الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي

جامعة 8 ماي 1945 قالمة

كلّية العلوم الانسانية والاجتماعية

قسم علم االاجتماع

مطبوعة بيداغوجية:

الإحصاء الوصفي والاستدلالي1

مقدّمة لطلبة ماستر علم الاجتماع تنظيم وعمل

السداسي الأوّل

الدكتور يخلف سهيل أستاذ محاضر أ

السنة الجامعية: 2024-2025

الفهـــــــــــــرس

<u>فهرس الجداول</u>

فهرس الأشكال

<u>فهرس المنحنيات</u>

<u>فهرس التمارين</u>

مقدمة

علم الإحصاء هو من بين فروع علوم الرياضيات، إلا أننا نجده يتقاطع مع العديد والعديد من العلوم نظرا للحاجة الماسة إليه، لأي عملية بحثية أو تجربة عليمة، خاصة في الدراسات التطبيقية التي

تهدف إلى الوصول إلى نتائج دقيقة تتسم بالمصداقية والموضوعية.

فمنذ القديم اهتم الانسان بالعدد ، خاصة ما وجد من تقنيات وأساليب وبعض الطرق التي عرفت عند الفراعنة من إحصاء بعض الأراضي على محيط النيل ، أو احصاء أجزاء من المجتمع كالجيوش والعبيد وغير ذلك، وتطور علم الإحصاء أكثر بعد القرن 19 بفعل تطور العلوم الأخرى خاصة الرياضيات والفيزياء والاحتمالات حتى نصف القرن الماضي أي بعد الستينيات عرفت دفعا آخر باستعمال الإحصاء بواسطة برمجيات و ألات خاصة بالإحصاء أين سهلت من طرائق الإحصاء وطورتها.

وتشير كلمة الإحصاء إلى عمليات جمع بعض الحقائق أو الظواهر التي تشمل حالات أو مشاهدات في الواقع يمكن لها أن تكون بيانات كمية أو كيفية وأساليب استعمالها ومعالجتها وتلخيصها، كما يتيح لنا فرص وطرق دراسة واستخلاص بعض الاستنتاجات من دراسة عينة صغيرة من أفراد المجتمع ، يمكن تعميم نتائجها على المجتمع الكلي.

ويهدف هذا المقرر إلى تسهيل وتمكين الطالب في علم الاجتماع إلى طرائق جمع البيانات وتلخيصها وتبويبها وعرضها بشكل صحيح في جداول تكرارية أو في رسومات بيانية مع القدرة على وصف وتحليل هذه البيانات من خلال مقايس النزعة المركزية أو مقايس التشتت والشكل، ويحتوي هذه المطبوعة على ثلاثة محاور وثلاثة أعشر محاضرة مع الشرح من . خلال أمثلة ونماذج.

يحتوي المحور الأول على مدخل للإحصاء بالتعريف وعلاقته بالعلوم الأخرى ومختلف مراحل الطرق

الإحصائية، مع كيفية عرض البيانات الإحصائية والمحور الثاني هو مخصص لمقاييس النزعة المركزية كالمتوسط الحسابي والوسيط وشبيهات الوسيط والمنوال والمحور الثالث هو مقاييس التشتت كالمدى والانحراف المتوسط والتباين والانحراف المعياري مع مقايس الشكل كمقاييس الالتواء والتفرطح.

الفصل الأوّل:

مدخل

إلى الإحصاء الوصفي

I. المحاضرة الأولى: تعريف الإحصاء وعلاقته بالعلوم الاجتماعية

1) تعريف الإحصاء

الاحصاء هو ذلك العلم الذي يهتم بتجميع البيانات حول الظواهر وتلخيصها وصولا إلى تقديم استنتاجات أوسع من حدود المجتمع المدروس، وقد تطور هذا العلم بشكل أوسع بعد القرن 18 على يد علماء منهم لا بالاس Laplace(1749-1829) و قاوس Gauss (1777-1855) الذين طوروا التحليل الاحصائي وإنشاء قوانين الاحتمالات والخوض في دراسة العلاقات بين الحوادث والظواهر المختلفة ، ومنذ ذلك استعمال الاحصاء في مختلف العلوم والتخصصات مثل الديموغرافيا و البيولوجيا و علم الاجتماع وعلم النفس والاقتصاد وغيرها.

المدلول الأول ونعني به تصنيف وتعداد الأشياء كقولنا عدد المدارس في الجزائر وعدد السكان وعدد المواليد ... إلخ ويمكن التعبير عن هذه الأعداء بمقاييس وبأدوات وطرق عديدة المعني الثاني ونقصد به ذلك العلم ال يهتم بتعميم النتائج المستخلصة من العينات على المجتمعات والذي يعتمد أساسا على قوانين ونظريات الاحتمالات، وعلى ضوء هذين المدلولين فإن الاحصاء ينقسم إلى قسمين أو أكثر[1]

1-1 الإحصاء الوصفي : ويهتم هذا النوع من الاحصاء بعمليات تصنيف وجمع وتنظيم وتلخيص البيانات العددية الرقمية بدلالة بعض المقاييس لأغراض الوصف و المقارنة فعلى سبيل المثال لدينا دراسة حول ظاهرة التسرب المدرسي في ولاية غرداية، نقوم بحصر وتجميع كل الحوادث التي وقعت في الولاية ثم نلخصها في جداول أو رسومات تبين شدة هذه الظاهرة أو مستواه ونسبتها من مجموع التلاميذ مثلا أو غاية مقارنتها مع نفس النسب والمعدلات بالمقارنة التاريخية أي تطور الظاهرة تاريخيا في نفس المنطقة أو مقارنتها جغرافيا . مجموعات سكانية أخرى أو ولايات أخرى أو مع دول أخرى.

2-1 الإحصاء الاستدلالي: إن الاحصاء الحديث يهتم بالنظرية والمنهجية لاستخلاص النتائج التي تتجاوز مجموعة البيانات الخاصة التي تم فحصها أي العينة، وهو بذلك يوفر الجهد والوقت والموارد للوصول إلى معلومات

دقيقة عن المجتمع بطرق أسهل دون اللجوء إلى الحصر الشامل، فمثلا إذ أراد طالب على مستوى الماستر دراسة أسباب التسرب المدرسي في ولاية غرداية فقد يتعذر عليه الوصول إلى كل مفردات البحث نظرا لتوسع المجال وكبر حجم المجتمع المدروس فإنه بطرق منهجية يستطيع أن يدرس مجموعة مصغرة يسقط نتائجها على المجتمع.[2]

---

[1] سالم عيسى بدر، عماد غصاب عبابنة، مبادئ الإحصاء الوصفي والاستدلالي، دار المسيرة، الأردن 2007، ص 12

[2] المرجع نفسه

2) العلاقة بين العلوم الاجتماعية والإحصاء

قد يكون هذا السؤال تبادر إلى ذهن جميع الطلبة أول أيام التحاقهم بهذا التخصص وتسألوا علنيا أو ضمنيا لماذا ندرس الاحصاء في تخصص يهتم بالظواهر الاجتماعية: فقد اهتم الانسان منذ القديم بمشاكله الاجتماعية أو التفكير في بعض المظاهر التي تطرأ على محيطه إلا أن إدراكه و فهمه لتلك الأمور كانت محدودة أو ذو صبغة فلسلفية أو لاهوتية محضة وتطبع عليها الأحكام المسبقة والعموميات، وفي وقتنا الحاضر أصبحت هذه الطريقة تقليدية وعاجزة على تقديم معرفة دقيقة حول الإشكالات والظواهر التي يهتم بها الانسان فابتدع طرائق ومناهج أسست للمعرفة العلمية،

فعندما نتكلم على العلم أو المعرفة العلمية، فإننا نتحدث عن شروط ومعايير وضوابط تقيد معرفتنا أو إدراكنا أو تعاملنا مع فهم هذه الظواهر، فإلى جانب التراكمية والموضوعية والتنظيم وغيرها من شروط المعرفة العلمية فهنالك شرط الدقة والاستدلال والقياس والتجريب، وكل هذه العناصر تستدعي بالضرورة التمكن من مادة الإحصاء للتعامل مع البحث العلمي. فعندما نقول كلمة علم فهنالك بالضرورة موضوع ومنهج وإحصاء إلى جانبه، خاصة إلى ميل الدراسات الحديثة إلى المناهج الكمية أكثر.[3]

وغالبا ما نجد مادة الإحصاء كمادة منهجية مدرسة في مختلف التخصصات التي بحثية علمية على اختلافها وتنوعها، فالباحث العلمي يحاول أن يلاحظ ظاهرته المدرسة عن طريق واقعها في الميدان وعن طريق دراسة مجموعة من الوحدات التي تشكل مجتمع الدراسة أو عن طريق عينة يختارها ليقوم عليها بالتجريب، أو صدق النظرية أو الفرضية التي وضعها من خلال قياس العلاقة بين المتغيرات والظواهر، بشكل رياضي وإحصائي فعلم الاحصاء يوفر لنا الطرق و الاساليب والمناهج الكفيلة بذلك.

فإذا أردنا مثلا الكلام عن ظاهرة ما في المجتمع ولنأخذ مثلا ظاهرة الطلاق فالإدراك البسيط العامي يحكم على هذه الظاهرة على حسب ما تخبره أحاسيسه وتجارب محيطه حول الموضوع، وغالبا ما تكون سطحية معممة.[4]

وحكمية، وإذا أخذنها بصفة علمية فإننا على الاقل سنحدد حجم ونسب هذه الظاهرة وفق معدلات ونقارنها التاريخ أو مع مجموعات سكانية أخرى مثلا، فهذه الطريقة الأخيرة ستقدم لنا معرفة أوضح وإدراك أفضل عن الظاهرة من المعرفة الأولى العامية.

كل هذا وبعد تطور العلوم الاجتماعية وانشقاقها عن الفلسفة جعلتها تستعين بعلم الاحصاء والأساليب والأدوات التي يوفرها هذا العلم فتطورت بسبب استخدامها وطورت معها المناهج الكمية التي تعتمد بشكل كبير

---

[3] عبد القادر حليمي مدخل إلى الإحصاء، ديوان المطبوعات الجامعية، ط 6 الجزائر، 2009 ص 17

[4] سالم عيسى بدر ، مرجع سابق ، ص 13-14، خالد حامد منهج البحث العلمي، دار ريحانة، الجزائر 2003 ص 16

على الاحصاء فانتقلت هذه العلوم من المعرفة الوصفية الانشائية إلى تكميم هذه الظواهر ومعرفتها وفهم علاقاتها بشكل رياضي ورقمي جد دقيق.

وفي الأخير فقد نقول أن الاحصاء يضفى على الموضوع صفة العلمية فالمعطيات والمعلومات الأكثر التي جمعت بعناية وتم تحليلها بدقة عن طريق البيانات الإحصائية والأساليب الاحصائية، كون العلوم الاجتماعية تعتمد على الأساليب الاحصائية كأسس قوية ودعامة موضوعية لتعزيز الفهم وبناء النظرية الاجتماعية وأغلب العلوم الاجتماعية والمعاهد والجامعات الحديثة تميل إلى الاعتماد والتركيز على البحوث الكمية أكثر من الكيفية، التي بدورها مرتبطة ارتباط وثيق بالإحصاء ومستحيلة أن تكون بدونها. 1

I. المحاضرة الثانية: خطوات الطريقة الإحصائية

لاستخدام الإحصاء واستعماله السليم في البحوث الاجتماعية وقبل الحصول على النتائج والبيانات الإحصائية وتحليلها وتوظيفها نشير إلى أنه هنالك مجموع من الضوابط والشروط المنهجية نتقيد بها وبعض الإجراءات والخطوات يجب تتبعها لكي نتحصل على بيانات سليمة تخدم البحث وتكون ذو مصداقية علمية.

ويمكن تلخيص هذه الخطوات والضوابط بشكل مختصر في هذه النقاط:

أ) اختيار الموضوع وتحديد أهداف الدراسة وإشكالية الدراسة بشكل دقيق حتى يتضح لنا ماذا نريد دراسته بالضبط وعن ماهية البيانات المطلوبة مع مراعات قابلية الانجاز من الامكانيات المادية والفنية المتوفرة لدى صاحب المشروع الوقت الزمني، الوصول إلى مفردات البحث، توفر المعطيات...إلخ).[5]

ب) صياغة فرضيات البحث التي تحمل في غالبيتها متغيرات علمية أو منهجية (متغيرات مستقلة ومتغير تابع) والتي تحمل علاقة سببية بين السبب وهو (المتغير المستقل) والنتيجة أو الظاهرة (المتغير التابع) ج) على ضوء تحديد الموضوع وفرضيات الدراسة يتعين علينا تحديد مفاهيمها بشكل إجرائي وبشكل نتمكن من إخضاع هذه المفاهيم للدراسة والقياس، بحيث يتعين تحديد أبعاد المفهوم ومتغيرات الأبعاد ثم مؤشرات المتغيرات، ويكمن هنا تمكن وذكاء وفطنة الباحث في قدرته على تحويل مفاهيم النظرية إلى متغيرات ومؤشرات قابلة للقياس والتجريب بإحدى أدوات البحث المعروفة

د) تحديد مجتمع الدراسة أي الإطار الذي يحتوي على جميع المفردات الخاضعة للدراسة الاحصائية، وعينة الدراسة وهي تلك المجموعة الصغيرة من الكل التي يتعامل معها الباحث، ووحدة الدراسة (الفرد) أي الوحدة الأساسية في البحث والخاضعة للفحص، فإذا ما أراد باحث دراسة على مستويات التحصيل الدراسي لدى تلاميذ الابتدائي في ولاية غرداية فإن مجتمع الدراسة الابتدائية الموجودة على تراب ولاية غرداية، وفي هذه الحالة لا يمكنه دراسة كل هذا المجتمع لكبر حجم المجتمع وتوسعها في الرقعة الجغرافية فإنه يأخد جزء من الكل وهي مجموعة يتم اختيارها بطرق منهجية تسمى العينة يخضعها للدراسة، أما وحدة الدراسة فهي التلميذ.[6]

هـ) تحديد مصادر البيانات والمعطيات المراد جمعها للظاهرة المدروسة. تنقسم إلى صنفين *بيانات مباشرة*: يتم جمعها من طرف الباحث بنفسه من عن طريق مفردات البحث المباشرة من الميدان نفسه أو بواسطة أشخاص يتم إعدادهم وتدريبهم للقيام بهذه المهمة بإحدى وسائل جمع المعطيات (المقابلة أو الاستمارة).

---

[5] مصطفى خلف عبد الجواد، الاحصاء الاجتماعي المبادئ والتطبيقات، دار المسيرة، الأردن 2013، ص 18-19

[6] أنظر كتاب موريس أنجرس منهجية البحث العلمي في العلوم الانسانية، تر مجموعة من الاساتذة، دار القصبة الجزائر 2004، ص 127

البيانات الغير مباشرة : والتي يتحصل عليها الباحث عن طريق سجلات مثل سجلات الولادات والوفيات في الحالة المدنية أو وثائق تاريخية أو جداول إحصائية جاهزة موجودة في نشريات خاصة لذلك مثل النشريات التي يقدمها الديوان الوطني للإحصائيات.

و) تحديد أسلوب جمع البيانات أي تحديد طريقة للوصول إلى مفردات البحث في حالة البيانات المباشرة عن طريق إما الحصر الشامل إن توفرت الإمكانية لذلك أو عن طريق أسلوب العينة وتختار بإحدى الطرق الإحصائية المناسبة على ضوء الغاية من الدراسة و طبيعة الموضوع و توفر قاعدة البيانات ، ثم تحديد المجال الزمني و الوقت الذي يتم فيه النزول إلى الميدان وتجميع الملاحظات أو الاستمارات.

ز) ضبط أداة جمع المعطيات بشكل المنهجي والتأكد من صحتها ومدى شموليتها وخدمتها للموضوع عن طريق دراسة استباقية أو عن طريق تصحيح وتصويب من مختصين

ح) النزول إلى الميدان وجمع هذه الملاحظات

ط) جمع هذه الاستمارات وفرزها من ناحية شمولية الاجابة و جديتها والأخطاء والشطب، ثم ترقيمها وتبويبها وترميزها ثم تفريغها في المجال المخصص لها المصفوفة

ي) جدولة هذه البيانات وعرضها بالطريقة المناسبة عن طريق الجداول الإحصائية أو الرسومات البيانية

ك) دراسة وتحليل هذه البيانات وقراءتها والتعليق عليها باستخدام الطرق الإحصائية كمقياس النزعة المركزية ومقاييس التشتت و الإنحدار والارتباط أو السلاسل الزمنية والأرقام الهندسية ... إلخ

ل) اختبار الفرضيات واستخلاص نتائج الدراسة وتحليلها عن طريق القراءة الإحصائية أو المعاملات الرياضية أو الاختبارات الاحتمالية، ثم تقديم المقترحات.[7]

تمرين 1: استخراج من المكتبة دراسة علمية أكاديمية منجزة منشورة

سواء في كتاب أو مذكرة ماجستير أو دكتورة ونقد هذه الخطوات الطريقة الاحصائية التي استعمالها الباحث بتبيان كل العناصر المشكلة لهذه الخطوات من موضوع الدراسة

تمرين 2: جد خطوات الطريقة الإحصائية

محاولة التفكير في موضوع بحث من خلال تحديد عنوانه وكل خطوات الطريقة الإحصائية كمشروع عمل يمكن إنجازه بشكل جماعي مع الفوج الذي ينتمي إليها الطالب، أو مشروع

يمكن أن يستمر فيه الطالب في السنة الثانية لإنجاز مذكرته للتخرج.

[7] سالم عيسي بدر ، مرجع سابق ، ص 19-20

II. المحاضرة الثالثة : مفاهيم احصائية

يفترض إدراك بعض الأمور الأساسية وبعض المفاهيم في الإحصاء ليسهل علينا تتبع مراحل المقبلة في الإحصاء وهنا سنعرض أهم المفاهيم ويستحسن على الطالب تعلمها من لغتين وأكثر.

1) المجتمع الإحصائي

أو مجتمع الدراسة وهو مفهوم احصائي يقصد به جميع العناصر والمفردات

التي تخص الدراسة، بحيث يمثل هذه العناصر كائنات أو أشياء، كما يمكن أن يكون المجتمع خاضع كله للدراسة وهنا تدعي بطريقة الحصر الشامل كالتعدادات وهي حصر شامل لجميع مفردات المجتمع الإحصائي، وإن تعذر على الباحث ذلك فإنه يلجأ إلى جزء من هذا المجتمع ويطلق عليه العينة

2) العينة

في غالب الحالات يتعذر على الباحث إقامة دراسته على المجتمع الاحصائي كليتا لإنها

مكلفة ومضيعة للوقت في الجمع وتحليل بياناتها ولهذا فإنه يأخذ جزء من الكل يقيم عليها دراسته، كما يمكنه إسقاطها على المجتمع الأصلي وتنقسم أنواع العينات إلى صنفين عشوائية وغير عشوائية ولها أنواع.[8]

3) المعطيات الإحصائية

تكون المعطيات الإحصائية في الأساس المعلومات الأساسية للإحصاء

وكل العلم والطرق والقواعد هي موجودة أساسا لهذه المعطيات التي تأخذ أساسا من الموضوع ومن الملاحظات أو من الاستمارة التي تم جمع بها هذه المعطيات وتنقسم متغيرات هذه المعطيات إلى أنواع منها :

أ) المعطيات الفعلية les données réelles : وهي التي يعبر فيها عن الظاهرة بأرقام فعلية من أحد القياسات الموجودة كالوزن والطول و غير ذلك.. كأربعة أشخاص لكل وزنه وطوله

ب) المعطيات الكاردينالية : Les données cardinales: وهي تلك المعطيات التي لا يمكن وضع لها نظام بين الظواهر وبالتالي فالقياس يكون للظاهرة فقط : مثل المعطيات التي تجمع . من ثلاثة مناطق جغرافية 1 و 2 و 3 فقد يكون جزء من مجموعة 2 تنتمي إلى المجموعة1 وبالتالي لا يمكن فصلها، ولهذا فتكون المجموعة الأولى المشكلة هي من مجموعة (1) و (2 و المجموع الثانية من المجموعة (3).

