانات المختلفة إذ أن مقاييس النزعة المركزية لا تكفي وحدها لوصف مجموعة البيانات أو مقارنة مجموعات البيانات المختلفة. ومن أشهر مقاييس التشتت نذكر:

المدى: Range

يعتبر المدى من أسهل مقاييس التشتت تعريفًا وحسابًا ويعطينا فكرة سريعة عن مدى تفرق البيانات.

تعريف (1):

نعرف المدى لمجموعة من البيانات على أنه الفرق بين أكبر قيمة وأصغر قيمة، أي أن المدى هو:

$$Range = X_{max} - X_{min}$$

حيث نعرف X_{max} و X_{max} كما يلي:

(أ) للبيانات المفردة:

X_{max} = أكبر قيمة

X_{min} = أصغر قيمة

(ب) للبيانات المبوبة:

X_{max} = مركز الفترة العليا

X_{min} = مركز الفترة الدنيا

ملاحظة (1):

تعرف بعض الكتب أكبر قيمة وأصغر قيمة للبيانات المبوبة كما يلي:

X_{max} = الحد الأعلى للفترة العليا

X_{min} = الحد الأدنى للفترة الدنيا

مثال (3):

أوجد المدى للمشاهدات التالية والتي هي عبارة عن أوزان (بالكيلوجرام) مجموعة مكونة من سبعة أشخاص:

25, 30, 40, 45, 35, 55, 50

الحل:

$$X_{max} = 55$$
$$X_{min} = 25$$
$$Range = X_{max} - X_{min} = 55 - 25 = 30 \quad \text{(كيلوجرامًا)}$$

مثال (4):

أوجد المدى لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (2) من الفصل الثاني.

الحل:

مستوى الهيموجلوبين	مركز الفترة x	التكرار f
12.95 – 13.95	**13.45**	3
13.9 5– 14.95	14.45	5
14. 95– 15.95	15.45	15
15. 95– 16.95	16.45	16
16. 95– 17.95	17.45	10
17. 95– 18.95	**18.45**	1

X_{max} = مركز الفترة العليا = 18.45

X_{min} = مركز الفترة الدنيا = 13.45

$$Range = X_{max} - X_{min} = 18.45 - 13.45 = 5.00$$

بعض مميزات وعيوب المدى:

أ- يتميز المدى بسهولة التعريف والحساب

ب- يعيب المدى العيوب التالية:

1. يتأثر المدى بالقيم الشاذة أو المتطرفة.
2. لا يأخذ المدى في الاعتبار جميع البيانات.

الانحراف المتوسط:

هو أحد مقاييس التشتت التي تخطر على البال هو مجموع انحرافات البيانات عن وسطها الحسابي. ولكن هذا المجموع يساوي دائما صفرا لأن مجموع الانحرافات الموجبة عن الوسط الحسابي يساوي مجموع الانحرافات السالبة.

لذلك لا بد من التخلص من الإشارة السالبة حتى نحصل على مقياس ذي معنى. أن إحدى الطرق المستخدمة في التخلص من الإشارة السالبة هي عن طريق أخذ القيمة المطلقة.

يعتمد الانحراف المتوسط على جميع مفردات البيانات وهو سهل التعريف وسهل الحساب إلا أنه لا يخضع للعمليات الجبرية بسهولة حيث يجب تعديل الإشارة ويجب معرفة المفردات بعينها إذا ما أردنا حساب قيمته.

ويتضح هذا من عدم وجود طريقة جبرية لحساب الانحراف المتوسط للمجموعة الناتجة عن دمج مجموعتين من البيانات إذا علم عدد مفردات كل منها ووسطها الحسابي وانحرافها المتوسط. ففي هذه الحالة يجب معرفة جميع المفردات لنتمكن من حساب انحرافها المتوسط، وبذلك فإنه من ادق مقاييس التشتت على الإطلاق.