---

[8] موساوي عبد النور، بركان يوسف، الإحصاء ،1 دار العلوم، الجزائر 2009، ص 9

ت) المعطيات الرتبية *Les données ordinales*: وهي المعطيات المرتبية تصاعديا أو تنازليا أو حسب معيار معين دون الأخذ بعين الاعتبار الفرق بين هذه الوحدات، كقولنا الكلية أ هي أكبر عدد من الكلية ب والكلية ب أكبر من ج في حجمه وعددها.

ث) المتغيرات الإسمية *les données nominales* وهي التي تصنف البيانات على حسب مرتبتهم وميزة مواتية لهم بحيث كل مفردة تدخل في صنف واحد دون الأخر مثل معطيات الجنس واللون .

4) المتغيرات الإحصائية

البيانات الإحصائية تأخذ في العادة ميزة وخاصية إما وصفية أو كمية.

أ) **المتغير الإحصائي الكيفي** : qualitative Data /variable qualitative وتسمى بالمتغيرات الوصفية كذلك وهي عبارة عن أنواع من البيانات تصف الظاهرة المدروسة بشكل غير رقمي، فهي غير قابلة للقياس بل تأخذ شكل المعطيات الأسمية وهي صفات ينقسم وفقها البيانات إلى مجموعات مثل الحالة الاجتماعية (متزوج ، أرمل، مطلق، عازب) الجنس (ذكر، أنثى) ...إلخ.[9]

ب) **المتغير الإحصائي الكمي** quantitative data /variable quantitative وهي البيانات التي تعبر عن الظاهرة بشكل رقم وحددت يقياس معين أو قيمة مأخوذة من وحدة قياس معينة وتنقسم إلى:

- متغير كمي منفصل discrete: وهي عبارة عن بيانات يتم التعبير عنها على شكل أعداد أو أرقام صحيحة ولا تحتوى على كسور مثل : عدد المواليد للمرأة أو أفراد الأسرة أو أعداد الطلاب إلى غير ذلك فهذه البيانات تأخذ أرقام صحيحة ولايمكن أن نقول عدد مواليد المرأة 4 أولاد ونصف أو عدد أفراد الأسرة و ربع أو عدد افراد الاسرة 7 و ربع او غير ذلك فهذا غير ممكن.

- متغير كمي متصل : عندما تكون البيانات متوزعة بفترات متباعدة بينها يتم تجميعها إلى فترات أو ما يطلق عليه الفئات مثل العمر وتحديد من [20-25] ومن [25-30] إلخ أو أجرة العامل من [150-200] و [200-250] ... إلخ فالمعطيات تتوزع هنا على حسب هذه الفترات من و إلى ويمكن تصور عدد كبير أو صغير داخل هذه المجالات.[10]

5) الفئات

في حالة المتغيرات الكمية المنفصلة الكبيرة الحجم والمتسعة في مجالاتها يستحسن عند عرضها أن تكون في مجالات فئات لأن عرضها بشكل منقطع يصعب لدى القارئ ولا عند الباحث تتبع جميع الملاحظات ولا قراءتها أو

[9] عبد القادر حلمي ، مرجع سابق ، ص 28-29

[10] سالم عيسى بدر ، مرجع سابق ، ص 31

استخلاص منها نتائج، ولهذا يتعين عليه تجميعها في فئات متساوية البعض يفضل أن تكون من مضاعفات 10 أو 5 والبعض يتبع الخطوات الإحصائية الموجودة والتي تتمثل في:

أ) استخراج المدى العام E وهو الفرق ما بين الحد الأعلى والحد الأدنى في البيانات ويمكن تلخيصها وفق القاعدة التالية E=Xmax-Xmin

ب) استخراج عدد الفئات k: ويتم ذلك وفق قاعدتين إما بقاعدة يول أو قاعدة ستورجس

- قاعدة يول Yule's : وتستخدم هذه المعادلة عندما يكون العينة أقل .حجم من 1000 وحدة وهنا نستخرجها من القاعدة 2,5 ضرب جدر أربع لعدد القيم ويمكن تلخيصها في القاعدة التالية :

$$K= 2.5 N$$

عن قاعدة ستورجس 1formule Sturges وتستخدم في حالة ما يزيد عدد القيم 1000 وحدة وهي تساوي 1 زائد 2.32 ضرب لوغاريتم عشري لـ عدد القيم وتلخيصها وفق القاعدة التالية: K= 1+2.32 log N

ج) تحديد طول الفئات أو مدى الفئة : وهو استخراج طول الفئة بقسمة المدى العام على عدد الفئات ويستحسن تقريب قيمة C إلى قيمة صحيحة، ويمكن تلخيص القاعدة على الشكل التالي: [11]

$$K=1+ 3,32 \log N$$

Xmax-Xmin E C

وبعد هذه الخطوات تحديد طول الفئة وعدد الفئات نقوم بتشكيل فئات الجدول بحيث نقوم بإضافة للحد الأول من البيانات مجال طول الفئات ويتشكل لنا الفئة الأولى وبعد تحديد الفئات الأولى يتكون لدينا حد أدنى للفئة وحد أعلى للفئة، وهذا الحد الأعلى للفئة الأولى هو الحد الأدنى للفئة الأولى الثانية ، ثم نضيف لها مجال أو طول الفئة ويتشكل الحد الأعلى للفئة الثانية وهكذا دواليك إلى غاية وصول الحد الأعلى للعينة.

6) التكرارات المطلقة: *frequencies absolutes*

بعد أن يجمع الباحث معطياته ينتقل إلى عملية عرضها بشكل مبسط فيأخذ في تصنيف هذه البيانات إلى فئات أو مجموعات متجانسة في خاصية ثم يقوم بتعداد وإظهار عدد المرات أو الحالات التي تدخل في كل فئة أو

[11] مصطفى خلف عبد الجواد ، مرجع سابق، ص 52

مجموعة وعدد هذه المرات أو الحالات هو ما يطلق علها في الإحصاء بالتكرارات المطلقة ومجموع كل هذه التكرارات هو حجم العينة.[12]

7) التكرارات النسبية *Relative frequencies*

تعبر عن منزلة أو نسبة أفراد الفيئة على مجموع أفراد العينة على أن يكون مجموعها يساوي الواحد وهي تساوي fi/N والتكرار النسبي له أهمية كبيرة خاصة أنه يسهل لنا القراءة وفهم التوزيع بشكل أفضل من التكرار المطلق خاصة إذا ما إذا أردنا المقارنة ووصف مجموعتين مختلفتين.[13]

8) التكرارات المتجمعة

نحصل علها بجمع التكرارات تصاعديا أو تنازليا بالتكرار المطلق أو التكرار النسبي المقابلة لكل فئة زائد تكرار الفئة الموالية إلى غاية وصول حد العينة عند التكرار المطلق أو الواحد عند التكرار النسبي وبهذا نجد لدينا نوعين من التكرارات المتجمعة

أ) التكرار المتجمع الصاعد fi نقوم بتجميع التكرارات من أول فئة زائد تكرار المطلق للفئة الموالية وعمليا الفئة الأولى هي تكرار المطلق للفئة الأولى والتكرار المتجمع الصاعد الثاني هو عبارة عن تكرار لمتجمع الأول + تكرار المطلق للفئة الثانية، ثم هكذا دوليك إلى غاية الوصول إلى حد العينة

ب) التكرار المتجمع النازل Fi : نقوم بوضع مجموع أفراد العينة في الفئة الأولى ثم الفئة الثانية تساوي تكرار المتجمع النازل ناقص التكرار المطلق للفئة الأولى ، الفئة الثالثة تساوي تكرار المتجمع النازل للفئة السابقة ناقص تكرارها المطلق وهكذا دواليك حتى وصول التكرار الأخير الذي يساوي نفس التكرار المطلق 1

ونفس الطريقة يمكن اتباعها مع التكرار النسبى المتجمع ولكن هنا الحد الأعلى الذي نصل إليه يساوي دائما واحد.[14]

---

[12] موساوي عبد النور ، مرجع سابق ص 14

[13] عبد القادر حلمي ، مرجع سابق ص 30

[14] موساوي عبد النور ، مرجع سابق ص 12

تمرين 3: جد مجموع من العمال الموزعين تصاعديا وتنازليا حسب أعمارهم في الجدول أسفله.

جدول 1: توزيع العمال حسب العمر

ع1	ع2	ع3	ع4	ع5	ع6	ع7	ع8	ع9	ع10
1	12	10	5	7	8	6	4	21	45
2	22	12	4	6	8	9	1	12	14
3	2	13	12	15	11	4	5	9	10
4	12	15	10	10	16	17	15	12	11

المطلوب :

- تشكيل فئات لهذه البيانات
- جدولة هذه المعطيات واستخراج التكرار المطلقة
- استخراج التكرارات النسبية في عمود آخر
- وضع عمود للتكرارات المتجمع الصاعد وآخر للتكرار المتجمع النازل
- ما نوع المتغير وما نوع المعطيات

II. المحاضرة الرابعة عرض البيانات الإحصائية عن طريق الجداول

بعدما حددنا بعض المفاهيم الإحصائية وضبطناها في المحور السابق، يبقى علينا عرض هذه البيانات التي جمعت من الميدان عن طريق الملاحظة أو الاستمارة لاستغلالها و استخراج الملاحظات حول هذه المعطيات والقيام بتحليلها أو تفسيرها على حسب الغاية منها ولهذا لدينا العديد من الطرق يمكن تقسيمها إلى مجموعتين وهو العرض عن طريق الجدول الإحصائي أو عن طريق الرسومات البيانية

وهنا يمكن أن نجد عدة أصناف من الجداول وسنركز على صنفين فقطين لشيوع استعمالهما والأكثر تداولا

1) الجداول التكرارية البسيطة

وهي الجداول التي تحمل غالبا متغير واحد وهي موجودة أساسا لغية الوصف وعرض نتائج وتجميع التكرارات ، ولأنشائها أولا علينا بتحديد الفئات أو تبويب المعطيات إن كانت غير مبوبة،، ثم إنشاء جدول ذو ثلاثة أعمدة في العمود الأول يكون فيه فئات المتغير والفئة الثانية للجدولة حيث نسجل فيه على شكل خطوط أو مربعات خماسية امام كل متغير تتكرر فيه الملاحظة، أما العمود الأخير فيتم تجميع تلك الخطوط في قيم تعبر عن تكرار تلك الفئة المقابلة لها على أن يكون مجموع قيم هذه التكرارات كلها يساوي حجم العينة كما هو موضح في الجدول الموالي لتوزيع 40 فرد على حسب نوع الجنس

جدول 2: نموذج لتوزيع المبحوثين على حسب نوع الجنس

اسم المتغير (الجنس)	العلامات	التكرارات fi
ذكر		15
أنثى		25
المجموع	-	40

وقد يستعين الباحث هنا أكثر في عرض هذا الجدول بإضمار تلك العلامات وإضافة التكرارات النسبية أو التكرار النسبي المئوية بضرب التكرارات النسبية في 100 وهذا يسهل غاية الوصف مادام ان هذه الجداول هدفها الوصف فقراءة النسبة المئوية أفضل وأصح من قراءة التكرار المطلق مثل ما هو مبين في الجدول الموالي :

جدول 3: توزيع المبحوثين على حسب الجنس

اسم المتغير (الجنس)	التكرارات fi	التكرار النسبي	تكرار النسبي المئوي
ذكر	15	0,375	%37,5
أنثى	25	0,625	%62,5
المجموع	40	1	%100

وهنا يتضح من الجدول توزيع أفضل من ناحية عدد الافراد حسب نوع الجنس وغالبا إذا تمت قراءة هذه المعطيات فإنها تقرأ التكرارات النسبية ونبدأ بأكبرها فأدناها مثل الجدول الذي يوضح توزيع أفراد العينة على

حسب الجنس بحيث يشكل جنس الإناث الأغلبية بـ 62,5% من أفراد العينة و 37,5 % من الذكور

2) الجداول التكرارية المركبة

يقصد بها تلك الجداول التكرارية للتوزيع ذو المتغيرين أو أكثر فالجداول التي تحمل متغيرين هي جداول مزدوجة والتي تحمل أكثر من متغيرين هي الرائزة وتسمى هذه الجداول المركبة أيضا بالجداول التفسيرية أو الجداول التحليلية، فالغاية من إدراج متغيرين في الجدول الواحد أحدهما مستقل والآخر تابع هو مقارنة تغير التوزيع وتأثيرها ما بين المتغيرين وفي الغالب يكون المتغير التابع وهو الظاهرة في الأعمدة والمتغير المستقل في الجانب على مستوى الخطوط ويصح العكس شريطة أن يكون التنسيب دائما في اتجاه المتغير المستقل.

- توجد عدة طرائق لقراءة الجداول واستخراج وتحليل النتائج أهمها والأكثر تداولات في علم الاجتماع هي تتبع الملاحظة عن طريق 4 خطوات أساسية يمكن أن تكشف لنا العلاقة ما بين المتغيرات الموجودة في الجدول، أولها وهي قراءة الاتجاه العام للظاهرة وهي قراءة النسب الموجودة في مجاميع الأعمدة المشكلة لاتجاه الظاهرة المدروسة والتركيز على أهم النسب ، القراءة الثانية يمكن استنتاجها من القراءة الوصفية التي نجدها غالبا في اتجاه المتغير المستقل أي في اتجاه التنسيب، والقراءة الثالثة و هي أهمها التي تكشف لنا العلاقة قراءة المقارنة ما بين النسب على مستوى الأعمدة بتأثير المتغير المستقل، أما القراءة الأخيرة فلا تكون إلا بعد القراءات السابقة وهي حصر شامل لاتجاهات وتطور النسب سواء من اليمين إلى اليسار أو العكس ومقارنتها مع الخطوط أو الأعمدة المقابلة لها ويمكن لنا عبرها استنتاج نوع العلاقة الموجودة ما بين المتغيرة أهي طردية أم عكسية .

في الأخير ندعم ذلك بمقايس واختبارات أخرى تأكد لنا العلاقة ما بين المتغيرين أو تنفيها سوف نتطرق لها في فصول السداسي الثاني، ثم نضفي عليها بالقراءة السوسيولوجية والتحليل النظري. لنأخذ المثال في الجدول الثالث: (أسفله)

جدول 4: توزيع أفراد العينة على حسب الجنس وعلاقة بسلوك الادخار

سلوك الإدخار / الجنس	يدخر	لا يدخر	المجموع
ذكر			
أنثى			
المجموع			

إذا ما لاحظنا في هذا الجدول فإننا نجد هنالك تقاطع بين متغيرين بين السلوك الاقتصادي وهو الإذخار و جنس المبحوثين ونريد من خلاله تتبع من هم الأشخاص الأكثر ادخار في العينة وأين يتمركز هذا السلوك.

ولندقق الملاحظة أكثر قسمنا الجدول إلى تكرارات مطلقة التي رمز لها بحرف ك وتكرار نسبي في الخانات الملونة بالأزرق ورمزت لها بالرمز النسبي المئوي في نفس المتغير وتم التنسيب نحو المتغير المستقل وهو جنس المبحوث أي على مستوى الخطوط مثال على ذلك تم تقسيم تكرار الخانة الأولى 45 على المجموع الموجود في أقصى اليسار وهو مجموع الخطوط 70 100 يعطي لنا النسبة المئوية للأشخاص الذكور الذين يدخرون وهو 64,28% وهكذا دواليك لكل خانات الجدول مع احترام اتجاه التنسيب، فالقراءة الإحصائية دائما تكون بالتكرار النسبي وليس بالتكرار المطلق وهنالك فرق سوف نلاحظه في الجدول الذي أمامنا.

أول قراءة هي قراءة الاتجاه العام وهي النسب الموجودة في مجموع الأعمدة أو مجموع المتغير التابع والذي يمثل الموضوع بحيث يشير لنا إلى تغير الظاهر المدروسة بشكلها الوصفي ونبدأ بالتكرار الأعلى ثم الأدنى، وإذا ما لاحظنا في الجدول فنقول أن الاتجاه العام لسلوك الادخار في العينة يشير إلى أن 51,53 % من أفراد العينة لديهم سلوك الادخار في حين أن 48,64% لا يدخرون قسط من دخلهم.

توجد قراءة وصفية أيضا في الجدول إذا ما أردنا استعمالها أو استغلالها في التحليل وهي هذا القراءة في نفس اتجاه التنسيب فالمثال الذي لدينا يمكن ملاحظة هذا إذا ما أردنا ملاحظة سلوك الادخار لدى جنس

معين، والقراءة الأهم وهي قراءة المقارنة وهي التي تحقق لنا هدف هذا الجدول وهو مقارنة التغير السلوك المدروس الإذخار ما بين الجنسين وهذه القراءة ترتكز على المقارنة عكس اتجاه التنسيب أو في اتجاه العمود لمقارنة تغير الظاهرة بمعية المتغير المستقل بحيث نلاحظ في الجدول أن الذكور أكثر إذخارا بنسبة 64,28% من نسبة الذكور في حين أن الإناث اللائي يدخرن لا يمثلن فقط 34,47% من مجموعهن وهذا ما نجد أن هذا السلوك لدى الذكور أكثر على الرغم أن التكرار المطلق لا يوحي بذلك 50 أنثى مقابل 45 ذكر ولكن هذا السلوك

يجب مقارنته بمعية وجوده في المجتمع هذه المقارنة تكشف لنا هنا كيف تتغير الظاهرة في المجتمع بمعية متغيرات أخرى مستقلة.

القراءة الأخيرة وهي قراءة الاستنتاج ماذا يمكن أن تستنتج من الجدول هل العلاقة طردية أم عكسية وهذا بتتبع تطور النسب في اتجاهها مقارنة مع تطورها في الخطوط أو الأعمدة المقابلة لها ثم كيف تأثر وأين يؤثر المستقل في التابع حتى يمكن لنا تفسير هذه العلاقة الموجودة في الفرضية.