حساب الانحراف المتوسط في حالة القيم غير مبوبة:

يمكن حساب الانحراف المتوسط بالمعادلة التالية:

$$\text{MD} = \frac{\sum|x - \bar{X}|}{N}$$

مثال:

$$\bar{X} = \frac{\sum x_i}{N}$$

$$\bar{X} = \frac{28}{5}$$

$$\bar{X} = 5.6$$

	$(x_i - \bar{X})$	
4	$(4 - 5.6) = -1.6$	1.6
5	$(5 - 5.6) = -0.6$	0.6
2	$(2 - 5.6) = -3.6$	3.6
10	$(10 - 5.6) = 4.4$	4.4
7	$(7 - 5.6) = 1.4$	1.4
28	0	11.6

حساب الانحراف المتوسط:

$$MD = \frac{\sum|x-\bar{X}|}{N}$$

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

$$MD = 2.32$$

حساب الانحراف المتوسط في حالة القيم المبوبة بالمعادلة التالية:

$$MD = \frac{\sum f_i|x-\bar{X}|}{N}$$

مثال:

أوجد الانحراف المتوسط للتوزيع التكراري التالي:

الفئات i	التكرار f_i	مركز الفئة x_i			$(x_i - \bar{X})$	$\lvert$	f
5−0	6	2.5		15	10.5 −	10.5	61.2
10−5	3	7.5		22.5	5.2 −	5.2	15.6
15−10	4	12.5	12.7	50	0.2 −	0.2	0.8
20−15	12	17.5		210	4.8	4.8	57.6
25−20	2	22.5		45	9.8	9.8	19.6
المجموع	27			342.5			154.8

$$MD = \frac{\sum f_i|x-\bar{X}|}{N}$$

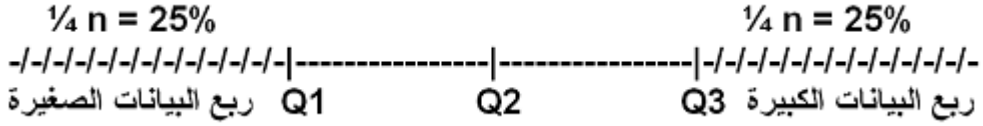
$$MD = \frac{154.8}{27}$$

$$MD = 5.73$$

المحاضرة التاسعة

الانحراف الربيعي

رأينا أن المدى يتأثر كثيرًا بالقيم الشاذة أو المتطرفة. ولذلك دعت الحاجة لإيجاد مقاييس أخرى للتشتت لا تتأثر بالقيم المتطرفة. وأحد هذه المقاييس هو الانحراف الربيعي. وحيث أن القيم المتطرفة هي تلك القيم الصغيرة جدًا أو الكبيرة جدًا فإنه عند حساب الانحراف الربيعي لا يؤخذ في الاعتبار ربع البيانات الصغيرة (25%) ولا ربع البيانات الكبيرة (25%). الشكل التالي يوضح موضع القيم الشاذة والمتطرفة.



يرمز للانحراف الربيعي بالرمز Q ويعرف بالصيغة التالية:

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

حيث أن Q_1 هو الربيع الأول و Q_3 هو الربيع الثالث وقد مر معنا كيفية إيجادهما للبيانات المبوبة بالطريقة الحسابية والبيانية.

مثال (5):

أوجد الانحراف الربيعي لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في موضح في الجدول التالي:

مستوى الهيموجلوبين	13.95 -12.95	14.95 -13.95	15.95 -14.95	16.95 -15.95	17.95 -16.95	18.95 -17.95	المجموع
عدد الأشخاص	3	5	15	16	10	1	50

الحل:

حساب الربيع الأول Q_1:

$$R = \frac{n}{4} = \frac{50}{4} = 12.5$$

$$A = 14.95, \quad L = 1.00, \quad F_1 = 8 \ , \ F_2 = 23$$

$$Q_1 = A + \left(\frac{R - F_1}{F_2 - F_1}\right) \times L = 14.95 + \left(\frac{12.5 - 8}{23 - 8}\right) \times 1.00 = 15.25$$

حساب الربيع الثالث Q_3:

$$R^* = \frac{3n}{4} = \frac{3\times 50}{4} = 37.5$$

$$A^* = 15.95,\quad L^* = 1.00,\quad F_1^* = 23\ ,\ F_2^* = 39$$

$$Q_3 = A^* + \left(\frac{R^* - F_1^*}{F_2^* - F_1^*}\right)\times L^* = 15.95 + \left(\frac{37.5-23}{39-23}\right)\times 1.00 = 16.86$$

وباستخدام الصيغة فإن الانحراف الربيعي هو:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{16.86 - 15.25}{2} = \frac{1.61}{2} = 0.805$$

	مستوى الهيموجلوبين	التكرار المتجمع الصاعد	
	12.95 – 13.95	3	
	13.95 – 14.95	8	
$Q_1 \Rightarrow$	14.95 – 15.95	23	
$Q_3 \Rightarrow$	15.95 – 16.95	39	$\leftarrow R = \frac{n}{4} = 12.5$
	16.95 – 17.95	49	$\leftarrow R^* = \frac{3n}{4} = 37.5$
	17.95 – 18.95	50	