III. المحاضرة الخامسة : عرض البيانات عن طريق الرسومات البيانية

في الكثير من الحالات الجداول التكرارية قد تكون عاجزة عن تقديم لنا ملاحظات أو يصعب تتبعها لكثرة التكرارات أو عدم وضوحها وهنا نلجأ لعرضها في شكل رسم رسم هندسي، ، هذه الرسومات الهندسية تسهل لنا الملاحظة بالعين المجردة وبطريقة أسهل، حتى أنها أقرب في العرض للقارء من الجدول، ويمكن حصر بعض أنواع الرسوم في الرسومات الديا غرامات les Diagrammes وهي تشمل الدوائر و المستطيلات والمقاطع وغيرها من الأشكال الهندسية، أما الرسومات البيانية des Graphiques وتشمل كل من الرسومات الكارتيزية les graphiques cartesiens والمدرجات التكرارية Histogrammes والمضلعات التكرارية Polygones والمنحنيات التكرارية les Courbes وغيرها ...[15]

طرائق عرض الرسومات البيانية هي متعددة ولكنها تخضع لشروط وضوابط ليكون الرسم صحيح وواضح ومن بين الشروط أن يكون لها عنوان واضح وله دلالة فيه نوع الرسم ونوع المتغير، ومصدر الرسم إن اقتضت الضرورة ويكون دائما في أسفل الرسم مباشرة ، مع وجود سلم ومفتاح للرسم يمكن للقارئ تتبع الظاهرة عن الرسم، أما نوع الرسم فهو يخضع أساسا إلى نوع المتغير الإحصائي ومنه يمكن تقسيم أنواع الرسوم إلى

1) الرسومات البيانية في حالة المتغير الكيفي

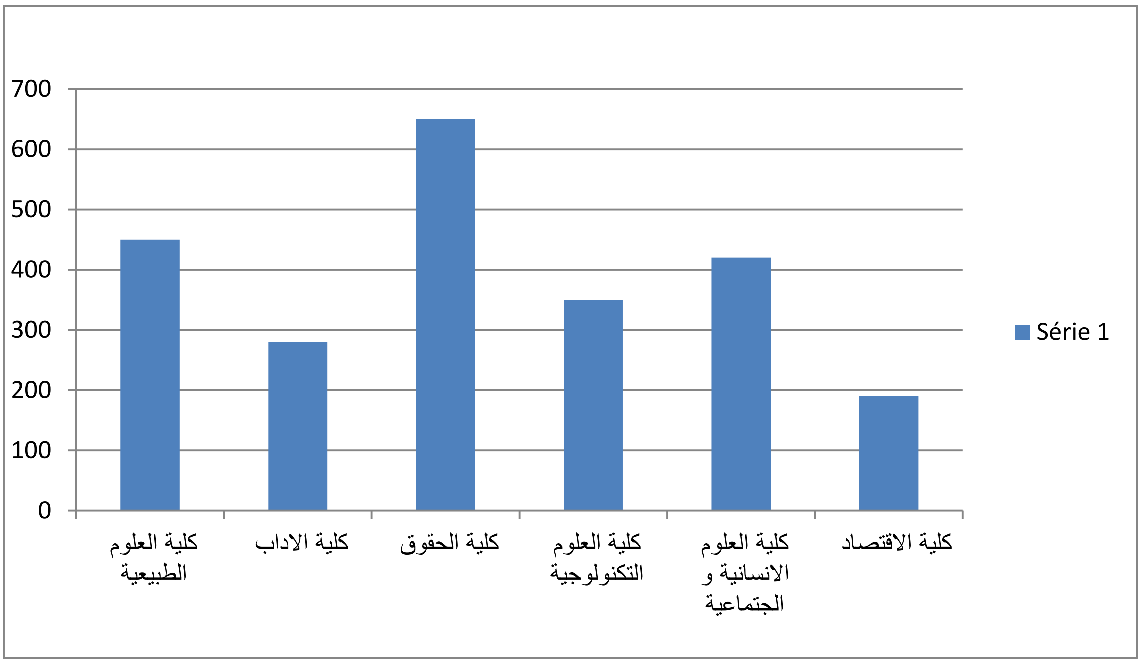
أ) **الأعمدة البيانية** : ينتمي هذا الرسم إلى الرسومات الكارتيزية بحيث يرسم على محوارين متعامدين متجانسين م س/ م ع يتقاطعان في النقطة م وتعرف بنقطة أصل الإحداثيات أو (ox / oy) بالحروف اللاتينية، في العمود الأفقي يمثل نوع المتغير الذي يكتب داخل المستطيل أو أسفله أو في مفتاح خاص به، أما المحور م ع العمودي أو الرأسي فيمثل فيه توزيع التكرار بطريقة تتناسب وحجم التكرار بحيث يمكن تقسيمها إلى وحدات كل وحدة تمثل 10 تكرارات أو 100 تكرارات، ونشير إلى ذلك في سلم الرسم، ثم نقوم برسم مستطيلات أفقية طولها يساوي عدد التكرار المقابل له 1 مثل ما هو موضح في المثال التالي:

جدول 5: توزيع الطلبة على حسب الكلية التي ينتمون لها

الكلية	كلية العلوم الطبيعية	كلية الأدب	كلية الحقوق	كلية ع التكنولوجية	كلية ع إنسانية والاجتماعية	كلية الاقتصاد
عدد الطلبة	**450**	**280**	**650**	**350**	**420**	**190**

[15] سالم عيسى بدر ، مرجع سابق ص 39

الشكل 1: أعمدة بيانية لتوزيع اعداد الطلبة عبى حسب الكليات



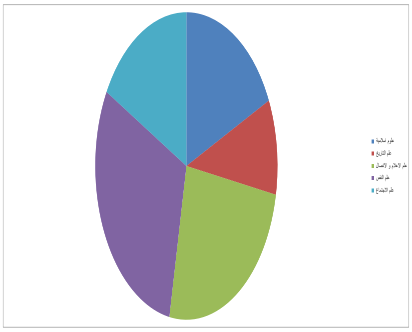
ب) **الدائرة النسبية** : ويتم ذلك بتقطيع الدائرة إلى أجزاء وقطع كل جزء يتناسب مع المعطيات الموجودة في الجدول ولنقوم بذلك يجب تحويل التكرارات المطلقة إلى تكرارات نسبية ثم نضرب التكرار النسبي في 360 درجة وتمثل مجموع زوايا الدائرة ويمكن ايضاح ذلك في الجدول الموالي:[16]

جدول 6: توزيع الطلبة على حسب الأقسام التابعة لها

القسم	التكرار	التكرار النسبي	تكرارنسبي ضرب 360
علم الاجتماع	50	0,17	60
علم النفس	90	0,30	108
علم الإعلام والاتصال	75	0,25	90
علم التاريخ	30	0,1	36
علوم اسلامية	55	0,18	66

[16] سالم عيسى بدر ، مرجع سابق، ص 41

الشكل 2: دائرة نسبية تمثل توزيع الطلبة على حسب الأقسام



علوم اسلامية - علم التاريخ - علم الإعلام والاتصال - علم النفس - علم الاجتماع -

2) الرسومات البيانية في حالة المتغيرات الكمية المتصلة

يتم عرضها عن طريق عدة رسومات منها المدرج التكراري والمضلع التكراري.

أ) **المدرج التكراري**: وهو نوع من الرسومات التي تنتمي إلى الرسومات الكارتيزية، بحيث يرسم محورين متعامدين متجانسين م س / م ع في محور م س قيم المتغير و في المحور م ع سلم للتكرارات ثم نمثل لأعمدة بيانية متلاصقة قاعدتها طول الفئة وارتفاعها يناضر التكرار المقابل له كما يوضح الجدول الموالي :[17]

---

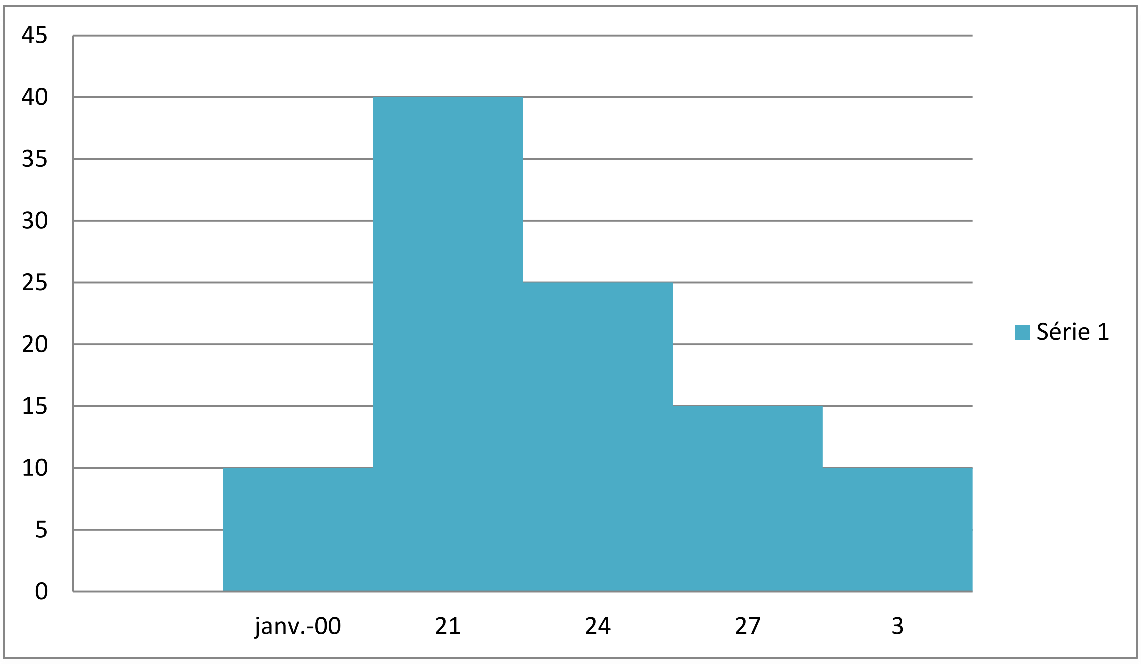
[17] موساوي عبد النور ، مرجع سابق ، ص 20

جدول رقم 6 : يمثل توزيع الطلبة حسب العمر

الفئات	التكرار
21 - 18	
24 - 21	
27 - 24	
30 - 27	
33 - 30	

ب) **المضلع التكراري**: يتم عرض البيانات المتصلة بالمضلع التكراري في محورين متعامدين متجانسين مثل الرسومات السابقة ثم نشكل احداثيات ثنائيتها (مركز الفئة ، التكرار المقابل له) مع وجود فئتين وهميتين قبل الفئة الأولى وبعد الفئة الأخيرة طولهما يساوي نفس الطول الفئة المجاورة لها وتكرارها يساوي الصفر، وبعد كل هذا نربط بين الإحداثيات بخط مستقيم 1 ، مثلما هو موضح في الشكل الموالي الممثل للجدول السابق رقم 06

الشكل 3: مدرج تكراري لتوزيع أفراد العينة



ج) **المنحنى البياني**: نفسه نفس الرسم السابق مع اختلافات بسيطة ، يتم عرضه في محورين متعامدين متجانسين ثم نقوم بإنشاء احداثيات الحد الأول مركز الفئة والحد الثاني التكرار المقابل ثم نقوم بربط هذه النقاط الثنائية بخط يدي ومنحني ويشكل لدينا الرسم الموالي2

الفصل الثاني

تنظيم البيانات وعرضها

I. تنظيم البيانات وتلخيصها وعرضها جدوليًا

بعد أن يتم تحديد مجتمع الدراسة والعينة وقياس قيم المتغير على أفراد العينة نحصل على البيانات في صورتها الأولية والتي تسمى البيانات الخام. إن البيانات في صورتها الخام تحتاج إلى تنظيم وتلخيص لكي نتمكن من عرضها بطريقة مناسبة يمكن الاستفادة منها. في هذه الفصل سنتطرق إلى عدة مفاهيم متعلقة بتنظيم وتلخيص وعرض البيانات وهي:

أ- الجداول التكرارية للبيانات الوصفية (النوعية)

ب- الجداول التكرارية للبيانات الكمية

ج- الجداول التكرارية النسبية والجداول التكرارية المئوية

د- مراكز الفترات (الفئات)

هـ- الجداول التكرارية المتجمعة الصاعدة

لتصميم جدول التوزيع التكراري للبيانات فإنه يلزمنا معرفة أن البيانات الإحصائية تنقسم إلى نوعين هما:

أ- **بيانات وصفية:** مثل لون الشعر، فصيلة الدم، الجنس، المستوى التعليمي، وغيرها.

ب- **بيانات كمية:** مثل الطول، الوزن، العمر، عدد الاولاد، وغيرها.

والبيانات الإحصائية سواءً كانت وصفية أم كمية فهي تنظم وتلخص في جداول تسمى الجداول التكرارية (أو جداول التوزيع التكرارية). والجدول التكراري عبارة عن جدول يلخص البيانات فيوزعها على فئات (أو طوائف أو فترات) ويحدد عدد البيانات التي تنتمي لكل فئة. وعدد البيانات التي تنتمي إلى فئة معينة يسمى بتكرار تلك الفئة.

1) الجدول التكراري للبيانات الوصفية

لإنشاء الجدول التكراري للبيانات الوصفية فإننا نقوم بعمل الخطوات التالية:

أ- نحصر جميع الفئات أو الطوائف وهي الصفات المختلفة في البيانات.

ب- ننشئ جدول تفريغ البيانات وهو عبارة عن جدول مكون من ثلاثة أعمدة هي: عمود الفئة (أو الصفة) وعمود العلامات وعمود التكرار.

ج- نوجد الجدول التكراري من جدول تفريغ البيانات وذلك بحذف عمود العلامات.

تمرين 4: جد الجدول التكراري لتقديرات عينة من (60) طالب

C	C	C	B	D	A	E	C	D	B	D	C	E	B	D
C	C	E	D	A	C	E	A	D	D	B	D	C	E	B
D	E	D	C	B	B	D	C	C	D	E	C	A	D	C
A	D	B	D	D	C	D	C	D	D	A	D	D	E	D

الحل:

أ- المتغير = تقدير الطالب (متغير وصفي / نوعي).

ب- حجم العينة = عدد البيانات = n = 60 .

ج- البيانات عبارة عن تقديرات الطلاب وهي بيانات وصفية.

د- القيم أو الصفات المختلفة للبيانات هي: A, B, C, D, E .

نقوم بتلخيص البيانات السابقة في توزيع تكراري وفق الخطوات التالية:

الشكل 4: تفريغ البيانات التكرارية

التكرار (f) (عدد الطلاب)	العلامات	الفئة أو الصفة (التقدير)
6	𝍸 |	A
8	𝍸 |||	B
16	𝍸 𝍸 𝍸 |	C
22	𝍸 𝍸 𝍸 𝍸 ||	D
8	𝍸 |||	E

الشكل 5: التوزيع التكراري لتقديرات (60) طالب

التكرار (f) (عدد الطلاب)	الفئة أو الصفة (التقدير)
6	A
8	B
16	C
22	D
8	E
$n = 60$	المجموع

ملاحظة (1): ينبغي ملاحظة أن مجموع التكرارات = عدد البيانات = عدد الطلاب = n = 60.

III. الجدول التكراري النسبي والجدول التكراري المئوي للبيانات الوصفية

نهتم كثيرًا بمعرفة التكرارات النسبية والتكرارات المئوية لكل فئة من فئات الجدول التكراري. فمعرفتنا بالتكرار النسبي للفئة أكثر فائدة من مجرد معرفة التكرار.

1) تعريف التكرارين النسبي والمئوي

نعرف التكرار النسبي والتكرار المئوي للفئة كما يلي:

أ- **التكرار النسبي للفئة** = تكرار الفئة ÷ عدد البيانات

= تكرار الفئة ÷ مجموع التكرارات

ب- **التكرار المئوي للفترة** = التكرار النسبي × 100%

تمرين 5: جد الجدول التكراري النسبي والجدول التكراري المئوي في التمرين (1).

الحل: (الجدول أسفله)

جدول 7: تكرارات نسبية ومئوية الخاصة بالتمرين(1)

الفئة أو الصفة (التقدير)	التكرار (f)	التكرار النسبي= f/n	التكرار المئوي= (f/n) * 100%
A	6	6/60 = 0.100	10.0%
B	8	8/60 = 0.133	13.3%
C	16	16/60 = 0.267	26.7%
D	22	22/60 = 0.367	36.7%
E	8	8/60 = 0.133	13.3%
المجموع	n = 60	1.000	100%

ملاحظة (2): ينبغي ملاحظة ما يلي في الجداول التكرارية:

أ- مجموع التكرارات = عدد البيانات = n

ب- مجموع التكرارات النسبية = 1.00

ج- مجموع التكرارات المئوية = 100%

2) الجدول التكراري للبيانات الكمية

تختلف طريقة إنشاء الجداول التكرارية للبيانات الكمية عنها للبيانات الوصفية. إذ لو اتبعنا الطريقة السابقة وذلك بتحديد القيم المختلفة وتكرارتها قد نجد أنفسنا، على سبيل المثال، أمام جدول يحوي جميع القيم كفئات وبتكرارات مساوية للواحد. وعليه نكون قد عقدنا عرض البيانات بدلاً عن تلخيصها. ولذلك فإننا

نلجأ إلى تقسيم مدى البيانات الكمية إلى فترات. ولإنشاء الجدول التكراري للبيانات الكمية فإننا لا بد من مراعاة ما يلي:

أ- نختار بطريقة مناسبة الفترات (الفئات) المختلفة لقيم المتغير وهي عبارة عن فترات على خط الأعداد الحقيقية. وينبغي ملاحظة ما يلي عند تحديد الفترات:

1. الفترات غير متداخلة: لا يمكن لمشاهدة أن تصنف في أكثر من فترة.
2. الفترات شاملة للبيانات: كل مشاهدة لابد أن تصنف في إحدى الفترات.
3. أصغر قيمة لابد أن تقع في الفئة الأولى (الفئة الدنيا) وأكبر قيمة لابد أن تقع في الفئة الأخيرة (الفئة العليا).
4. لايوجد طريقة واحدة لاختيار الفترات إذ أن اختيار الفترات يحدده الباحث بطريقة مناسبة لبحثه.
5. بداية الفئة يسمى بالحد الأدنى للفترة ونهاية الفئة يسمى بالحد الأعلى للفترة

ب- نقوم بتوزيع البيانات على هذه الفترات ثم نقوم بحصر أو تحديد عدد البيانات الواقعة في كل فترة من الفترات.

ج- ننشئ جدول تفريغ البيانات وهو عبارة عن جدول مكون من ثلاثة أعمدة هي: الفترات والعلامات والتكرار.

د- نوجد الجدول التكراري من جدول تفريغ البيانات وذلك بحذف عمود العلامات.

تمرين 6: في أحد البحوث التي أجريت لدراسة مستوى الهيموجلوبين قام الباحث باختيار عينة مكونة من خمسين شخصا فحصل على البيانات المبينة في الجدول أسفله:

جدول 8: مستوى الهيموجلوبين لعينة المفحوصين

17.3	15.7	17.1	16.2	16.2	15.9	17.7	17.0
16.1	16.4	16.2	13.7	16.4	15.3	15.8	14.6
18.3	15.5	16.1	17.8	14.9	16.4	16.2	14.0
15.0	17.4	16.3	15.9	16.8	13.9	15.3	15.9
16.3	14.4	16.5	17.4	15.1	15.7	16.1	14.2
15.8	17.0	13.5	15.8	15.1	16.7	15.9	16.3
						17.3	17.5

تمرين 7: جد جدول التوزيع التكراري لبيانات مستوى الهيموجلوبين لهؤلاء الأشخاص باستخدام الفترات التالية:

$$12.95-13.95,\ 13.95-14.95,\ 14.95-15.95,$$
$$15.95-16.95,\ 16.95-17.95,\ 17.95-18.95.$$

الحل:

أ- المتغير = مستوى الهيموجلوبين (متغير كمي).

ب- حجم العينة = عدد البيانات = n = 50.

ج- البيانات عبارة عن قيم مستوى الهيموجلوبين وهي بيانات كمية.

د- أكبر قيمة= 18.3 وأصغر قيمة = 13.5.

نقوم بتلخيص البيانات السابقة في توزيع تكراري وفق الخطوات التالية:

جدول 9: تفريغ العلامات التكرارية لبيانات مستوى الهيموجلوبين

التكرار (f) (عدد الأشخاص)	العلامات	مستوى الهيموجلوبين (فترة الفئة)
3	ǀǀǀ	12.95 – 13.95
5	~~ǀǀǀǀ~~	13.9 5– 14.95
15	~~ǀǀǀǀ~~ ~~ǀǀǀǀ~~ ~~ǀǀǀǀ~~	14. 95– 15.95
16	~~ǀǀǀǀ~~ ~~ǀǀǀǀ~~ ~~ǀǀǀǀ~~ ǀ	15. 95– 16.95
10	~~ǀǀǀǀ~~ ~~ǀǀǀǀ~~	16. 95– 17.95
1	ǀ	17. 95– 18.95

جدول 10: التكرارات الفئوية لمستوى الهيموجلوبين

التكرار(f) (عدد الأشخاص)	**مستوى الهيموجلوبين (فترة الفئة)**
3	12.95 – 13.95
5	13.9 5– 14.95
15	14. 95– 15.95
16	15. 95– 16.95
10	16. 95– 17.95
1	17. 95– 18.95
n = 50	**المجموع**

3) طريقة عمل الفترات للجدول التكراري المنتظم للبيانات الكمية

يقال بأن الجدول التكراري منتظمًا إذا تساوت أطوال فتراته. وفي الحقيقة، فإنه لا يوجد قواعد ثابته ومتفق عليها لتحديد عدد الفترات وأطوالها وبداياتها ونهاياتها. كما أن هناك عدة عوامل قد تؤثر على طريقة اختيار الفترات، مثل، عدد البيانات، n، وطبيعة البيانات وأهداف البحث وطريقة اختيار الفترات للدراسات السابقة الشبيهة التي يراد مقارنة نتائجها بنتائج الدراسة بين أيدينا.