المحاضرة العاشرة

التباين والانحراف المعياري

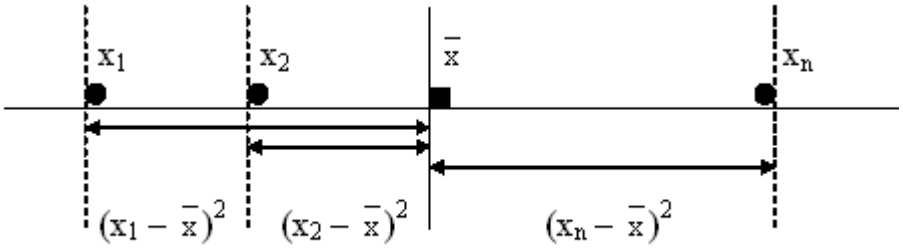
التباين (Variance) والانحراف المعياري (Standard Deviation):

يعتبر التباين والانحراف المعياري من أهم وأفضل مقاييس التشتت ومن أكثرها شيوعًا واستخدامًا في التحليل الإحصائي وذلك لما يتمتعان به من خصائص وصفات إحصائية جيدة.

التباين: Variance

فكرة التباين تعتمد على تشتت أو تباعد البيانات عن متوسطها. فالتباين يكون كبيرًا إذا كانت البيانات متباعدة عن متوسطها والعكس بالعكس.

ويعرف التباين بأنه متوسط مربع انحرافات القيم عن الوسط الحسابي ويرمز له بالرمز S^2. الشكل التالي يبين مربعات انحرافات القيم عن الوسط الحسابي.



والجدول التالي يلخص طريقة حساب انحرافات ومربعات انحرافات القيم عن المتوسط.

القيم (البيانات)	x_n	⋯	x_2	x_1
انحرافات القيم عن المتوسط	$x_n - \bar{x}$	⋯	$x_2 - \bar{x}$	$x_1 - \bar{x}$
مربع انحرافات القيم عن المتوسط	$(x_n - \bar{x})^2$	⋯	$(x_2 - \bar{x})^2$	$(x_1 - \bar{x})^2$

تستخدم مربعات انحرافات القيم عن الوسط الحسابي لحساب التباين. فمن الواضح أن تشتت البيانات يزداد بزيادة مربعات الانحرافات والعكس بالعكس. والسؤال الذي يتبادر للأذهان هو أي واحد من هذه المربعات ينبغي أن يستخدم لقياس التشتت؟ إن من المنطقي أن نبحث عن قيمة نموذجية تمثل هذه المربعات لاستخدامها لقياس التباين. وبناءً على دراستنا السابقة لمقاييس النزعة المركزية فإن المتوسط من أفضل المقاييس التي تمثل مجموعة البيانات. وعليه فإن متوسط مربعات انحرافات القيم عن الوسط الحسابي $\sum_{i=1}^{n}(x_i - \bar{x})^2 / n$ يعتبر أحد المقاييس التي يمكن استخدامها لقياس التباين.

الانحراف المعياري: Standard Deviation

إن التباين من أهم وأفضل مقاييس التشتت ولكنه يقاس بوحدة البيانات الأصلية المربعة. وفي كثير من الأحيان نرغب في استخدام مقياس للتشتت يقاس بوحدة البيانات الأصلية ويتمتع بخصائص إحصائية جيدة مثل التباين. وأحد هذه المقاييس هو الانحراف المعياري. ويعرف الانحراف المعياري على أنه الجذر التربيعي الموجب للتباين ويرمز له بالرمز s.

حساب التباين والانحراف المعياري:

سنستعرض طرق حساب التباين والانحراف المعياري في حالة البيانات المفردة وفي حالة البيانات المبوبة.

أولاً: التباين والانحراف المعياري للبيانات المفردة (غير المبوبة):

إذا كانت $x_1, x_2, \ldots, x_n$ عينه حجمها n وكان متوسطها هو $\bar{x}$ فإن تباين العينة يعرف كما يلي:

$$s^2 = \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \mathrm{L} + (x_n-\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} \qquad (3)$$

لاحظ أننا قسمنا على المقدار $(n-1)$، وهو ما يسمى بدرجات الحرية، بدلاً من القسمة على عدد البيانات n في الصيغة السابقة وذلك لكي نحصل على مقياس يتمتع بصفات إحصائية جيدة.

وأما الانحراف المعياري فإنه يعرف بالصيغة:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} \qquad (4)$$

ملاحظة (4):

1. $s^2 \geq 0$ (دائمًا) وكذلك $s \geq 0$ (دائمًا).
2. $s = 0 \Leftrightarrow s^2 = 0 \Leftrightarrow$ جميع قيم العينة متساوية (لا يوجد اختلاف بين القيم).
3. وحدة التباين، s^2، هي وحدة البيانات الأصلية المربعة.
4. وحدة الانحراف المعياري، s، هي نفس وحدة البيانات الأصلية.
5. يمكن حساب التباين بإحدى الصيغتين الحسابيتين التاليتين:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^{\,2} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right\} \qquad (5)$$

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^{\,2} - n\left(\bar{x}\right)^2\right\} \qquad (6)$$

ولحساب تباين العينة باستخدام الصيغتين الحسابيتين السابقتين فإننا نحتاج إلى معرفة الكميات التالية فقط دون الحاجة لمعرفة البيانات الأصلية:

1. حجم العينة = n.
2. مجموع البيانات = $\sum_{i=1}^{n} x_i$.
3. مجموع مربعات البيانات = $\sum_{i=1}^{n} x_i^2$.

والصيغتين الحسابيتين السابقتين تستخدمان لحساب تباين العينة وذلك لسببين هما:

1. لأنهما أكثر سهولة من صيغة التعريف.
2. لأنها أكثر دقة في الحساب من صيغة التعريف عندما يكون هناك تقريب في حساب متوسط العينة.

مثال (6):

أوجد تباين العينة والانحراف المعياري لمجموعة الأوزان (بالكيلوجرام) التالية:

7.1, 2.5, 2.5, 5.4, 8.3

الحل:

نلخص الحل في الجدول التالي:

x_i	الانحراف $(x_i - \bar{x}) = (x_i - 5.16)$	مربع الانحراف $(x_i - \bar{x})^2$	x^2
7.1	1.94	3.7636	50.41
2.5	–2.66	7.0756	6.25
2.5	–2.66	7.0756	6.25
5.4	0.24	0.0576	29.16
8.3	3.14	9.8596	68.89
$\sum_{i=1}^{n} x_i = 25.8$	0.00	$\sum_{i=1}^{n} (x_i - \bar{x})^2 = 27.8$	$\sum_{i=1}^{n} x_i^2 = 160.9$

من هذا الجدول نوجد الكميات التالية:

$$n = 5,\ \sum_{i=1}^{n} x_i = 25.8,\ \sum_{i=1}^{n} (x_i - \bar{x})^2 = 27.832,\ \sum_{i=1}^{n} x_i^2 = 160.96$$

إن متوسط العينة هو:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{25.8}{5} = 5.16 \text{ (كيلوجرامًا)}$$

حساب تباين العينة:

(أ) باستخدام التعريف:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{27.832}{5-1} = 6.958 \text{ (كيلوجرامًا مربعًا)}$$

(ب) باستخدام الصيغة الحسابية الأولى:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i{}^2 - \left(\sum_{i=1}^{n} x_i\right)^2 / n\right\} = \frac{1}{5-1}\left\{160.96 - (25.8)^2 / 5\right\}$$

$$= \frac{160.96 - 133.128}{4} = \frac{27.832}{4} = 6.958$$

(ج) باستخدام الصيغة الحسابية الثانية:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i{}^2 - n\ (\bar{x})^2\right\} = \frac{1}{5-1}\left\{160.96 - (5)(5.16)^2\right\}$$

$$= \frac{160.96 - 133.128}{4} = \frac{27.832}{4} = 6.958$$

الانحراف المعياري هو:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = s = \sqrt{6.958} = 2.6378 \text{ (كيلوجرامًا)}$$

ثانياً: التباين والانحراف المعياري للبيانات المبوبة:

إذا كان لدينا بيانات عددها n وكانت هذه البيانات ملخصة في جدول تكراري بحيث أن:

أ- عدد الفترات هو k.

ب- مراكز الفترات هي $x_1, x_2, \ldots, x_k$.

ج- تكرارات الفترات هي $f_1, f_2, \ldots, f_k$.