يعتمد عدد الفترات k على عدد البيانات. وينبغي ألا يكون عدد الفترات كبيرًا حتى لا نفقد الهدف المرجو من الجدول التكراري والمتمثل في تلخيص البيانات. ومن جهة أخرى، ينبغي ألا يكون عدد الفترات صغيرًا فيفقد جدول التوزيع التكراري معالمه ونفقد بذلك كثيرًا من المعلومات. وكقاعدة عامة، ينبغي أن لا تقل عدد الفترات عن خمس فترات و لا تزيد عن عشرين فترة. ويمكن استخدام قاعدة ستارج (Sturges' rule) لتحديد عدد الفترات k كما يلي:

$$k = 1 + 3.322 \log(n)$$

حيث أن n هو عدد البيانات، ومع ملاحظة تقريب القيمة الناتجة للأعلى لأقرب عدد صحيح. (الجدول أسفله)

جدول 11: أعداد الفترات لأحجام عينات مختارة

n	25	45	100	200	350	700	1500
k	6	7	8	9	10	11	12

يرمز لطول الفئة في الجدول التكرار المنتظم بالرمز w، وهو الفرق بين الحد الأعلى والحد الأدنى للفترة الفعلية. ولتحديد طول الفئة للجدول التكراري المنتظم، يحسن بنا التطرق إلى مدى البيانات ووحدة دقة البيانات. فأما مدى البيانات R فهو الفرق بين أكبر قيمة وأصغر قيمة للبيانات. أي أن:

$$R = X_{\max} - X_{\min}$$

حيث أن $X_{\max}$ هي أكبر قيمة و $X_{\min}$ هي أصغر قيمة. وأما وحدة دقة البيانات، u، فهي أصغر وحدة (خانة عشرية) مستخدمة لتقريب البيانات الأصلية. فعلى سبيل المثال، إذا كانت البيانات على الشكل:

2.39, 3.05, 3.20, 1.17, 1.90, 2.55, 3.16

فإن وحدة الدقة هي u=0.01.

بعد تحديد عدد الفترات k، نحدد طول الفئة للجدول التكراري المنتظم باستخدام العلاقة التالية:

$$w = \frac{R}{k}$$

مع ملاحظة تقريب القيمة الناتجة للأعلى إلى أقرب وحدة دقة.

يوجد نوعان من الفترات، هي الفترات التقريبية والفترات الفعلية. يتم أولاً تحديد الفترات التقريبية ومن ثم

تحديد الفترات الفعلية. ولتحديد الفترات التقريبية، لنفرض أن الفئة التقريبية الأولى هي $L_1 - U_1$ والثانية هي $L_2 - U_2$، وهكذا، حتى الفئة التقريبية الأخيرة رقم k وهي $L_k - U_k$. حيث يمثل L_i الحد الأدنى ويمثل U_i الحد الأعلى للفترة التقريبية رقم i مقربة بنفس وحدة تقريب البيانات الأصلية. ويجب أن يكون الحد الأدنى للفترة التقريبية الأولى أقل من أو يساوي أصغر قيمة، أي أن $L_1 \le X_{\min}$. ويجب أن يكون الحد الأعلى للفترة التقريبية الأخيرة أكبر من أو يساوي أكبر قيمة، أي أن $U_k \ge X_{\max}$. ويجب أن تكون الفترات التقريبية غير متداخلة بحيث تحقق:

$$L_1 < U_1 < L_2 < U_2 < \cdots < L_k < U_k$$

ولتكوين الفترات التقريبية، نحدد أولاً الحد الأدنى للفترة التقريبية الأولى (الدنيا) بالقيمة L_1 مقربة بنفس وحدة تقريب البيانات الأصلية، ومن ثم نستخدم القوانين التالية لتحديد حدود الفترات التقريبية الباقية:

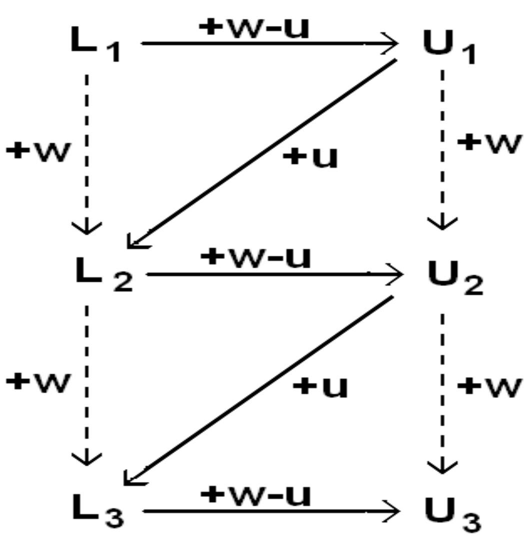
$$U_i = L_i + w - u; \quad i = 1, 2, \ldots, k$$

$$L_i = U_{i-1} + u \ ; \ i = 2, 3, \ldots, k$$

$$L_i = L_{i-1} + w \ ; \ i = 2, 3, \ldots, k$$

$$U_i = U_{i-1} + w \quad i = 2, 3, \ldots, k$$

الشكل 6: طريقة عمل الفترات التقريبية



شكل (1): طريقة عمل الفترات التقريبية

ينبغي التنويه إلى أن الفترات التقريبية غير متلاصقة بحيث أن الفرق بين الحد الأعلى للفترة التقريبية والحد الأدنى للفترة التي تليها يساوي وحدة دقة.

بعد تحديد حدود الفترات التقريبية، نقوم بتحديد حدود الفترات الفعلية على النحو التالي: نضيف نصف وحدة الدقة للحد الأعلى التقريبي للحصول على الحد الأعلى الفعلي، ونطرح نصف وحدة الدقة من الحد الأدنى التقريبي للحصول على الحد الأدنى الفعلي. فلو فرضنا أن $L_i^* - U_i^*$ هي الفئة الفعلية رقم i، فإن:

$$L_i^* = L_i - \frac{u}{2}; \quad i = 1,\ 2,\ \ldots,\ k$$

$$U_i^* = U_i + \frac{u}{2}; \quad i = 1,\ 2,\ \ldots,\ k$$

ويمكن التحقق من أن $L_i^* = L_{i-1}^* + w$ و $U_i^* = U_{i-1}^* + w$. وينبغي التنويه إلى أن الفترات الفعلية متلاصقة بحيث أن الحد الأعلى للفترة الفعلية هو الحد الأدنى للفترة التي تليها. كما تجدر الإشارة إلى استحالة تساوي أي قيمة من البيانات بأي حد من حدود الفترات الفعلية، وذلك لأن الخانات العشرية للحدود الفعلية للفترات تزيد بمقدار خانة عشرية عن الخانات العشرية للبيانات.

تمرين 8:

لنفرض أن لدينا بيانات عددها n=45 ولنفرض أن هذه البيانات مسجلة إلى ثلاث خانات عشرية بحيث أن أصغر قيمة هي 2.517 وأكبر قيمة هي 2.548. والمطلوب هو تكوين الفترات التقريبية والفترات الفعلية بفرض أن الحد الأدنى للفترة التقريبية الأولى (الدنيا) هو $L_1 = 2.515$.

الحل: من المعلومات المعطاة نستنتج ما يلي:

$$k = 1 + 3.322\ \log(45) = 6.49 \approx 7$$

$$R = X_{\max} - X_{\min} = 2.548 - 2.517 = 0.031$$

$$u = 0.001$$

$$w = \frac{R}{k} = \frac{0.031}{7} = 0.004443 \approx 0.005$$

باستخدام القيمة المعطاة للحد الأدنى للفترة التقريبية الأولى (الدنيا) $L_1 = 2.515$ ، وباستخدام القوانين التالية:

$$U_i = L_i + 0.005 - 0.001 \; ; i = 1, 2, \ldots, 7$$

$$L_i = U_{i-1} + 0.001 \; ; \; i = 2, 3, \ldots, 7$$

$$L_i^* = L_i - \frac{0.001}{2} = L_i - 0.0005 ; \; i = 1, 2, \ldots, 7$$

$$U_i^* = U_i + \frac{0.001}{2} = U_i + 0.0005 ; \; i = 1, 2, \ldots, 7$$

جدول 12: الفترات التقريبية والفترات الفعلية

الفترات التقريبية		الفترات الفعلية	
الحد الأعلى	الحد الأدنى	الحد الأعلى	الحد الأدنى
2.519	2.515	2.5195	2.5145
2.524	2.520	2.5245	2.5195
2.529	2.525	2.5295	2.5245
2.534	2.530	2.5345	2.5295
2.539	2.535	2.5395	2.5345
2.544	2.540	2.5445	2.5395
2.549	2.545	2.5495	2.5445

نلاحظ أن أصغر قيمة 2.517 تقع في الفئة الأولى (الدنيا)، وأكبر قيمة 2.548 تقع في الفئة الأخيرة (العليا). كما نلاحظ أن الفترات التقريبية غير متلاصقة (الفرق بين كل فترة والتي تليها مقدار وحدة دقة 0.001)، وأما الفترات الفعلية فهي متلاصقة.

4) الجدول التكراري النسبي والجدول التكراري المئوي للبيانات الكمية

كما مر معنا سابقًا، فإننا نعرف التكرار النسبي والتكرار المئوي للفئة كما يلي:

أ- التكرار النسبي للفئة = تكرار الفئة ÷ عدد البيانات

= تكرار الفئة ÷ مجموع التكرارات

ب- التكرار المئوي للفترة = التكرار النسبي × 100%

تمرين 9

أوجد الجدول التكراري النسبي والجدول التكراري المئوي في مثال (2).

الحل:

جدول 13: التكرارات النسبية المئوية لمستوى الهيجلوبين

مستوى الهيموجلوبين (فترة الفئة)	التكرار (f)	التكرار النسبي= f/n	التكرار المئوي= (f/n) * 100%
12.95 – 13.95	3	3/50= 0.06	6%
13.9 5– 14.95	5	5/50= 0.10	10%
14. 95– 15.95	15	15/50= 0.30	30%
15. 95– 16.95	16	16/50= 0.32	32%
16. 95– 17.95	10	10/50= 0.20	20%
17. 95– 18.95	1	1/50= 0.02	2%
المجموع	$n = 50$	1.00	100%

5) مركز الفئة وطول الفئة

نعرف في الجدول التكراري تكرار كل فترة من الفترات ولكننا لا نعرف البيانات الفعلية لكل فترة إلا إذا كانت البيانات الخام متوفرة لدينا. ولذلك فإننا بحاجة إلى تقدير قيم البيانات لكل فترة وذلك بما يسمى مركز الفئة.
إذا كانت فترات الجدول التكراري ذات أطوال متساوية، فإننا نقول بأن الجدول منتظم في أطوال الفترات.

تعريف (2): نعرف مركز الفئة على أنه منتصف الفئة، أي أن:

$$\text{مركز الفئة} = \frac{\text{الحد الأعلى + الحد الأدنى}}{2}$$

تعريف (3): نعرف وطول الفئة على أنه الفرق بين الحد الأعلى والحد الأدنى للفترة، أي أن:
طول الفئة (L) = الحد الأعلى للفترة مطروحًا منه الحد الأدنى للفترة

تمرين 10

جدول 14: التوزيع التكراري متضمنا مراكز الفترات لمستوى الهيموجلوبين في مثال (2):

مستوى الهيموجلوبين (فترة الفئة)	مركز الفئة (m)	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

طول الفئة لهذا الجدول التكراري هو:

$$L = 13.95 - 12.95 = 1.00$$

ملاحظة (3):

1. مركز الفئة = مركز الفئة السابقة + طول الفئة.
2. إن عدد البيانات (التكرار) الواقعة في كل فترة يكون معروفا للتوزيع التكراري. ونظرًا لأن البيانات الأصلية في كل فترة غير معروفة فإن مركز الفئة يستخدم كقيمة تقريبية لجميع البيانات الواقعة في تلك الفئة. أي أن مركز الفئة يعتبر قيمة مثالية أو نموذجية لجميع البيانات الواقعة في تلك الفئة.

6) الجدول التكراري المتجمع الصاعد

نستطيع من خلال الجدول التكراري معرفة عدد البيانات الواقعة في كل فترة (تكرار الفئة). وأما أهمية الجدول التكراري المتجمع الصاعد فتكمن في استطاعتنا من خلاله معرفة عدد البيانات التي تقل عن أو تساوي قيمة معينة وكذلك معرفة عدد البيانات التي تقع بين قيمتين معيانتين.

تعريف: التكرار المتجمع الصاعد للفترة يساوي عدد البيانات التي تقل عن أو تساوي الحد الأعلى للفترة. أي أن التكرار المتجمع الصاعد للفترة ما هو إلا تكرار الفئة مضافًا إليه مجموع تكرارات الفترات السابقة لها.

ويمكن وبشكل مشابه تعريف التكرار النسبي المتجمع الصاعد والتكرار المئوي المتجمع الصاعد. وكما ذكرنا سابقًا، إننا نستخدم الجدول التكراري النسبي المتجمع الصاعد لإيجاد نسبة البيانات التي تكون قيمتها أصغر من أو يساوي قيمة معينة. كما يستخدم لإيجاد نسبة البيانات التي تكون قيمتها محصورة بين قيمتين معينتين.

تمرين 11: في مثال (2):

عدد الأشخاص الذين مستوى الهيموجلوبين لهم يقل عن أو يساوي 12.95 يساوي 0.

عدد الأشخاص الذين مستوى الهيموجلوبين لهم يقل عن أو يساوي 13.95 يساوي 3.

عدد الأشخاص الذين مستوى الهيموجلوبين لهم يقل عن أو يساوي 14.95 يساوي 8.

عدد الأشخاص الذين مستوى الهيموجلوبين لهم يقل عن أو يساوي 15.95 يساوي 23.

عدد الأشخاص الذين مستوى الهيموجلوبين لهم يقل عن أو يساوي 16.95 يساوي 39.

عدد الأشخاص الذين مستوى الهيموجلوبين لهم يقل عن أو يساوي 17.95 يساوي 49.

عدد الأشخاص الذين مستوى الهيموجلوبين لهم يقل عن أو يساوي18.95 يساوي 50.

إذا وضعنا هذه المعلومات في جدول فإننا نحصل على الجدول التكراري المتجمع الصاعد كما يلي:

جدول 15: التكرارات المتجمعة الصاعدة

التكرار المتجمع الصاعد	مستوى الهيموجلوبين	
0	أقل من	12.95
3	أقل من	13.9 5
8	أقل من	14. 95
23	أقل من	15. 95
39	أقل من	16. 95
49	أقل من	17. 95
$50 = n$	أقل من	18.95

من هذا الجدول نستطيع استخراج كثير من المعلومات، فعلى سبيل المثال:

1. عدد الأشخاص الذين مستوى الهيموجلوبين لهم يقل عن أو يساوي 16.95 هو 39 شخصًا.
2. عدد الأشخاص الذين مستوى الهيموجلوبين لهم يتراوح بين 14.95 و 16.95 هو 31 شخصًا حيث يحسب كما يلي: $31 = 8 - 39$.

7) العرض البياني للجداول التكرارية

فيما سبق قدمنا طريقة لتنظيم وتلخيص وعرض البيانات في جداول تكرارية. وفي هذه الفقرة سوف نستعرض التمثيل أو العرض البياني للجداول التكرارية والجداول التكرارية المتجمعة الصاعدة.

1.7. عرض وتمثيل الجدول التكراري بيانيًا

لعرض وتمثل الجدول التكراري أو الجدول التكراري النسبي أو الجدول التكراري المئوي بيانيًا فإننا نستخدم الرسوم التالية:

أ- المدرج التكراري:

المدرج التكراري عبارة عن رسم بياني مكون من محورين. المحور الأفقي يمثل الفترات للمتغير والمحور العمودي يمثل التكرار. يتم رسم المدرج التكراري برسم مستطيلات متلاصقة قواعدها عبارة عن الفترات وارتفاعاتها عبارة عن التكرارات المقابلة.

ب- المضلع التكراري:

المضلع التكراري عبارة عن رسم بياني مكون من محورين. المحور الأفقي يمثل مراكز الفترات والمحور العمودي يمثل التكرار. يتم رسم المضلع التكراري برسم نقاط فوق كل مركز فترة. ارتفاع النقطة عن مركز الفئة عبارة عن تكرار تلك الفئة. وبعد رسم جميع النقاط يتم توصيلها بخطوط مستقيمة ومن ثم غلق المضلع التكراري لكي تكون المساحة تحت المضلع التكراري مساوية لمساحات المستطيلات في المدرج التكراري.

عملية غلق المضلع تتم من خلال افتراض وجود فترة قبل الفئة الدنيا بتكرار مساو للصفر وفترة بعد الفئة العليا بتكرار مساو للصفر. ويتم تحديد مركزي هاتين الفترتين على المحور الأفقي للمضلع التكراري ومن ثم رسم نقطتين فوق هذين المركزين بارتفاع يساوي الصفر. وبعد تحديد هاتين النقطتين يتم وصلهما بالمضلع، الأولى مع الجهة اليسرى والثانية مع الجهة اليمنى.

2.7. عرض وتمثيل الجدول التكراري المتجمع الصاعد بيانيًا

لعرض وتمثل التوزيع التكراري المتجمع الصاعد بيانيًا فإننا نستخدم الرسم المسمى بالمضلع التكراري المتجمع الصاعد. والمضلع التكراري المتجمع الصاعد عبارة عن رسم بياني مكون من محورين. المحور الأفقي يمثل الفترات والمحور العمودي يمثل التكرار المتجمع الصاعد. يتم رسم المضلع التكراري المتجمع برسم نقاط فوق حدود الفترات. ارتفاع النقطة عن حد الفئة عبارة عن التكرار المتجمع الصاعد عند ذلك الحد. وبعد رسم جميع النقاط يتم توصيلها بخطوط مستقيمة.

تمرين 12

يوضح الجدول التكراري المتجمع الصاعد لمستوى الهيموجلوبين لخمسين شخصًا في التمرين2.

1. مثل الجدول التكراري بيانيا باستخدام المدرج التكراري و المضلع التكراري.

2. مثل الجدول التكراري المتجمع الصاعد بيانيا باستخدام المضلع التكراري المتجمع الصاعد.

المطلوب هو تمثيل الجدول التكراري والجدول التكراري المتجمع الصاعد بيانيًا.

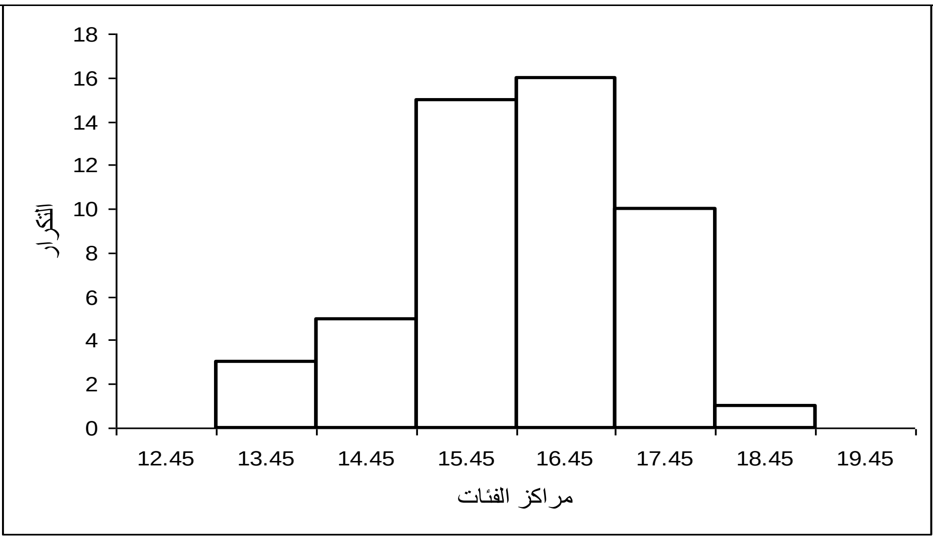
جدول 16: التكرارات المتجمعة الصاعدة لمستوى الهيموجلوبين لـ (50) شخصًا

التكرار المتجمع الصاعد	مستوى الهيموجلوبين
0	أقل من 12.95
3	أقل من 13.9 5
8	أقل من 14. 95
23	أقل من 15. 95
39	أقل من 16. 95
49	أقل من 17. 95
50	أقل من 18.95

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

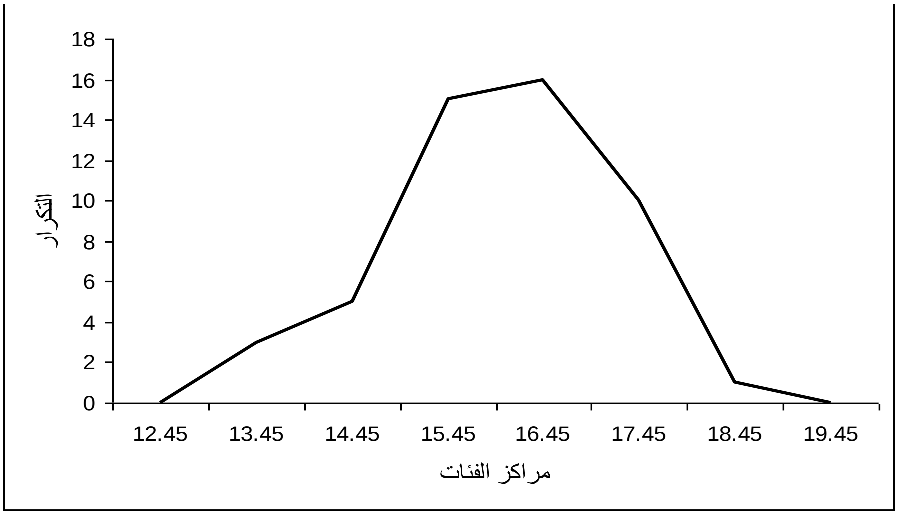
الحل: (1) تمثيل الجدول التكراري بيانيا باستخدام المدرج التكراري و المضلع التكراري:

أ- أولاً: المدرج التكراري:



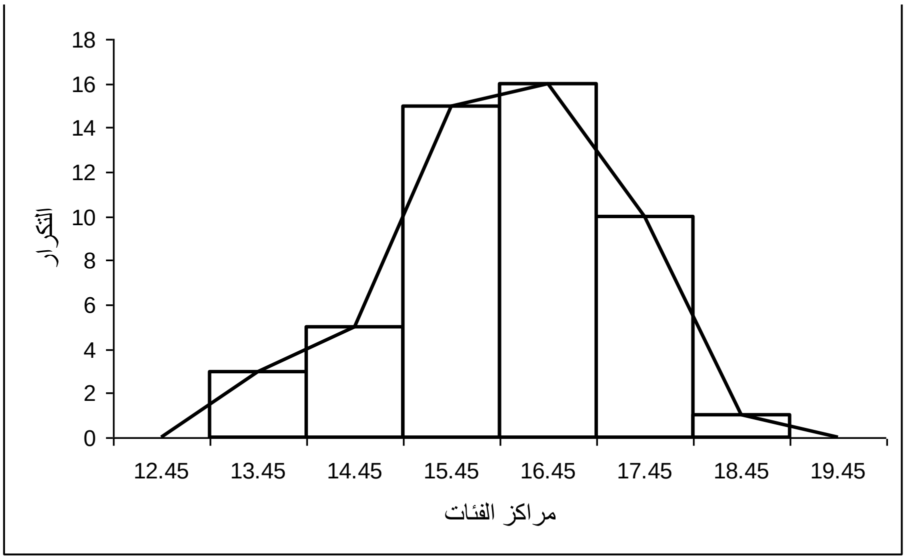
الشكل 7: المدرج التكراري لمستوى الهيموغلوبين في الدم

ب- ثانيًا: المضلع التكراري:



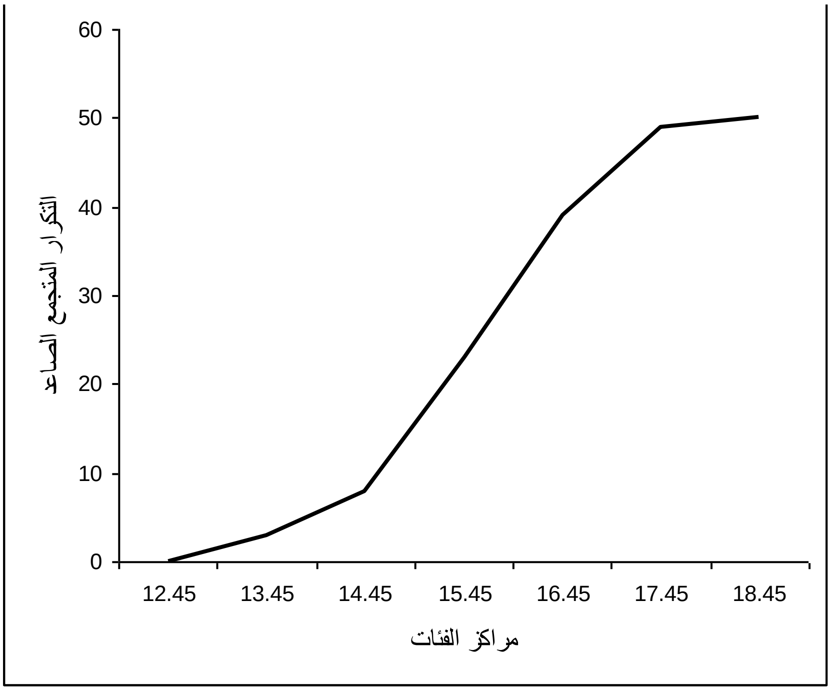
الشكل 8: المضلع التكراري لمستوى الهيموغلوبين في الدم

ج- ثالثًا: المدرج التكراري والمضلع التكراري معًا:



الشكل 9: المدرج التكراري والمضلع التكراري معًا لمستوى الهيموغلوبين في الدم

(2) تمثيل الجدول التكراري المتجمع الصاعد بيانيا باستخدام المضلع التكراري المتجمع الصاعد:



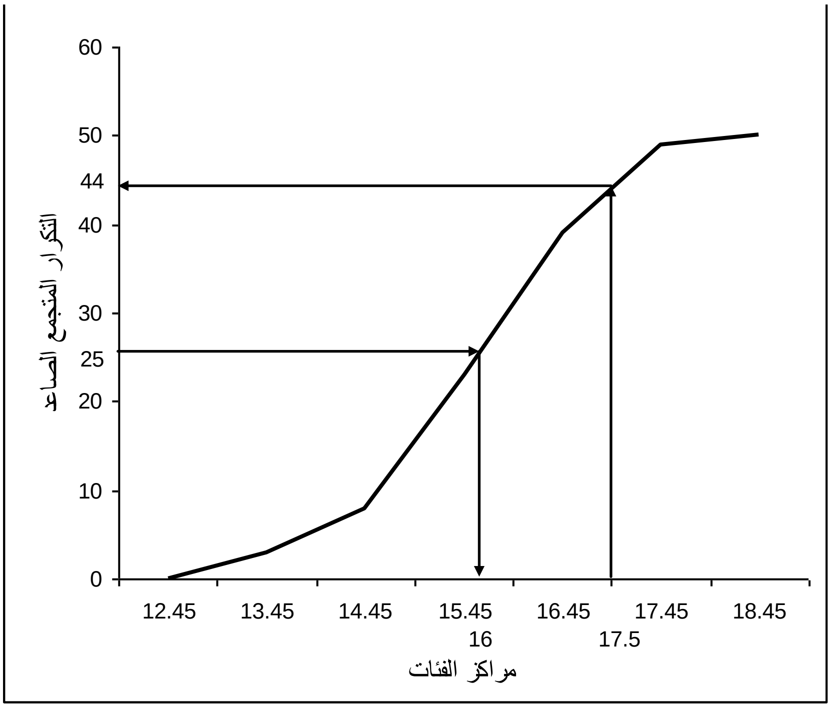
الشكل 10: المضلع التكراري المتجمع الصاعد

ملاحظة (4):

نستطيع بواسطة المضلع التكراري المتجمع الصاعد، كما هو مبين في الشكل (5) أن نحدد ما يلي:

أ- عدد الأشخاص الذين مستوى الهيموجلوبين لهم يقل عن أو يساوي 17.5 هو 44 شخصًا.

ب- مستوى هيموجلوبين نصف الأشخاص (25 شخصًا) يقل عن أو يساوي 16.



الشكل 11: استخدام المضلع التكراري المتجمع الصاعد لمستوى الهيموغلوبين في الدم

الفصل الثالث:

مقاييس النزعة المركزية

(أوالمتوسطات)

Central Tendency Measurements

مقدمة

تعتبر مقاييس النزعة المركزية (أو المتوسطات) من أهم المقاييس الإحصائية التي يفكر الباحث في حسابها، بل هي أول مقاييس إحصائية يفكر فيها الباحث السياسي عموماً. فمقياس النزعة المركزية لظاهرة سياسية ما تعني التعرف على القيمة التي تقع عادة عند مركز التوزيع العددي للقيم المبحوثة.

إن متوسط أي ظاهرة يعبر عن المستوى العام لهذه الظاهرة. فمتوسط مجموعة من القيم هو القيمة التي تعبر عن جميع القيم، أو هو القيمة التي تدور (أو تتركز) حولها باقي القيم. فمتوسط الدخل لأي بلد يعبر عن المستوى العام للدخل في هذا البلد كما أن متوسط دخل عمال أحد المصانع هو الدخل الذي تتركز حوله دخول العمال بهذا المصنع. ولذا تسمى المتوسطات بمقاييس النزعة المركزية. وذلك لأن قيم أي ظاهرة – عادة – تميل أو تنزع للتركز حول قيمة معينة هي متوسط هذه الظاهرة أو مقياس نزعتها المركزية. فأطوال البالغين تتركز حول رقم معين هو متوسط الطول، وكذلك أوزانهم ومعدلات ذكائهم، وأي ظاهرة أخرى. فالمتوسط – بصفة عامة – هو الذي يعبر عن المستوى العام للظاهرة أي هو الذي يعبر عن جميع قيمها، بمعنى أنه القيمة التي تتركز حولها باقي القيم.

في هذا الفصل نستخدم تعبير " المتوسطات " التي تستخدم عادة في البحث السياسي كمرادف لمقاييس النزعة المركزية وذلك لسهولته ووضوح معناه وشيوع استخدامه.

وسوف نتناول في هذا الفصل أهم المتوسطات وهي : الوسط الحسابي والوسيط والمنوال. وفيما يلي عرض لهذه المتوسطات نبين فيه تعريف كل مقياس وكيفية حسابه ومزاياه وعيوبه.

I. الوسط الحسابي The Mean

يعتبر الوسط الحسابي أكثر المتوسطات شهرة وأكثرها استخداماً، بل لعله من أهم المقاييس الإحصائية على الإطلاق، وذلك لما يتمتع به من مزايا وخواص، ولدخوله في حساب الكثير من المقاييس الإحصائية الأخرى كما سيتضح فيما بعد.

الفكرة الأساسية في حساب الوسط الحسابي لمجموعة من القيم أنه يساوي خارج قسمة مجموع القيم على عددها:

$$\text{الوسط الحسابي لمجموعة قيم} = \frac{\text{مجموع هذه القيم}}{\text{عددها}}$$

(ويعرف الوسط الحسابي لمجموعة من القيم بأنه القيمة التي لو حلت محل جميع القيم لا يتغير مجموعها).

تمرين 13: حساب الوسط الحسابي

استغرقت مفاوضات السلام بين بلدين خمس جولات، وكانت كل جولة تستغرق عدة أيام كما يلي :

7, 10, 12, 8, 9

أحسب الوسط الحسابي لعدد الأيام في هذه الجولات.

الحل :

1 – لدينا خمس جولات أو خمس قيم، أي أن عدد القيم = 5 جولات.

2 – مجموع الأيام أو مجموع القيم هو : 10+7+9+8+12= 46

3 – الوسط الحسابي : 5/46 = 9.2 يوماً

أي أن متوسط عدد الأيام في هذه الجولات من المفاوضات هو 9.2 يوماً وللباحث السياسي بعدها حرية إعطاء التفسير لطول أو قصر هذه المدة.

والوسط الحسابي – والذي يقال عنه أحياناً " **الوسط** " **أو** " **المتوسط** " - يمكن أن يكتب بالرموز كما يلي

- نفترض أن عدد القيم هو n

- وأن هذه القيم هي :

$X_1, X_2, X_3, \dots, X_n$

حيث أن :

X_1 تعني القيمة الأولى.

X_2 تعني القيمة الثانية.

X_3 تعني القيمة الثالثة.

X_n تعني القيمة الخيرة رقم n

ومجموع هذه القيم هو : $X_1 + X_2 + X_3 + ... + X_n$

والذي يمكن أن يكتب اختصاراً $\sum X$ أي مجموع القيم حيث :

X ترمز للقيم

والرمز اللاتيني " سيجما " $\sum$ يرمز للمجموع

أي أن :

$\sum X = X_1 + X_2 + X_3 + ... + X_n$

وبالتالي فإن الوسط الحسابي والذي يرمز له بالرمز $\overline{X}$ (والذي ينطق X بار) هو:

$$\overline{X} = \frac{\sum X}{n} \quad (1)$$

فإذا عدنا إلى المثال السابق رقم (1) نجد أن :

- عدد القيم يساوي 5 أي أن : n = 5

والقيم هي :

$X_1 = 7$
$X_2 = 10$
$X_3 = 12$
$X_4 = 8$
$X_5 = 9$

مجموع هذه القيم هو:

$\sum X = X_1 + X_2 + X_3 + ... + X_n$

$= 7+10+12+8+9= 46$

وبالتالي فإن الوسط الحسابي $\overline{X}$ هو :

$$\overline{X} = \frac{\sum X}{n} = \frac{46}{5} = 9.2$$

تمرين 14: حساب الوسط الحسابي

أخذت عينة عشوائية لعدد من سكان أحد الأحياء الفقيرة في دولة نامية حجمها 10 أشخاص، وكانت دخولهم اليومية بالدولار هي :

3.6 , 4.2 , 2.9 , 3.7 , 4.8 , 2.5 , 3.1 , 3.9 , 3.4 , 4.5

أحسب الوسط الحسابي لدخول هؤلاء الأشخاص.

الحل :

1 – عدد الأشخاص يساوي 10 أي أن : $n = 10$

2 – مجموع القيم (مجموع دخولهم اليومية) هو :

$$\sum X = 3.6 + 4.2 + 2.9 + 3.7 + 4.8 + 2.5 + 3.1 + 3.9 + 4.5$$
$$\sum X = 36.6$$

3 – الوسط الحسابي لدخول هؤلاء الأشخاص هو :

$$\overline{X} = \frac{\sum X}{n} = \frac{36.6}{10} = 3.66$$

وهو معدل يعكس بلا شك مجموعة من الحقائق قد يكون أهمها الصعوبات الاقتصادية التي تواجه هذه الدولة. ومن الشرح والأمثلة السابقة يتضح ما يلي :

أولاً : أنه لحساب الوسط الحسابي يجب أن تكون لدينا **بيانات كمية**. أي لا يصلح الوسط الحسابي إذا كانت البيانات وصفية أسمية أو ترتيبية، إذ لا معنى له في هذه الحالات.

ثانياً : أن جميع القيم – بلا إستثناء – تدخل في حساب الوسط الحسابي. أي أنه يعبر عن جميع القيم فعلاً. ولذا فإنه إذا كان من بين القيم قيمة شاذة أو متطرفة (بمعنى أنها كبيرة جداً أو صغيرة جداً بالنسبة لباقي القيم) فإنها سوف تؤثر في قيمة الوسط الحسابي. أي أنه يتأثر بالقيم الشاذة أو المتطرفة. ولتوضيح هذه النقطة بشيء من التفصيل، نورد المثال التالي :

تمرين 15: حساب الوسط الحسابي

إذا كانت أعداد الطلعات الجوية التدريبية الشهرية للمقاتلات الحربية في بلد ما لست من مقاتلاتها هي :

10 , 81 , 84 , 83 , 82 , 80

أحسب الوسط الحسابي الشهري لأعداد هذه الطلعات.

الحل : قبل الشروع في الحل نلاحظ أن عدد هذه الطلعات متقاربة جداً وكلها بأعداد عالية (80 فأكثر) باستثناء طائرة واحدة بلغ عدد طلعاتها (10) طلعات فقط. وهي تمثل قيمة شاذة (أو متطرفة) بالنسبة لباقي طلعات الطائرات الأخرى (حيث أنها صغيرة جداً بالنسبة للطلعات الأخرى). وسوف نرى فيما يلي تأثيرها على قيمة الوسط الحسابي.

1 – عدد القيم (عدد الطلعات) هو 6 أي أن : $n = 6$

2 – مجموع القيمة (أي مجموع الطلعات) هو :

$$\sum X = 80 + 82 + 83 + 84 + 10$$
$$\sum X = 420$$

3 – الوسط الحسابي لعدد الطلعات هو :

$$\overline{X} = \frac{\sum X}{n} = \frac{420}{6} = 70$$

فكأن الوسط الحسابي لعدد الطلعات قد أنخفض إلى 70 طلعة على الرغم من أن كل الطائرات (باستثناء طائرة واحدة شاذة) كانت **80** فاكثر.

الخلاصة : أن الوسط الحسابي يتأثر بالقيم الشاذة أو المتطرفة، فوجود قيمة كبيرة جداً بالنسبة لباقي القيم يرفع قيمة الوسط، والعكس وجود قيمة صغيرة جداً يقلل من قيمة الوسط. لذا فإنه يقال أن الوسط في هذه الحالات قد يكون مضللاً أي لا يعبر عن الغالبية العظمى من القيم. ففي المثال السابق رقم (3) إذا أهملنا القيمة الشاذة نجد أنّ :

$$n = 5$$
$$\sum X = 80 + 82 + 83 + 84 + 81 + = 410$$
$$\overline{X} = \frac{\sum X}{n} = \frac{410}{5} = 82$$

فإهمال القيم الشاذة رفعت قيمة الوسط الحسابي للطلعات من **70** إلى **82** طلعة وهي التي تعبر فعلاً عن جميع القيم (باستثناء القيمة الشاذة). ونخلص من ذلك إلى أنه في حالة وجود قيم شاذة فإن الوسط الحسابي قد يكون مضللاً أي لا يعبر عن غالبية القيم. وفي هذه الحالة فإنه لا يفضل حساب الوسط الحسابي بل نبحث عن متوسط أخر لا يتأثر بهذه القيم الشاذة. أو – كما يرى البعض – نهمل القيمة الشاذة ونحسب الوسط الحسابي لباقي القيم (بدون القيمة الشاذة).