بطريقة مشابهة لحسب المتوسط للتوزيع التكراري فإن التباين للتوزيع التكراري المبوب يمكن حسابه بشكل تقريبي بالصيغة التالية:

$$s^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{n-1}$$

حيث أن:

$$n = \sum_{i=1}^{k} f_i \ , \quad \bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum xf}{n} \ ,$$

كما يمكن استخدام إحدى الصيغتين الحسابيتين التاليتين:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{k} x_i^2 f_i - \left(\sum_{i=1}^{k} x_i f_i\right)^2 / n\right\}$$

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{k} x_i^2 f_i - n(\bar{x})^2\right\}$$

ويكمن تلخيص عمليتي إيجاد المتوسط والتباين باستخدام الجدول التالي:

الفترة	مركز الفترة x	التكرار f	xf	$x^2 f$	$f(x-\bar{x})^2$
الفترة رقم 1	x_1	f_1	$x_1 f_1$	$x_1^2 f_1$	$f_1 (x_1-\bar{x})^2$
الفترة رقم 2	x_2	f_2	$x_2 f_2$	$x_2^2 f_2$	$f_2 (x_2-\bar{x})^2$
⋮	⋮	⋮	⋮	⋮	⋮
الفترة رقم k	x_k	f_k	$x_k f_k$	$x_k^2 f_k$	$f_k (x_k-ر)^2$
المجموع		$\sum f = n$	$\sum xf$	$\sum x^2 f$	$\sum_{i=1}^{k} f_i (x_i-\bar{x})^2$

مثال (10):

أوجد التباين والانحراف المعياري لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (2) من الفصل الثاني.

الحل:

نلخص إيجاد التباين باستخدام الجدول التالي:

مستوى الهيموجلوبين	مركز الفترة x	التكرار f	xf	$x^2 f$	$f(x-\bar{x})^2$ $f(x-16.01)^2$
12.95 – 13.95	13.45	3	40.35	542.708	19.6608
13.9 5– 14.95	14.45	5	72.25	1044.013	12.1680
14. 95– 15.95	15.45	15	231.75	3580.538	4.7040
15. 95– 16.95	16.45	16	263.20	4329.640	3.0976
16. 95– 17.95	17.45	10	174.50	3045.025	20.7360
17. 95– 18.95	18.45	1	18.45	340.403	5.9536
المجموع		$\sum f = 50$	$\sum xf = 800.5$	$\sum x^2 f = 12882.33$	$\sum_{i=1}^{k} f_i (x_i - \bar{x})^2 = 66.320$

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum xf}{n} = \frac{800.5}{50} = 16.01$$

حساب التباين باستخدام صيغة التعريف:

$$s^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{n-1} = \frac{66.320}{50-1} = 1.3535$$

حساب التباين باستخدام الصيغة الحسابية الأولى:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{k} x_i^2 f_i - \left(\sum_{i=1}^{k} x_i f_i\right)^2 / n\right\}$$

$$= \frac{1}{50-1}\left(12882.33 - \frac{(800.5)^2}{50}\right)$$

$$= \frac{1}{49}(12882.33 - 12816.005)$$

$$= \frac{66.325}{49} = 1.3536$$

حساب التباين باستخدام الصيغة الحسابية الثانية:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{k} x_i^2 f_i - n(\bar{x})^2\right\}$$

$$= \frac{1}{50-1}\{12882.33 - (50)(16.01)^2\}$$

$$=\frac{1}{49}\{12882.33-12816.005\}$$
$$=\frac{66.325}{49}=1.3536$$

حساب الانحراف المعياري:

$$s=\sqrt{s^2}=\sqrt{1.3536}=1.163$$

المحاضرة الحادية عشرة

الالتواء

عند تمثيل بيانات ظاهرة ما في شكل منحنى تكراري، فإن هذا المنحنى يأخذ أشكالا مختلفة، بحيث قد يكون هذا المنحنى متماثل يعني أن مقاييس النزعة المركزية المتوسط والوسيط والمنوال كلهم يقعون على نقطة واحدة، ولكن في كثير من الحالات يكون هناك قيم كبيرة في البيانات تجذب اليها المتوسط الحسابي، وهذا معناه أن المنحنى التكراري يدل على وجود التواء جهة اليمين، وكذلك العكس عندما تكون هناك قيم صغيرة في البيانات تجذب إليها المتوسط الحسابي، مما يدل أن المنحنى التكراري يوجد له التواء نحو اليسار، إلا أن هناك مقاييس كثيرة لوصف البيانات تعتمد في حسابها على مقاييس النزعة المركزية ومقاييس التشتت معا، ومنها مقاييس الالتواء.