جـ - إذا كان لدينا عينتين، حجم الأولى n_1 وحجم الثاني n_2 وكان الوسط الحسابي لكل منهما $\overline{X}_1$ ، $\overline{X}_2$

فإنه يمكن حساب الوسط الحسابي للعينتين معاً بالاستفادة من متوسط كل منهما كما يلي :

$$\boxed{\overline{X} = \frac{n_1 . \overline{x}_1 + n_2 \overline{x}_2}{n_1 + n_2}} \quad (2)$$

يسمى المتوسط في هذه الحالة " المتوسط المرجح The weighted average " ونلاحظ أنه تم ترجيح كل متوسط بحجم العينة المحسوب منها، أي إعطاء كل متوسط وزن يساوي حجم العينة الخاصة به، ثم القسمة على مجموع حجم العينتين معاً. **<u>أي أنه لحساب الوسط الحسابي للعينتين معاً. لا نجمع الوسطين ونقسم على اثنين إلا في حالة واحدة فقط وهي إذا كان حجم العينتين متساويين.</u>**

أما في الحالة العامة – وهي اختلاف حجم العينتين – فيتم ضرب كل متوسط في حجم العينة الخاص به، ونجمع، ثم نقسم على حجم العينتين معاً $(n_1 + n_2)$ ويمكن تعميم هذا القانون لأي عدد من العينات. فمثلاً إذا كان لدينا ثلاث عينات أحجامها هي n_1 , n_2 , n_3 ومتوسطاتها هي على الترتيب $\overline{X}_1 , \overline{X}_2 , \overline{X}_3$ فإن الوسط المرجح للعينات الثلاث معاً هو :

$$\boxed{X = \frac{n_1 . \overline{x}_1 + n_2 . \overline{x}_2 + n_3 . \overline{x}_3}{n_1 + n_2 + n_3}} \quad (3)$$

تمرين 16: حساب الوسط الحسابي المرجح

إذا كانت لدينا مجموعتين من الطلاب تدرسان المقرر نفسه. وكان عدد الطلاب في المجموعتين هو :

$n_1 = 40$ $\qquad$ $n_2 = 25$

وكان الوسط الحسابي لدرجات الطلاب في المجموعتين هو : $\overline{X}_1 = 75 . \overline{X}_2 = 80$

فما هو الوسط الحسابي لدرجات الطلاب في المجموعتين معاً ؟

الحل : المتوسط المرجح $\overline{X}$ يحسب باستخدام العلاقة رقم (2) :

$$\overline{X} = \frac{n_1 . \overline{x}_1 + n_2 . \overline{x}_2}{n_1 + n_2}$$

$$= \frac{40 \times 75 + 25 \times 80}{40 + 25}$$

$$\frac{3000 + 2000}{65} =$$

$$\frac{5000}{65} = 76.9$$

تمرين 17: حساب المتوسط الحسابي المرجح

في المثال السابق إذا كان لدينا المتوسطان نفسهما، ولكن عدد الطلاب في كل من المجموعتين متساوي وليكن يساوي **40** في كل منهما :

$$\overline{X}_1 = 75 \ , \ \overline{X}_2 = 80$$
$$n_1 = n_2 = 40$$

أحسب الوسط المرجح للمجموعتين معاً ؟

الحل : نلاحظ أن حجمي العينتين (أو المجموعتين) متساويان وبالتالي فإن المتوسط المرجح – في هذه الحالة الخاصة هو:

$$\overline{X} = \frac{\overline{x}_1 + \overline{x}_2}{2}$$
$$= \frac{75 + 80}{2} = 77.5$$

ولو استخدمنا العلاقة رقم (2) وهي الحالة العامة لحصلنا على النتيجة نفسها وذلك كما يلي :

$$\overline{X} = \frac{n_1 \ . \ \overline{x}_1 + n_2 \ . \ \overline{x}_2}{n_1 + n_2}$$
$$= \frac{40 \text{ x } 74 + 40 \, x \, 80}{40 + 40}$$
$$= \frac{3000 + 3200}{80}$$
$$= \frac{6200}{80} = 77.5$$

وهذه النتيجة نفسها بطبيعة الحال. أي أنه في حالة تساوي أحجام العينات فقط تجمع المتوسطات وتقسم على عددها. أما في حالة اختلاف احجام العينات فنحسب المتوسط المرجح (باستخدام العلاقة (2) أو (3) أو الحالة العامة لهما وذلك على حسب عدد العينات).

حساب الوسط الحسابي في حالة البيانات المبوبة ولكن دون فئات

أحياناً تكون البيانات المتوافرة لدى الباحث مبوبة بمعنى أنها تأخذ شكل قيم وتكرارات كما يلي :

تمرين 18: الوسط الحسابي في حالة البيانات المبوبة ولكن دون فئات

الجدول التالي يعطي عينة من الأسر تمثل نمط الإنجاب في واحدة من دول العالم الثالث، والمطلوب حساب الوسط الحسابي لعدد الأطفال في الأسرة في هذه العينة؟

جدول 17: نمط الإنجاب في واحدة من دول العالم الثالث

أعداد الأسر (التكرارات) F	عدد الأطفال بالأسر X
2	4
6	5
10	6
15	7
9	8
5	9
3	10
50	**المجموع**

الحل: أرقام الجدول أعلاه تقول أنّ العيّنة تحوي 50 أسرة منهم أسرتان بكل منهما 4 أطفال، 6 أسر بكل منهم 5 أطفال، 10 أسر لدى كل واحد منهم 6 أطفال... وهكذا.

لكي نحسب الوسط الحسابي نحصل أولاً على مجموع الأطفال ثم نقسم على إجمالي عدد أفراد العينة (والذي يساوي في هذا المثال 50 أسرة). ولكي نحصل على مجموع عدد الأطفال نضرب عدد الأطفال في عدد الأسر ثم نجمع لكل الأسر. فمثلاً الأسرتان اللتان لدى كل منهما 4 أطفال مجموع أطفالهما 8 = 2 × 4 والأسر الست التي لدى كل منهم 5 أطفال مجموع عدد أطفالهم 30 = 6 × 5 وهكذا بالنسبة لباقي الجدول. ويمكن تنظيم ذلك في عمود جديد يضاف إلى الجدول السابق كما يلي:

جدول 18: نمط الإنجاب في واحدة من دول العالم الثالث (حاصر ضرب)

عدد الأطفال بالأسر	أعداد الأسر (التكرارات)	حاصل الضرب (مجموع الأطفال)
X	f	Xf
4	2	$4 \times 2 = 8$
5	6	$5 \times 6 = 30$
6	10	$6 \times 10 = 60$
7	15	$7 \times 15 = 105$
8	9	$8 \times 9 = 72$
9	5	$9 \times 5 = 45$
10	3	$10 \times 3 = 30$
المجموع	$\sum f = 50$	$\sum xf = 350$

نعلم أن الوسط الحسابي لعدد الأطفال بالأسرة يساوي مجموع الأطفال على عدد الأسر، أي مجموع القيم على عددها. وبالتالي فإن المعادلة تصبح :

$$\boxed{\overline{X} = \frac{\sum xf}{\sum f}} \quad (4)$$

وبالتعويض في قانون الوسط الحسابي رقم (4) نحصل على الوسط الحسابي لحجم الأسرة كما يلي :

$$\overline{X} = \frac{350}{50} = 7$$

أي أن الوسط الحسابي لحجم الأسرة بعينة هذا البلد يساوي 7 أطفال، وهو متوسط مرتفع دون شك.

8) الوسط الحسابي في حالة الفئات

كما قد تكون البيانات مبوبة على **فئات** وتكرارات كما يلي :

تمرين 19: حساب الوسط الحسابي في حالة الفئات

احسب الوسط الحسابي لمعطيات الجدول أسفله.

جدول 19: توزيع مجموعة من الطلاب حسب فئات الدرجات

فئات الدرجات Classes	أعداد الطلاب f
2 – 4	3
4 – 6	9
6 – 8	10
8 – 10	5
المجموع	$\sum f = 27$

المطلوب حساب الوسط الحسابي لدرجات الطلاب.

الحــل : الجـــدول يقـــول أن 3 طـــلاب حصـــل كـــل مـــنهم علـــى درجـــة تتـــراوح بـــين 2 **و أقـــل مـــن** 4 (لكن لا نعلم ما درجة كل منهم بالتحديد)، 9 طلاب حصل كل منهم على درجة تتراوح بين 4 **و أقل من** 6 (لكن لا نعلم ما درجة كل منهم بالتحديد)، وهكذا بالنسبة لباقي الفئات. وفي هذه الحالة نحسب مراكز الفئات كأحسن قيم تمثل هذه الفئات. ومركز الفئة هو القيمة التي تقع في منتصف الفئة، أي أن :

$$\text{مركز الفئة} = \frac{\text{الحد الأدنى للفئة + الحد الأعلى لها}}{2}$$

أي أننا نستعيض عن الفئات بمراكزها وهي التي تمثل القيم (كما في الأمثلة السابقة) وسوف نرمز لها بالرمز X، ثم نكمل الحل كما في المثال السابق :

جدول 20: توزيع مجموعة من الطلاب حسب مجموع الدرجات

فئات الدرجات	أعداد الطلاب	مراكز الفئات	حاصل الضرب (مجموع الدرجات)
Classes	f	x	x. f
2 – 4	3	$\frac{2+4}{2}=3$	$3\times 3=9$
4 – 6	9	$\frac{4+6}{2}=5$	$5\times 9=45$
6 – 8	10	$\frac{6+8}{2}=7$	$7\times 10=70$
8 – 10	5	$\frac{8+10}{2}=9$	$9\times 5=45$
المجموع	$\sum f=27$		$\sum xf=169$

بالتعويض في قانون الوسط الحسابي رقم (4) نحصل على :

$$\overline{X}=\frac{\sum xf}{\sum f}$$
$$=\frac{169}{27}=6.26$$

أي أن الوسط الحسابي يساوي 6.26 درجة.

ملاحظة مهمة: عند حساب الوسط الحسابي في حالة الفئات نحسب أولاً مراكز الفئات كأحسن قيم تمثل الفئات – كما ذكرنا – ولذلك يقال أن قيمة الوسط الحسابي في حالة **الفئات** قيمة **تقريبية** (وليست دقيقة exact) وذلك لأننا نفترض – على سبيل التقريب – أن مركز الفئة هو أحسن قيمة تمثل الفئة لأنه ليست لدينا الدرجات الدقيقة التفصيلية لكل طالب. ومن ذلك نستنتج أنه إذا كانت هناك فئة مفتوحة (بمعنى عدم معرفة أحد حديها) فإنه لا يمكن حساب مركز هذه الفئة، وبالتالي لا يمكن حساب الوسط الحسابي في هذه الحالة.

تمرين 20: شامل على المتوسطات

الجدول التالي يمثل أهم الحروب التي شهدها العالم من عام 1945 وحتى عام 1980م. أحسب المتوسطات الثلاثة لهذه البيانات.

جدول 21: أهم الحروب التي شهدها العالم من عام 1945م وحتى عام 1980م

الرقم	العام	العدد	المكان
1	1945	4	سوريا – لبنان، أندونيسا، الصين، ماليزيا
2	1946	2	الهند الصينية، اليونان
3	1947	3	مدغشقر، الهند والباكستان، كشمير
4	1948	4	الفلبين، الحرب العربية الإسرائيلية الأولى. حيدر اباد، بورما
5	1949	0	

الرقم	العام	العدد	المكان
6	1950	3	كوريا، فرموزا، التبت
7	1951	0	
8	1952	1	كينيا
9	1953	0	
10	1954	2	جواتيمالا، الجزائر
11	1955	0	السودان، قبرص
12	1956	3	سينا. هنغاريا، السويس
13	1957	0	
14	1958	2	لبنان، كوبا
15	1959	4	فيتنام، هملايا، راوندا، لاوس
16	1960	2	الكونغو، كولومبيا
17	1961	3	كوبا (خليج الخنازير) جيو، انغولا
18	1962	3	غرب غينيا الجديدة، اليمن، غينيا الأسبانية
19	1963	4	الجزائر – المغرب، قبرص، ماليزيا، الصومال – كينيا
20	1964	3	جنزيار، تايلند، موزنبيق
21	1965	3	الهند – الباكستان، جمهورية الدومينيكان، أندونيسا
22	1966	1	بيافرا
23	1967	1	الحرب العربية الإسرائيلية الثانية
24	1968	1	تشيوكوسلفاكيا
25	1969	4	ماليزيا، السلفادور، تشاد، شمال إيرلندا
26	1970	1	أثيوبيا (أرثيريا)
27	1971	2	كمبوديا – بانجلاديش / كشمير
28	1972	1	برونداي
29	1973	1	الحرب العربية الإسرائيلية الثالثة (حرب أكتوبر)
30	1974	2	العراق (الأكراد) قبرص
31	1975	3	أنغولا، تايمور، لبنان
32	1976	1	أسبانيا / المغرب
33	1977	4	الصومال، أثيوبيا، أثيوبيا (ارثيريا، سوريا – لبنان، ليبيا – مصر
34	1978	6	إيران، نيكاراغو، فيتنام – لاوس، تشاد، زائير، روديسيا (زمبابوي)
35	1979	6	اليمن الشمالية – اليمن الجنوبية، أوغندا – تنزانيا، الصين – فيتنام، فيتنام – كمبوديا، نيكاراغوا، جنوب أفريقيا – انغولا
36	1980	3	روسيا – أفغانستان، العراق – إيران، السلفادور
	المجموع	**85**	

الحل : حساب الوسط الحسابي

1 – عدد القيم (عدد السنوات) $n= 36$

2 – مجموع الحروب خلال تلك الفترة $\sum x = 85$

3 – الوسط الحسابي لعدد الحروب خلال تلك الفترة : $X = \frac{\sum x}{n} = \frac{85}{36} = 2.36$

أي أن الوسط الحسابي لعدد الحروب يساوي 2.36 حرباً في السنة الواحدة.

II. الوسيط The Median

1) تعريف الوسيط

هو القيمة التي تقع في منتصف القيم بعد ترتيبها (تصاعدياً أو تنازلياً). فالوسيط هو القيمة التي تتوسط القيم بعد ترتيبها. فإذا كان عدد القيم فردياً فإنه توجد قيمة واحدة في المنتصف (بعد الترتيب) تكون هي الوسيط. أما إذا كان عدد القيم زوجياً فإنه توجد قيمتان في المنتصف نجمعهما ونقسم على 2 فنحصل على قيمة الوسيط. وبديهي أننا سنحصل على النتيجة نفسها لو كان الترتيب تصاعدياً أو تنازلياً.

تمرين 21: أعمار مجموعة من الناخبين

32 24 20 35 29 فما هو وسيط العمر ؟

الحل : أولاً : نرتب هذه الأعمار تصاعدياً كما يلي :

35 32 **29** 24 20

ثانياً : نلاحظ أن عدد القيم فردي (يساوي 5) وأنه توجد قيمة واحدة في المنتصف هي 29 وبالتالي فإن قيمة الوسيط تساوي 29 سنة.

تمرين 22: دخول بعض الأفراد اليومية بالدولار الأمريكي في إحدى الدول.

11 19 14 18 12 15 أحسب وسيط هذه الدخول ؟

الحل : أولاً : نرتب هذه الدخول تصاعدياً كما يلي :

19 18 **15 14** 12 11

ثانياً : نلاحظ أن عدد القيم زوجي (يساوي 6) وأنه توجد قيمتان في المنتصف هما 15، 14 لذلك نجمعهما ونقسم على 2. أي أن الوسيط يساوي :

$$\frac{14+15}{2}=14.5$$ دولاراً

2) بعض خصائص الوسيط

لا يتأثر الوسيط بالقيم الشاذة أو المتطرفة. وهذا منطقي لأنه يقع في منتصف القيم، والقيم الشاذة إما أن تكون في أول القيم أو أخرها (بعد ترتيب القيم تصاعدياً أو تنازلياً). ففي المثال التالي لدينا عدة مجموعات من القيم المرتبة.

تمرين 23: **10 9 <u>8</u> 6 4**

100 9 **<u>8</u>** 6 4

1000 9 **<u>8</u>** 6 4

نلاحظ أن قيمة الوسيط في الحالات الثلاث تساوي 8 (سواء كانت أكبر قيمة تساوي 10 أو 100 أو 1000) أي لم تتأثر قيمة الوسيط بوجود قيمة شاذة أو متطرفة.

يمكن إيجاد قيمة الوسيط في بعض حالات البيانات الترتيبية Ordinal Data. والمثال التالي يوضح ذلك.

تمرين 24: تقديرات عينة مختارة من الناخبين لاحتمال فوز أحد المرشحين في أحد الانتخابات

good , v.good , fair , good , excellent , fair , good

ولحساب وسيط هذه التقديرات نتبع الخطوات التالية :

الحل : رغم أن البيانات غير كمية إلا أنها ترتيبية أي يمكن ترتيبها (تصاعدياً أو تنازلياً). وترتيبها تصاعدياً يكون كما يلي :

fair , fair , good , good , goo , v.good , excellent

حيث أن التقدير good هو الذي يقع في منتصف التقديرات بعد ترتيبها تصاعدياً فإن وسيط التقديرات هو good أو جيد.

ملاحظة مهمة : ***نلاحظ أن الوسيط هو القيمة التي تقع في منتصف القيم. لذلك يمكن حساب " ترتيب الوسيط " أو" موضع الوسيط " أو رقمه في الترتيب قبل معرفة أو حساب قيمته وذلك حسب القاعدة التالية :***

ترتيب الوسيط = $\frac{\text{عدد القيم} + 1}{2}$ أو يساوي $\frac{n + 1}{2}$

في المثال رقم (21)، عدد القيم 5 لذا فإن ترتيب الوسيط هو $\frac{5+1}{2}=\frac{6}{2}=3$ فالوسيط هو القيمة رقم 3 (بعد ترتيب القيم تصاعدياً) والقيمة **<u>الثالثة</u>** في الترتيب هي 29 وهي تمثل قيمة الوسيط لأعمار الناخبين في المثال الأول.

وفي المثال رقم (22) عدد القيم المعطاة هي 6 لذا فإن ترتيب الوسيط هو : $\frac{6+1}{2}=\frac{7}{2}=3.5$ أي أنه يقع بين القيمتين الثالثة والرابعة، لذلك فقد جمعنا هاتين القيمتين وقسمنا على 2، أي أن قيمة الوسيط لدرجات الطلاب في هذا المثال هي : $\frac{14+15}{2}=14.5$

وفي المثال رقم (24) عدد القيم 7 لذا فإن ترتيب الوسيط هو :

$\frac{7+1}{2}=\frac{8}{2}=4$ أي أن الوسيط هو القيمة **<u>الرابعة</u>** في الترتيب، لذا فإن وسيط التقديرات هو good.

والخلاصة إنه يمكن تلخيص خطوات حساب الوسيط في حالة البيانات غير المبوبة كما يلي :

1 – ترتيب البيانات تصاعدياً.

2 – حساب ترتيب الوسيط والذي يساوي $\frac{n+1}{2}$ حيث n هي عدد القيم.

3 – إذا كان عدد القيم فردياً فإنه توجد قيمة واحدة في المنتصف تكون هي قيمة الوسيط. وإذا كان عدد القيم زوجياً فإنه توجد قيمتان في المنتصف نجمعهما ونقسم على 2 فنحصل على قيمة الوسيط.

3) حساب الوسيط

1 – **ترتيب البيانات تصاعدياً** (لاحظ أن العدد هو 36) أي ترتب عدد الحروف تصاعدياً :

<u>66</u> <u>444444</u> <u>333333333</u> <u>2222222</u> <u>11111111</u> <u>0000</u>

2 – **ترتيب الوسيط** $\frac{n+1}{2}$

$$\frac{n+1}{2}=\frac{36+1}{2}=\frac{37}{2}=18.5$$

أي أن الوسيط يقع بين القيمتين الثامنة عشرة والتاسعة عشرة.

3 – **قيمة الوسيط**

(نجمع القيمتين الثامنة عشرة والتاسعة عشرة ونقسم على اثنين): $=\frac{2+2}{2}=2$

أي أن وسيط الحروف يساوي (2) حرباً في السنة.

III. المنوال The Mode

1) تعريف المنوال

المنوال وهو ثالث المتوسطات ويعرف بأنه القيمة الأكثر تكراراً أو شيوعاً بين القيم، فهو القيمة التي تتكرر أكثر من غيرها. وأحياناً يسمى المنوال **"القيمة الشائعة"** أي الأكثر شيوعاً بين القيم. والمنوال من أكثر المتوسطات استخداماً في الحياة التجارية. حيث تعتمد – على سبيل المثال – مصانع الملابس الجاهزة على المقاييس الشائعة بين الناس لتحديد المقاييس المختلفة لهذه الملابس.

ويتميز المنوال بالسهولة والبساطة سواء في فكرته أو في إيجاد قيمته. وكما سنرى في الأمثلة التالية أنه لا ترتب البيانات ولا تجمع ولا أي شيء من هذا القبيل. فقط نبحث عن القيمة التي تتكرر أكثر من غيرها لتكون منوال القيم.

2) تمارين

تمرين 25: أعمار مجموعة من الناخبين :

36 , **29** , 50 , **29** , 42 , 36 , **29** , 34 , **29** , 25

فما هو منوال هذه الأعمار ؟

الحل : بما أن العمر **29** سنة هو العمر الذي تكرر أكثر من غيره من الأعمار (تكرر 4 مرات) فإن :   منوال العمر = 29 سنة

(لاحظ أن البيانات في هذا المثال كمية)

تمرين 26: منوال تقديرات مجموعة من الطلاب في أحد المقررات؟

*fair , **good** , fair , v.good , **good** , excellent , **good***

فما هو منوال هذه التقديرات ؟

الحل : منوال التقديرات هو التقدير "***good***" لأنه تكرر أكثر من غيره **(تكرر ثلاث مرات)**.