مقاييس الالتواء:

هناك طرق كثيرة لقياس الالتواء ومنها مايلي:

معامل الالتواء لبيرسون (α_p):

لحساب معامل الالتواء لبيرسون لا بد أن نأخذ في الاعتبار العلاقة بين المتوسط والوسيط والانحراف المعياري، في حالة ما إذا كان التوزيع قريب من التماثل وليس شديد الالتواء، بحيث تكون العلاقة كما يلي:

$$\boldsymbol{\alpha_p=\frac{3(\bar{X}-Med)}{S}}$$

بحيث:

α_p : معامل الالتواء لبيرسون

$\bar{X}$: المتوسط الحسابي

Med : الوسيط

S : الانحراف المعياري

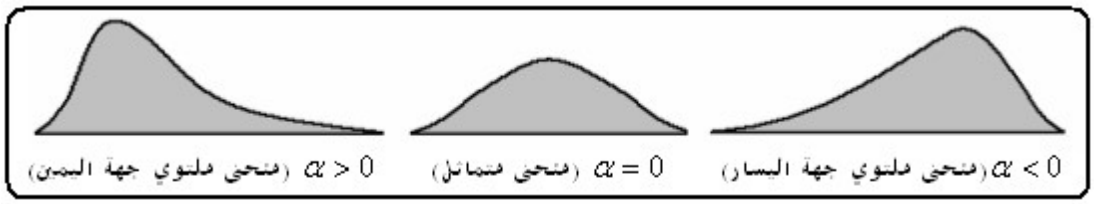
ويمكن من خلال الإشارة التي يأخذها معامل الالتواء لبيرسون الحكم على شكل الالتواء كما يلي:

* اذا كان (المتوسط = الوسيط) كانت قيمة المعامل ($\alpha_p = 0$)، ويدل ذلك على أن منحنى التوزيع التكراري متماثل.

* اذا كان (المتوسط > الوسيط) كانت قيمة المعامل ($\alpha_p > 0$)، ويدل ذلك على أن منحنى التوزيع التكراري ملتوي جهة اليمين.

* اذا كان (المتوسط < الوسيط) كانت قيمة المعامل ($\alpha_p < 0$)، ويدل ذلك على أن منحنى التوزيع التكراري ملتوي جهة اليسار.

أشكال التواء البيانات



معامل الالتواء الربيعي:

الربيع هو ينتج عن ترتيب البيانات تصاعديا، بحيث تقسم البيانات إلى أربعة أجزاء، يفصل بينها قيم تسمى الربيعيات وهي (الربيع الأول (Q_1) والربيع الثاني (Q_2) والربيع الثالث (Q_3)).

ونعتمد فكرة الربيع في قياس الالتواء على مدى قرب الربيع الأول والربيع الثالث من الربيع الثاني والذي هو الوسيط.

ويمكن من خلال الإشارة التي يأخذ معامل الالتواء الربيعي الحكم على شكل الالتواء كما يلي:

* إذا كان بعد الربيع الثالث عن الربيع الثاني يساوي بعد الربيع الأول عن بعد الربيع الثاني كان التوزيع متماثل.

* إذا كان بعد الربيع الثالث عن الربيع الثاني أكبر من بعد الربيع الأول عن بعد الربيع الثاني كان التوزيع ملتوي نحو اليمين.

* إذا كان بعد الربيع الثالث عن الربيع الثاني أصغر من بعد الربيع الأول عن بعد الربيع الثاني كان التوزيع ملتوي نحو اليسار.

وبشكل عام يمكن الحكم على شكل التوزيع باستخدام معامل الالتواء الربيعي، ويأخذ المعادلة التالية:

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

مثال:

كانت درجات 8 طلاب في الاختبار النهائي في مادة الإحصاء التربوي وهي كما يلي:

58 74 91 80 78 52 85 66

المطلوب:

1- حساب معامل الالتواء لبيرسون.

2- حساب معامل الالتواء الربيعي.

الحل:

لحساب معامل الالتواء لبيرسون لا بد من حساب المتوسط الحسابي والانحراف المعياري والوسيط.

66	4356
85	7225
52	2704
78	6084
80	6400
91	8281
74	5476

3364	58
43890	584

$\sum x_i = 584$

$\sum x_i^2 = 43890$

$\bar{X} = \frac{\sum x_i}{n} = \frac{584}{8} = 73$

$\bar{X} = 73$

$$S = \sqrt{\frac{\sum x_i^2 - n\bar{X}^2}{n-1}} = \sqrt{\frac{43890 - 8(73)^2}{8-1}} = \sqrt{\frac{1258}{7}}$$

$S = \sqrt{179.71} = 13.41$

$S = 13.41$

91 85 80 78 74 66 58 52

$\frac{n}{2} = \frac{8}{2} = 4$

$\frac{n}{2} + 1 = \frac{8}{2} + 1 = 5$ $\quad$ $Med = \frac{74+78}{2} = 76$

$$\alpha_p = \frac{3(\bar{X} - Med)}{S}$$

$$\alpha_p = \frac{3(\overline{73} - 76)}{13.41}$$

$$\alpha_p = -0.67$$

وعليه قيمة معامل الالتواء سالبة يعني أن التوزيع التكراري للدرجات ملتوي نحو اليسار.

حساب معامل الالتواء الربيعي كما يلي:

حساب الربيع الأول

$(n + 1)/4 = (8+1)/4 = 2.25$

$Q_1 = \frac{58+66}{2} = 62$

حساب الربيع الثالث:

$(n + 1)3/4 = (8 + 1)3/4 = 6.75$

$Q3 = \frac{80+85}{2} = 82.5$

$Q2 = Med = 76$

$$\alpha_Q = \frac{(Q_3-Q_2)-(Q_2-Q_1)}{(Q_3-Q_1)} = \frac{(82.5-76)-(76-62)}{(82.5-62)}$$

$$\alpha_Q = \frac{-7.5}{20.5} = -0.37$$

وعليه قيمة معامل الالتواء الربيعي سالبة وتعني أن الربيع الثالث يقترب إلى الربيع الثاني، وهذا ما يدل على التوزيع التكراري للدرجات ملتوي نحو اليسار.

المحاضرة الثانية عشرة

معامل الاختلاف والدرجات المعيارية

معامل الاختلاف (التغير): Coefficient of Variation:

ذكرنا سابقًا أن التباين والانحراف المعياري من المقاييس المفيدة لقياس التشتت لتوزيع متغير ما. ولكن في كثير من الأحيان نكون مهتمين بمقارنة التشتت والاختلاف لتوزيعي متغيرين مختلفين. وبما أن التباين والانحراف المعياري مقياسان يعتمدان على وحدة البيانات فإنه يصعب استخدامهما لمقارنة تجانس المجموعات المختلفة من البيانات وذلك لاختلاف الوحدة المستخدمة. وبشكل عام فإن مقاييس التشتت التي ذكرناها آنفًا تكون غير مناسبة لمقارنة تجانس مجموعات البيانات المختلفة في الحالتين التاليتين:

1. إذا كانت وحدتا المتغيرين مختلفتين حيث لا نستطيع مقارنة الوحدات المختلفة ببعض.
2. إذا كان متوسطا المتغيرين مختلفين وذلك لأن تباين توزيع المتغير ذي المتوسط الصغير ينزع لأن يكون صغيرًا والعكس بالعكس.

لذلك دعت الحاجة إلى مقياس لا يعتمد على وحدة المتغير ويقيس ما يسمى بالتشتت النسبي. وأحد هذه المقاييس هو ما يسمى بمعامل الاختلاف أو معامل التغير. فمعامل الاختلاف هو أحد مقاييس التشتت النسبي وهو مقياس عديم الوحدة ويستخدم لمقارنة التشتت النسبي أو التجانس لمجموعات البيانات المختلفة. فمجموعة

البيانات ذات معامل الاختلاف الأكبر يكون تشتتها النسبي أكبر أي أنها تكون أقل تجانسًا والعكس بالعكس. ويعرف معامل الاختلاف للعينة التي متوسطها $\overline{x}$ وانحرافها المعياري s بالصيغة التالية:

$$C.V. = \frac{s}{\overline{x}}$$

مثال (11):

الجدول أدناه يتضمن بيانات إحدى الدراسات التي طبقت على خمسة أشخاص لقياس الوزن (بالكيلوجرام) والطول (بالسنتيمتر). أي البيانات أكثر تشتتًا نسبًا (أقل تجانسًا)؛ بيانات الأوزان أم بيانات الأطوال؟

رقم الشخص	1	2	3	4	5
الوزن	69	59	65	67	65
الطول	164	162	155	165	158

الحل:

أولاً نوجد المتوسط $\overline{x}$ والانحراف المعياري s لكل من بيانات الأوزان وبيانات الأطوال كما مر معنا سابقاً. نلخص الحسابات في الجدول التالي:

البيانات	المتوسط $\overline{x}$	الانحراف المعياري s	معامل الاختلاف $C.V. = \frac{s}{\overline{x}}$
الأوزان	65.0 kg	3.7417 kg	0.0576
الأطوال	160.8 cm	4.2071 cm	0.026

بما أن معامل الاختلاف لبيانات الأوزان أكبر من معامل الاختلاف لبيانات الأطوال فإن التشتت النسبي لبيانات الأوزان أكبر من التشتت النسبي لبيانات الأطوال. أي أن بيانات الأوزان أقل تجانسًا من بيانات الأطوال.