(لاحظ أن البيانات في هذا المثال وصفية ترتيبية)

تمرين 27: منوال تكرارات الجنسيات

أنظر بيانات الجدول أسفله:

جدول 22: توزيع فوج من السائحين لإحدى الدول حسب جنسياتهم

الجنسية	عدد السائحين
ألمانية	50
فرنسية	80
أمريكية	120
إيطالية	90

من هذا الجدول نجد أن منوال الجنسية (أي الجنسية الشائعة أو التي تكررت أكثر من غيرها) **هي الجنسية الأمريكية** (120 سائحاً). **(لاحظ أن البيانات في هذا المثال وصفية اسمية Nominal)**

❖ **بعض الملاحظات على المنوال**

- لاحظنا من الأمثلة السابقة أنه يمكن إيجاد المنوال لكل انواع البيانات (كمية أو ترتيبية أو أسمية).
- حسب تعريف المنوال قد لا تتكرر قيمة أكثر من غيرها، وبالتالي قد لا يوجد منوال لبعض البيانات.

تمرين 28: منوال **أعمار مجموعة من الناخبين**

25 32 48 39 55 40

الحل: لا يوجد منوال لهذه الأعمار.

❖ أيضاً قد يوجد أكثر من منوال واحد للبيانات.

تمرين 29: منوال توزيع مجموعة من الناخبين حسب العمر؟

جدول 23: توزيع مجموعة من الناخبين حسب العمر

الأعمار	أعداد الناخبين	الملاحظة
25	3	
30	5	
المنوال الأول = 35	9	أكبر تكرار (التكرار المنوالي)
40	4	
المنوال الثاني = 45	9	أكبر تكرار (التكرار المنوالي)
50	2	

في هذا الجدول نلاحظ أن العمر 35 تكرر 9 مرات (وهو أكبر تكرار) وكذلك العمر 45 تكرر أيضاً 9 مرات (وهو أكبر تكرار) لذلك فإن :

المنوال الأول = 35 سنة، والمنوال الثاني = 45 سنة.

3) حساب المنوال

يعرف المنوال بأنه القيمة التي تكررت أكثر من غيرها وحيث أن القيمة 3 هي التي تكررت أكثر من غيرها (تكررت تسع مرات) فإن منوال الحروب يساوي (3) حروب في السنة. لاحظ أن المتوسطات الثلاثة ليست بالضرورة متساوية.

ملاحظة : أبحث عن المكان الذي تكرر أكثر من غيره خلال تلك الفترة. أي منوال المكان (أو الدولة أو المنطقة).

الفصل الرابع

مقاييس التشتت

(التباين)

Measures of Dispersion (Variation)

مقدمة

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.

I. مقاييس التشتت *Measures of Dispersion*

في دراستنا لمقاييس النزعة المركزية نسعى – نحن الباحثين – إلى معرفة قيمة وسطية معينة يمكن بوساطتها التعّبير عن البيانات من جهة، ومن جهة أخرى يمكن أن تكون أداة مقارنة بيانات مجموعة واحدة، أو مقارنة بيانات هذه المجموعة مع بيانات مجاميع أخرى. إلا أنّ مقياس النزعة المركزية و قد لا تفي بوصف البيانات بشكل كامل، إذ هي لا تعطي صورة واضحة عن كيفية تجميع أو تشتت هذه البيانات حول نقاطها المركزية. بمعنى أنها لا توفر لنا فهماً كاملاً عن مدى تشتت قيم المجموعة عن الوسط، لذا نحتاج بالإضافة إلى المقاييس الوسطية إلى تقدير أو مقياس أخر مثل التباين، أو التشتت تتوزع في مداه مجموعة القيم. وما علينا في هذا الباب إلا أن نناقش هذا المفهوم (التشتت) والمقاييس التي تعبر عنه.

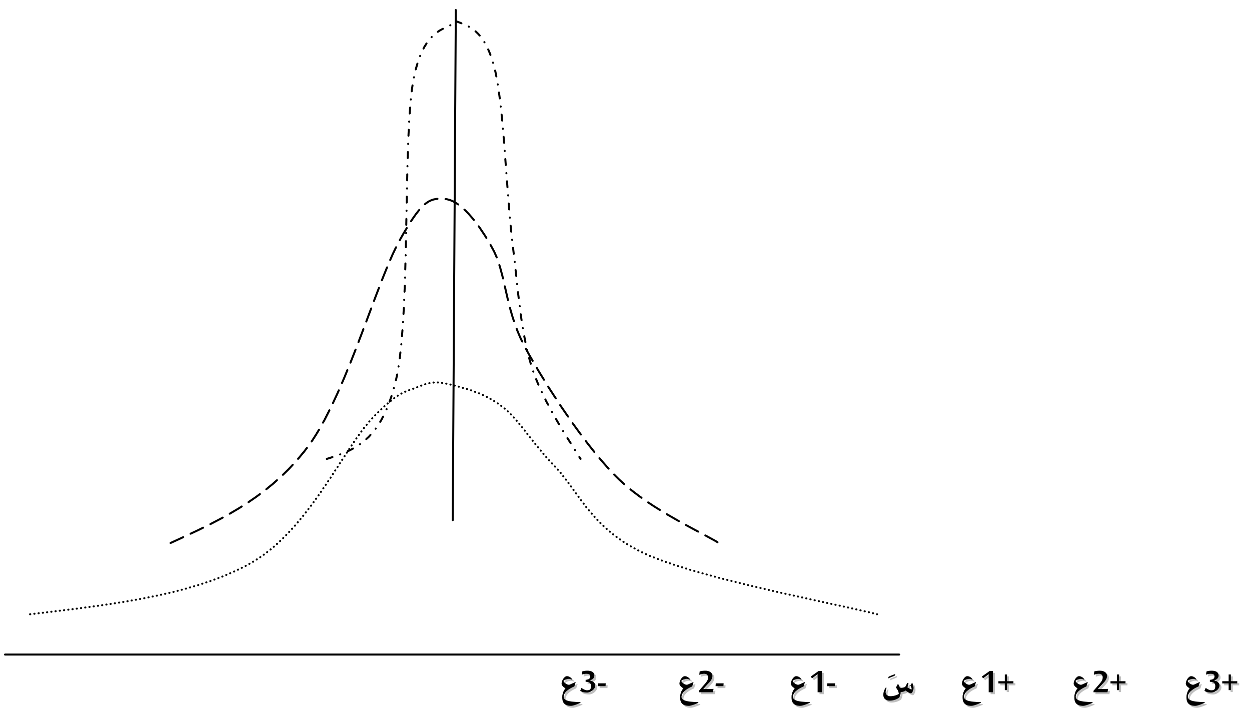
المهم، أن معنى التشتت بمفهومه العام، هو تباعد القيم عن بعضها. لكن، مثل هذا التباعد يحمل في طياته عدة تساؤلات وذلك لعدم تجانس البيانات في أحيان معينة، لذا اتفق بعض من المختصين على أن يكون هنالك نقطة ثابتة لقياس التباعد أو التقارب عن هذه النقطة، ووجد أن القيمة الوسيطية كالوسط الحسابي هي خير من يمثل هذه النقطة. إذ أن غالبية النقاط تكون قريبة من هذه، وعلى هذا قد يكون البعد :

- كبيراً مما يعبر عن تبعثر البيانات.
- قليلاً مما يعبّر عن عدم تبعثر البيانات وانتظامها.
- متساوٍ مما يشير إلى عدم وجود تشتت أو اختلاف.

من كل ما ورد أنفاً نرى أن مقاييس النزعة المركزية تعطينا فكرة عن القيمة التي تتجمع حولها بقية القيم الأخرى في مجموعة من المشاهدات تعبّر عن ظاهرة معينة، إلاّ أن مثل هذه القيمة (الوسطية) لا تدلنا على كيفية توزيع هذه المشاهدات أو على درجة انتشارها وتباعدها عن بعضها البعض من جانب، وعن القيمة المركزية لها من جانب آخر. يضاف إلى هذا كله، قد يكون في أحيان معينة هدفنا مقارنة مجموعتين من البيانات، وعند هذه الحالة لا نكتفي – بمقارنة الأوساط أو القيم الوسطية، لأنه قد يكون للمجموعتين نفس القيمة الوسطية ولكنهما يختلفان في درجة تشتتهما. إذ نجد مثلاً أن قيم إحدى المجموعتين متجمعة حول وسط المجموعة في حين أن مفردات المجموعة الأخرى منتشرة ومتباعدة عن وسطها، عند هذا تكون المجموعة الأولى أقل تشتتا من المجموعة الثانية. وعلى هذه، ولكي نتمكن من وصف توزيع مفردات مجموعة ما والمقارنة بين قيمها بصورة دقيقة، يجدر بنا أن نعرف مقدار التباعد ما بين القيمة الوسطية ومدى انتشار القيم الأخرى في المجموعة (التشتت بين قيم مفردات المجموعة)، إذ أن مقياس التشتت (درجة التباعد)، هو الذي يعوّل عليه في إعطاء الفكرة الواضحة عن صلاحية القيم الوسيطية (المتوسطات) في تمثيل مفردات المجموعة الواحدة. فإذا كانت نتيجة مقياس التشتت كبيرة، كان التباعد بين قيم المجموعة واسعاً، وعندئذ لا يمكن الاعتماد على المتوسط في

تمثيل مفرداته، أما إذا كانت نتيجة مقياس التشتت صغيرة، فان هذا يعني أن القيم متقاربة وقريبة من متوسطها. ولم يقف الأمر عند هذا وإنما إذا ما كانت نتيجة مقياس التشتت تساوي (صفراً) فمعنى هذا أن القيمة الوسيطية (المتوسط) يساوي كل مفردة من المفردات (القيم) في المجموعة. لهذا سيكون اعتمادنا الأساس في المعالجات الإحصائية المعنية بالمقارنات على القيم الوسيطية (المتوسطات) المعنية بالمجموعات ذات المفردات القريبة من هذه القيمة المتوسطة فقط.

لتوضيح الأمر بشكل أفضل نتناول الشكل أدناه ونفسر مضامينه إذ يبين لنا أن هنالك بيانات إحصائية متعددة لها نفس الوسط الحسابي إلاّ أن توزيع هذه البيانات يختلف من توزيع لآخر مما يشير إلى الوسط الحسابي لوحدة وإن كان مشتركاً في توزيعات مختلفة (أي أن الوسط الحسابي نفسه) لا يعطي وصفاً إحصائياً دقيقاً وكافياً للبيانات الإحصائية.



نخلص إلى أنّه من الضروري الاهتمام في دراسة الفروقات في توزيع مفردات أي ظاهرة مقاسة بجانب قياس قيمتها الوسطية، ويقاس ذلك بمقاييس مختلفة، تختلف فيما بينها في الدقة وطرائق الحساب. ومن مقاييس التشتت ما يأتي :

II. المـدى *L'étendu*

1) تعريف المدى

يقصد بالمدى " الفرق بين أكبر قيمة من مجموعة قيم الظاهرة ما وأصغرها في حالة المفردات غير المبوبة، أو هو الفرق بين الحد الأدنى للفئة الدنيا والحد الأعلى للفئة العليا في حالة البيانات المبوبة ". وهذا يعني أن المدى يعبّر عن مقدار الفرق الموجود بين أعلى وأوطأ قيمتين في التوزيع، فلو وضعت القيم في ترتيب تصاعدي وعلى خط مستقيم لكان المدى هو المسافة الموجودة ما بين النهايتين.

يعد المدى من أبسط أنواع مقاييس التشتت، لاسيما عندما تستخدم المقاييس الإحصائية في التعرف على مدى تباعد التوزيعات، وهو أقلها دقة من حيث اتخاذه قيمة معبرّة عن وصف المجموعة (مجموعة القيم) أو عندما يراد المقارنة بين المجموعات الإحصائية. فعندما يستعمل المدى في مقارنة مجموعتين من القيم فإن المجموعة التي تكون قيمة مداها (مقدار تشتتها) أقل، تكون هي الأكثر تجانساً من المجموعة الثانية. وعلى قيمة المدى وتبويب بيانات المجموعة أو المجاميع تتحدد تسميته، فقد يسمى بالمدى المقصور عندما يحسب من بيانات غير مبوبة لا حدود عليا أو دنيا لأي من قيمه الكبرى أو الصغرى، في حين يسمى بالمدى المطلق أو الشامل عندما يحسب من الحدود الدنيا أو العليا للقيم الصغرى أو الكبرى. وعلى هذا سيكون (المدى المقصور) هو الفرق ما بين أكبر وأصغر قيمتين في توزيع مجموعة قيم، أما (المدى المطلق) فهو الفرق ما بين الحد الأدنى للقيمة الصغرى والحد الأعلى للقيمة الكبرى.

بناءاً على ما جاء آنفاً، نرى أن سبب بساطة المدى وقلة دقته يعود إلى جملة من الأسباب أهمها ما يأتي :

1/ قيمة المدى لا تعطينا أية فكرة عن طبيعة ومدى انتشار قيم التوزيع حول أحد مقاييس النزعة المركزية.

2/ إن المدى يتأثر بالقيم المتطرفة فقط، وقد لا يأخذ بنظر الاعتبار القيم الأخرى في التوزيع مما يعطي صورة غير صحيحة عن البيانات الإحصائية التي نقوم بوصفها وتفسيرها. فمثلاً قد تظهر في بعض من البيانات قيم متطرفة جداً يتأثر بها المدى مما يجعل البعد بين القيمتين الصغرى والكبرى كبيراً، ولهذا هنالك من ينصح من الإحصائيين بحذف القيم المتطرفة عن طريق استخدام مفاهيم معينة، منها المدى المئيني، نصف المدى المئيني، المدى العشري، المدى الربيعي، ونصف المدى الربيعي... وغير ذلك.

3/ لا يمكن الاستفادة من المدى للمقارنة بين عدة مجموعات في مدى تباينها وتشتتها.

4/ لا يمكن للمدى أن يعطينا أية صورة واضحة عن شكل التوزيع لأنه لا يهتم إلاّ بقيمتين فقط.

بقي أن نقول : إن ما يعاب على المدى، اعتماده على القيمتين المتطرفتين فقط، واللتان غالباً ما تكونا شاذتين عن

قيم المجموعة. ولكونه مضللاً عند مقارنة المجموعات التي تختلف عدد مفرداتها اختلافاً كبيراً. كذلك صعوبة حسابه من الجداول التكرارية، لاسيما المفتوحة منها. زيادة على هذا كله أن المدى يتمتع بخصائص معينة تشتمل على مزاياه وعيوبه، منها :

1/ يعد مقياساً بسيطاً وسهل الحساب.

2/ يمكن استخدامه بشكل واسع لقياس التشتت في العينات الصغيرة.

3/ يتأثر وينحاز للقيم الشاذة وبخاصة المتطرفة منها مما يجعله أن يعطي فكرة خاطئة ونتائج مضللة.

4/ من الصعب استخدامه في التوزيعات التكرارية المفتوحة.

بعد هذا كله، نرى أنه من المناسب أن نتعرض لطرائق حساب المدى بشيء من التفصيل وبحسب الآتي :

2) في حالة البيانات غير المبوبة

عندما تكون هنالك مجموعة من القيم، صغيرة العدد ولا تحتاج إلى تبويب، من الممكن حساب البعد بين أكبر قيمة وأصغر قيمة بين هذه المجموعة من القيم. عن طريق حساب المدى لها. وعلى هذا سيكون المدى، عبارة عن الفرق بين اكبر القيم وأصغرها، أي إن :

المدى = أكبر قيمة – أصغر قيمة

تمرين 30: أوجد المدى للبيانات

(3، 5، 7، 12، 19، 27، 29، 34، 38)

الحل :

- ❖ نحدد أعلى قيمة في المجموعة بعد ترتيب القيم تصاعدياً، وهي (38) وكذلك أدنى قيمة. وهي (3).
- ❖ نطبق المعادلة المشار إليها أنفاً، فتكون قيمة المدى على النحو الآتي :

المدى = أعلى قيمة – أدنى قيمة

المدى = 38 – 3

المدى = 35

3) في حالة البيانات المبوبة

إن المدى للبيانات المبوبة، هو الفرق ما بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الأولى. ويراه بعض الإحصائيين أنه الفرق بين مركز الفئة الأخيرة من التوزيع التكراري ومركز الفئة الأولى منه. وعلى هذا يمكن حسابه من العلاقة الآتية :

المدى = الحد الأعلى للفئة العليا – الحد الأدنى للفئة الدنيا

تمرين 31: جد المدى المطلق لهذا التوزيع؟

في توزيع تكراري لعدد سنوات ممارسي الألعاب الرياضية في نادي رياضي اجتماعي

جدول 24: سنوات ممارسي الألعاب الرياضية في نادي رياضي اجتماعي

الفئة	- 5	- 10	- 15	- 20	- 25	- 30	40-35	**المجموع**
التكرار	4	9	12	17	8	5	2	**57**

الحل :

- ❖ نوجد قيم الحد الأدنى للفئة الأخيرة، والحد الأدنى للفئة الأولى من جدول التوزيع التكراري.

قيمة الحد الأعلى للفئة الأخيرة = 40

قيمة الحد الأدنى للفئة الأولى = 5

- ❖ نحسب قيمة المدى من المعادلة آنفة الذكر.

المدى = 40 – 5 = 35

تمرين 32: جد تشتت القيم

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

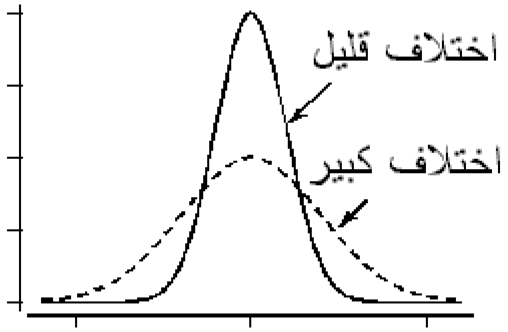
جدول 25: شكل انتشار القيم

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.

تمرين 33

الشكل التالي يوضح المضلعين التكراريين لمجموعتين من البيانات لهما نفس مقاييس النزعة المركزية ولكنهما مختلفتان في طبيعة التشتت.



الشكل 12: 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:

1. المدى: Range
2. نصف المدى الربيعي Semi-Inter-quartile Range
3. التباين: Variance
4. الانحراف المعياري: Standard Deviation
5. معامل الاختلاف (أو التغير): Coefficient of Variation

4) حسابات المدى: Range

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

1.4. تعريف (1):

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

$$Range = X_{max} - X_{min} \qquad (1)$$

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

تمرين 34: المدى لعدد من المشاهدات

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

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

الحل:

$$X_{max} = 55$$

$$X_{min} = 25$$

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

تمرين 35: المدى لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا

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

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

5) 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.

ملاحظة (2):

1. وحدة المدى هي نفس وحدة البيانات الأصلية.
2. نظرًا لأن المدى يعتمد فقط على أكبر وأصغر قيمة ولا يأخذ في الاعتبار القيم الأخرى فهو مقياس غير جيد لقياس التشتت.

III. نصف المدى الربيعي Semi-Inter-quartile Range

1) تعريف المدى 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 نصف المدى الربيعي. وحيث أن القيم المتطرفة هي تلك القيم الصغيرة جدًا أو الكبيرة جدًا فإنه عند حساب نصف المدى الربيعي لا يؤخذ في الاعتبار ربع البيانات الصغيرة (25%) ولا ربع البيانات الكبيرة (25%). الشكل التالي يوضح موضع القيم الشاذة والمتطرفة.

```
    ¼ n = 25%                                              ¼ n = 25%
-/-/-/-/-/-/-/-/-/-/-/-/-/-|----------------|----------------|-/-/-/-/-/-/-/-/-/-/-/-/-/-
ربع البيانات الصغيرة  Q1             Q2              Q3  ربع البيانات الكبيرة
```

يرمز لنصف المدى الربيعي بالرمز Q ويعرف بالصيغة التالية:

$$Q=\frac{Q_3-Q_1}{2} \qquad (2)$$

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

تمرين 36: أوجد نصف المدى الربيعي لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين
تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (2) من الفصل الثاني باستخدام:

(أ) الطريقة الحسابية.

(ب) الطريقة البيانية.