(4-7) الدرجات (القيم) المعيارية : Standard Scores (Values)

نستطيع بكل يسر وسهولة استخدام قيمتي مشاهدتين في نفس المجموعة لمقارنتهما ببعض. فمثلاً، نستطيع أن نقول بأن أداء الطالب الحائز على الدرجة 85 في اختبار مقرر ما أفضل من أداء الطالب الحائز على الدرجة 80 في نفس الاختبار إذا كان الطالبان في نفس الشعبة. وفي المقابل، لا نستطيع القول بأن أداء الطالب الحائز على الدرجة 85 في اختبار مقرر ما في الشعبة التي يدرسها المدرس (أ) أفضل من أداء طالب آخر حائز على الدرجة 80 في نفس الاختبار ولكنه في شعبة أخرى يدرسها المدرس (ب). من هنا، نرى أنه من الضروري إيجاد

قيم لا تعتمد على الوحدات ويمكن استخدامها لمقارنة البيانات في المجموعات المختلفة. هذه القيم التي لا تعتمد على الوحدات نسميها بالقيم (أو الدرجات) المعيارية.

تعريف (2):

لتكن $x_1, x_2, \ldots, x_n$ عينه من البيانات حجمها n ومتوسطها $\bar{x}$ وانحرافها المعياري s. نعرف الدرجة (القيمة) المعيارية للمشاهدة x_i بالصيغة التالية:

$$z_i = \frac{x_i - \bar{x}}{s}; i = 1, 2, \text{K}, n \qquad (11)$$

أي أن الدرجات المعيارية للبيانات $x_1, x_2, \ldots, x_n$ هي:

$$z_1 = \frac{x_1 - \bar{x}}{s},\ z_2 = \frac{x_2 - \bar{x}}{s}, \ldots,\ z_n = \frac{x_n - \bar{x}}{s}$$

ملاحظة (6):

1. القيمة $z_i = \frac{x_i - \bar{x}}{s}$ هي الدرجة المعيارية للمشاهدة الأصلية x_i.
2. المشاهدة الأصلية للدرجة المعيارية z_i هي $x_i = \bar{x} + S\, z_i$.
3. الدرجات المعيارية هي قيم عديمة الوحدة ولذلك فإنها تستخدم للمقارنة بين المشاهدات المختلفة في المجموعات المختلفة للبيانات.
4. متوسط الدرجات المعيارية =0.
5. الانحراف المعياري للدرجات المعيارية يساوي =1.

مثال (14):

إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x} = 7$ وانحرافها المعياري s =5 فأوجد:

1. الدرجة المعيارية للقيمة $x = 9$.
2. القيمة الأصلية للدرجة المعيارية $z = 0.1$.

الحل:

1. الدرجة المعيارية للقيمة $x = 9$ هي:

$$z = \frac{x - \bar{x}}{s} = \frac{9-7}{5} = \frac{2}{5} = 0.4$$

2. القيمة الأصلية للدرجة المعيارية $z = 0.1$ هي:

$$x = \bar{x} + s\, z = 7 + 5 \times 0.1 = 7 + 0.5 = 7.5$$

مثال (15):

إذا كانت درجة أحد الطلاب في مقرر الإحصاء تساوي 82 ودرجته في مقرر الرياضيات تساوي 89، وإذا كان

متوسط درجات الطلاب في مقرر الإحصاء يساوي 75 بانحراف معياري يساوي 10 ومتوسط درجات الطلاب في مقرر الرياضيات يساوي 81 بانحراف معياري يساوي 16، ففي أي المقررين كان أداء الطالب أفضل؟

الحل:

نلخص إيجاد الدرجات المعيارية في الجدول التالي:

المقرر	المتوسط $\bar{x}$	الانحراف المعياري s	الدرجة x	الدرجة المعيارية $z=\frac{x-\bar{x}}{s}$
الإحصاء	75	10	82	$z=\frac{82-75}{10}=0.7$
الرياضيات	81	16	89	$z=\frac{89-81}{16}=0.5$

بما أن الدرجة المعيارية لمقرر الإحصاء 0.7 أكبر من الدرجة المعيارية لمقرر الرياضيات 0.5 فإن أداء الطالب في مقرر الإحصاء أفضل من أدائه في مقرر الرياضيات بالرغم من أن درجته في مقرر الإحصاء أقل من درجته في مقرر الرياضيات.