الحل:

(أ) الطريقة الحسابية:

حساب الربيع الأول Q_1:

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

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

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

2) 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 نصف المدى الربيعي هو:

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

جدول 26: معطيات نصف المدى الربيعي

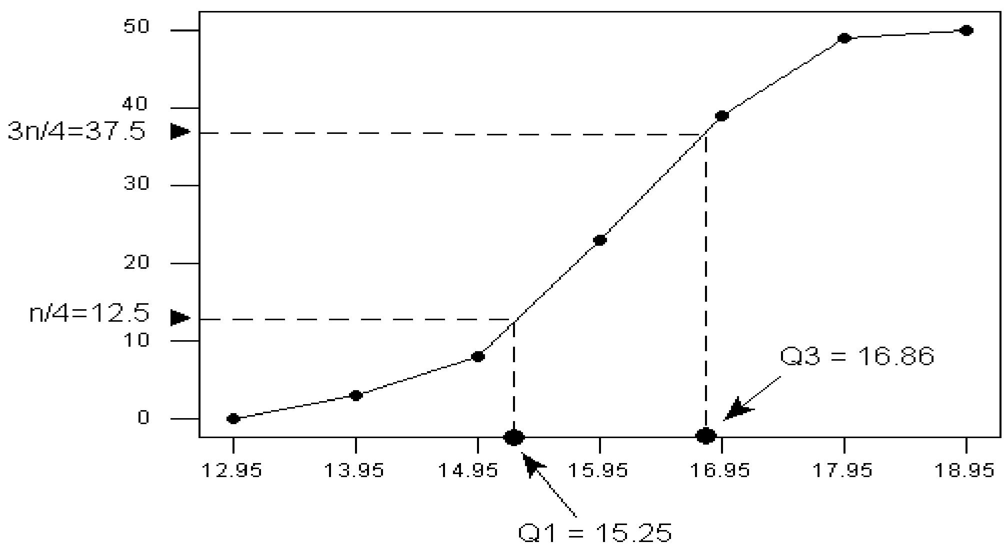
	التكرار المتجمع الصاعد	مستوى الهيموجلوبين	
	0	أقل من	12.95
	3	أقل من	13.9 5
$R = \frac{n}{4} = 12.5$	$8 = F_1$	أقل من	14. 95= A
$R^* = \frac{3n}{4} = 37.5$	$23 = F_2 = F^*_1$	أقل من	$15.95 = A^*$
	$39 = F^*_2$	أقل من	16. 95
	49	أقل من	17. 95
	50	أقل من	18.95

$Q_1 \Rightarrow$

$Q_3 \Rightarrow$

(ب)الطريقة البيانية:

الشكل 13: معطيات نصف المدى الربيعي



وباستخدام الصيغة فإن نصف المدى الربيعي هو:

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

3) بعض مميزات وعيوب نصف المدى الربيعي

1. من المميزات أنه لا يتأثر بالقيم الشاذة أو المتطرفة.
2. من العيوب أنه لا يأخذ في الاعتبار جميع البيانات.

ملاحظة (3):

وحدة نصف المدى الربيعي هي نفس وحدة البيانات الأصلية.

4) 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.

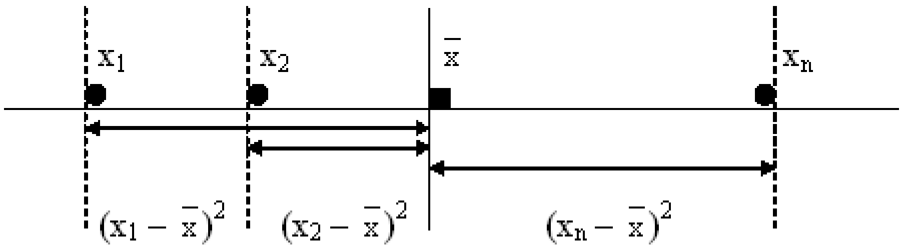
IV. التباين: Variance

1) تعريف التباين

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 14: مربعات انحرافات القيم عن الوسط الحسابي.



جدول 27: طريقة حساب انحرافات ومربعات انحرافات القيم عن المتوسط.

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

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.

V. الانحراف المعياري: Standard Deviation

1) تعريف الانحراف المعياري

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.

2) حساب التباين والانحراف المعياري:

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

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

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

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

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:

1. $s^2 \geq 0$ (دائمًا) وكذلك $s \geq 0$ (دائمًا).
2. $s^2 = 0 \Leftrightarrow s = 0 \Leftrightarrow$ 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:

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

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

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.

تمرين 37: أوجد تباين العينة والانحراف المعياري

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:

الشكل 15: الانحراف المعياري لمجموع الأوزان (كغ)

x^2	مربع الانحراف $(x_i-\bar{x})^2$	الانحراف $(x_i-\bar{x})=(x_i-5.16)$	x_i
50.41	3.7636	1.94	7.1
6.25	7.0756	−2.66	2.5
6.25	7.0756	−2.66	2.5
29.16	0.0576	0.24	5.4
68.89	9.8596	3.14	8.3
$\sum_{i=1}^{n} x_i^2 = 160.96$	$\sum_{i=1}^{n}(x_i-\bar{x})^2 = 27.832$	**0.00**	$\sum_{i=1}^{n} x_i = 25.8$

من هذا الجدول نوجد الكميات التالية:

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

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:

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

(ب) باستخدام الصيغة الحسابية الأولى:

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

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

(ج) باستخدام الصيغة الحسابية الثانية:

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

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

الانحراف المعياري هو:

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

4) بعض خصائص التباين والانحراف المعياري:

1. يخضع التباين والانحراف المعياري لبعض العمليات الجبرية. فإذا كان s^2 و s هما على الترتيب التباين والانحراف المعياري للبيانات $x_1, x_2, \ldots, x_n$ وكان a و b مقدارين ثابتين، فإن:

أ- تباين البيانات $x_1 \pm b, x_2 \pm b, \cdots, x_n \pm b$ هو s^2 وأما انحرافها المعياري فهو S. لذلك فإن التباين والانحراف المعياري لا يتأثران بإضافة أو طرح مقدار ثابت من جميع المشاهدات.

ب- تباين البيانات $ax_1, ax_2, \cdots, ax_n$ هو $a^2 s^2$ وأما انحرافها المعياري فهو $|a|s$. حيث $|a|$ هي القيمة المطلقة للقيمة a.

ج- تباين البيانات $ax_1 \pm b, ax_2 \pm b, \cdots, ax_n \pm b$ هو $a^2 s^2$ وأما انحرافها المعياري فهو $|a|s$.

د- تباين المقدار الثابت يساوي الصفر.

جدول 28: بعض خصائص التباين والانحراف المعياري

المشاهدات	التباين	الانحراف المعياري
$x_1, x_2, \ldots, x_n$	s^2	s
$x_1 \pm b, x_2 \pm b, \cdots, x_n \pm b$	s^2	s
$ax_1, ax_2, \cdots, ax_n$	$a^2 s^2$	$\lvert a \rvert s$
$ax_1 \pm b, ax_2 \pm b, \cdots, ax_n \pm b$	$a^2 s^2$	$\lvert a \rvert s$

تمرين 38:

	المشاهدات	الانحراف المعياري	التباين
$: x$	2, 6, 4, 3, 5	$s = 1.581$	$s^2 = 2.5$
$: x+5$	7, 11, 9, 8, 10	1.581	2.5
$: 3x$	6, 18, 12, 9, 15	$\lvert 3 \rvert \times 1.581 = 4.743$	$9 \times 2.5 = 22.5$
$: 3x+5$	11, 23, 17, 14, 20	$\lvert 3 \rvert \times 1.581 = 4.743$	$9 \times 2.5 = 22.5$

تمرين 39:

إذا كان التباين للمشاهدات $x_1, x_2, \ldots, x_n$ هو 36 فإن التباين للمشاهدات $\frac{x_1 - 10}{2}, \frac{x_2 - 10}{2}, \ldots, \frac{x_n - 10}{2}$ هو $\left(\frac{1}{2}\right)^2 36 = \frac{36}{4} = 9$ وأما الانحراف المعياري فهو $\sqrt{9} = 3$.

2. إذا كان لدينا مجموعتان من البيانات بحيث أن عدد بيانات المجموعة الأولى n_1 ومتوسطها $\bar{x}_1$ وتباينها s_1^2 وكان عدد بيانات المجموعة الثانية n_2 ومتوسطها $\bar{x}_2$ وتباينها s_2^2 وإذا كان $\bar{x}_1 = \bar{x}_2$ (أي أن متوسطي المجموعتين متساويان) فإن تباين المجموعة الكلية المكونة من دمج هاتين المجموعتين يمكن حسابه بالصيغة التالية:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 1}$$

تمرين 40: جد تباين المجموعة الكلية المكونة من دمج مجموعتين

جدول 29: تباين المجموعة الكلية المكونة من دمج المجموعتين التاليتين

	المجموعة الأولى	المجموعة الثانية
حجم العينة	$n_1 = 4$	$n_2 = 6$
المتوسط	$\bar{x}_1 = 5$	$\bar{x}_2 = 5$
التباين	$s_1^2 = 3$	$s_2^2 = 3.5$

الحل: نلاحظ أن متوسطي المجموعتين متساويان.

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 1}$$
$$= \frac{(4-1)(3) + (6-1)(3.5)}{4+6-1} = 2.944$$

5) 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:

$$s^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{n-1} \qquad (7)$$

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:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{k} x_i^2 f_i - \left(\sum_{i=1}^{k} x_i f_i\right)^2 / n\right\} \qquad (8)$$

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{k} x_i^2 f_i - n(\bar{x})^2\right\} \qquad (9)$$

جدول 30: ملخص لعمليتي إيجاد المتوسط والتباين

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	$x_k f_k$	$x_k^2 f_k$	$f_k(x_k-\text{ر})^2$
المجموع		$\sum f = n$	$\sum xf$	$\sum x^2 f$	$\sum_{i=1}^{k} f_i(x_i-\bar{x})^2$

تمرين 41: جد التباين والانحراف المعياري لمستوى الهيموجلوبين في الدم لعينة

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:

جدول 31: ملخص إيجاد التباين

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

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

حساب التباين باستخدام صيغة التعريف:

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

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:

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

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

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

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

حساب الانحراف المعياري:

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

VI. 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 $\bar{x}$ وانحرافها المعياري s بالصيغة التالية:

$$C.V. = \frac{s}{\bar{x}} \qquad (10)$$

تمرين 42: جد البيانات الأكثر تشتتًا نسبًا (أقل تجانسًا)؟

الجدول أدناه يتضمن أي البيانات أكثر تشتتًا نسبًا (أقل تجانسًا)؛ بيانات الأوزان أم بيانات الأطوال؟

جدول 32: بيانات إحدى الدّراسات التي طبقت على خمسة أشخاص لقياس الوزن والطول

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

الحل: أولاً نوجد المتوسط $\bar{x}$ والانحراف المعياري s لكل من بيانات الأوزان وبيانات الأطوال كما مر معنا سابقاً.

جدول 33: متوسط و انحراف بيانات الأوزان و الأطوال

البيانات	المتوسط $\bar{x}$	الانحراف المعياري s	معامل الاختلاف $C.V. = \frac{s}{\bar{x}}$
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.

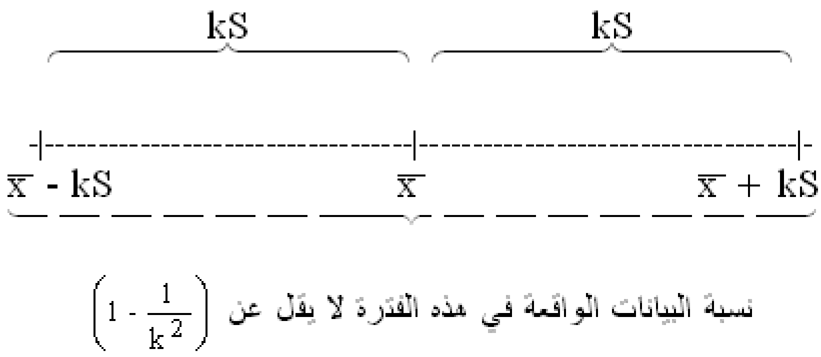
VII. نظرية اللامساواة (متراجحة) تشيبيشيف *Chebychev Inequality*

1) تعريف نظرية اللامساواة

إن نظرية ***تشيبيشيف*** من النظريات المفيدة إذا أنها تعطينا حدًا أدنى لنسبة البيانات الواقعة في فترة معينة عند معرفة متوسط البيانات وانحرافها المعياري دون الحاجة لمعرفة البيانات الأصلية أو التوزيع الذي أخذت منه العينة.

نظرية (1): إذا كان لدينا عينة من البيانات متوسطها $\bar{x}$ وانحرافها المعياري s فإن نسبة البيانات الواقعة في الفترة $(\bar{x}-ks,\bar{x}+ks)$ لايقل عن $\left(1-\frac{1}{k^2}\right)$ حيث أن $k>1$.

الشكل التالي يوضح فكرة نظرية تشيبيشيف.



ملاحظة:

1. في بعض الأحيان نكتب الفترة $(\bar{x}-ks,\bar{x}+ks)$ على الصورة $\bar{x}\pm ks$.
2. تطبق نظرية تشيبيشيف للفترات التي منتصفها (مركزها) هو المتوسط $\bar{x}$.
3. تستخدم نظرية تشيبيشيف بطريقتين(في كلا الحالتين لابد من معرفة قيمة k):
   - أ- تحديد النسبة (التقريبية) لعدد البيانات الواقعة في فترة معينة.
   - ب- تحديد فترة يقع فيها ما لا يقل عن نسبة معينة.

تمرين 43: جد نسبة البيانات الو اقعة في محال معيّن؟

إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x}=7$ وانحرافها المعياري s=5 فما هي نسبة البيانات الواقعة في الفترة $(-4,18)$؟

الحل:

أولاً نلاحظ أن منتصف الفترة المعطاة $(-4,18)$ هو المتوسط $\bar{x}=7$ لذلك نستطيع تطبيق نظرية تشيبيشيف. والآن:

$$(\bar{x}-ks,\bar{x}+ks)=(-4,18) \quad \Rightarrow \quad \bar{x}+ks=18$$

$$\Leftrightarrow \quad 7+k(5)=18$$

$$\Leftrightarrow \quad 5k=11$$

$$\Leftrightarrow \quad k = 11/5$$

$$\Leftrightarrow \quad \left(1 - \frac{1}{k^2}\right) = \left(1 - \frac{1}{(11/5)^2}\right)$$

$$\Leftrightarrow \quad \left(1 - \frac{1}{k^2}\right) = 0.7934$$

لذلك فإن نسبة البيانات الواقعة في الفترة (4,18−) لا تقل عن 79.34%.

تمرين 44: جد نسبة البيانات الو اقعة في محال معيّن؟

إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x} = 7$ وانحرافها المعياري s =5 فأوجد فترة يقع فيها ما لا يقل عن 75% من البيانات.

الحل:

$$\left(1 - \frac{1}{k^2}\right) = 0.75 \quad \Leftrightarrow \quad \frac{1}{k^2} = 1 - 0.75$$

$$\Leftrightarrow \quad \frac{1}{k^2} = 0.25$$

$$\Leftrightarrow \quad k = \sqrt{\frac{1}{0.25}}$$

$$\Leftrightarrow \quad k = 2$$

وبالتالي فإن الفترة التي يقع فيها ما لا يقل عن 75% من البيانات هي:

$$(\bar{x} - ks, \bar{x} + ks) = (7 - 2\times5, 7 + 2\times5)$$

$$= (7 - 10, 7 + 10)$$

$$= (-3, 17)$$

VIII. الدرجات (القيم) المعيارية : Standard Scores (Values)

1) تعريف القيم المعيارية

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

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

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

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

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.

تمرين 45: حد الدرجة المعيارية لقيمة معينة؟

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:

$$x=\bar{x}+s\,z=7+5\times0.1=7+0.5=7.5$$

تمرين 46: جد الدرجات المعيارية للمعطيات؟

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:

جدول 34: الدرجات المعيارية

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.

IX. أمثلة الفصل الرابع باستخدام حزمة "*إكسل*"

تمرين 47: جد المدى للمشاهدات المعطاة؟

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

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

الحل:

جدول 35: إدخال البيانات في صفحة من إكسل

=MAX(A2:A8) - MIN(A2:A8)

	A	B	C	D	E	F	G	H	I	J	K	L
1	الوزن											
2	25	المدى =	30									
3	30											
4	40											
5	45											
6	35											
7	55											
8	50											

لحساب المدى استخدمنا التعريف $R = MAX(x_1,x_2,...,x_n) - MIN(x_1,x_2,...,x_n)$ لاحظ اننا أدخلنا المشاهدات على شكل مجال A2:A8.

تمرين 48: جد المدى لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا؟

تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (2) من الفصل الثاني.

الحل:

جدول 36: نتائج محصّلة بالطرح المباشر

	A	B	C	D	E	F
1	مستوى الهيموجلوبين	مركز الفترة	التكرار			
2	12.95 – 13.95	13.45	3	المدى =	5	من مركز الفئات
3	13.9 5– 14.95	14.45	5	أو المدى =	6	من حدود الفئات
4	14. 95– 15.95	15.45	15			
5	15. 95– 16.95	16.45	16			
6	16. 95– 17.95	17.45	10			
7	17. 95– 18.95	18.45	1			

الحل بطريقة اخرى:

يتميز إكسل بمقدرته على التعامل مع عدد كبير جدا من المشاهدات والتي لانحتاج إلى تلخيصها في جدول تكراري إلا إذا كنا نحتاج ذلك لغرض إيجاد توزيع المشاهدات. لذلك من الأفضل التعامل مع المشاهدات كما هي في إكسل.

جدول 37: المدى لمستوى الهيموجلوبين من المشاهدات 50 مباشرة

=MAX(A2:A51) - MIN(A2:A51)

	A	B	C	D	E	F	G	H	I	J	K
1	مستوى الهيموجلوبين										
2	17.0	المدى =	4.8								
3	17.7										
4	15.9										

وهذه النتيجة أدق من النتيجة السابقة.

تمرين 49: جد نصف المدى الربيعي لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا.

الحل:

نستخدم الصيغة: =(QUARTILE(A2:A51,3) - QUARTILE(A2:A51,1))/2

جدول 38: المدى الربيعي من خلال الصيغة الرياضية المباشرة

=(QUARTILE(A2:A51,3) -QUARTILE(A2:A51,1))/2

	A	B	C	D	E	F	G	H	I	J	K
1	مستوى الهيموجلوبين										
2	17.0	نصف المدى الربيعي =		0.65							
3	17.7										

تمرين 50: جد تباين العينة والانحراف المعياري لمجموعة الأوزان (بالكيلوجرام)

7.1, 2.5, 2.5, 5.4, 8.3

الحل:

التبـاين يحسـب بالصـيغة $=VAR(x_1,x_2,...,x_n)$ حيـث تـدخل المشـاهدات على شـكل مجـال A2:A51 ويحسـب الإنحراف المعياري من الصيغة $=STDEV(x_1,x_2,...,x_n)$ حيث تدخل المشاهدات على شكل مجال A2:A51 ايضا.

ملاحظـة: بعـد حسـاب التبـاين يمكـن حسـاب الإنحـراف المعيـاري بأخـذ الجـزر التربيعي بإسـتخدام الصـيغة =SQRT(Number) حيث Number هو قيمة التباين.

جدول 39: التباين يحسب بالصيغة

	A	B	C
1	الوزن		
2	7.1	التباين =	=VAR(A2:A51)
3	2.5	الإنحراف المعياري =	=STDEV(A2:A51)
4	2.5	الإنحراف المعياري	
5	5.4	من التباين =	=SQRT(C2)
6	8.3		

تمرين 51: جد التباين والإنحراف المعياري؟

حساب التباين والانحراف المعياري لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا.

الحل: من المشاهدات 50 نوجد التباين والإنحراف المعياري كما في المثال السابق.

جدول 40: حساب التباين والإنحراف المعياري

	A	B	C
1	مستوى الهيمجلوبين		
2	17	التباين =	=VAR(A2:A51)
3	17.7	الإنحراف المعياري =	=STDEV(A2:A51)
4	15.9		
5	16.2		

	A	B	C
1	مستوى الهيمجلوبين		
2	17.0	التباين =	1.216526531
3	17.7	الإنحراف المعياري =	1.102962615
4	15.9		

تمرين 52: جد البيانات الأكثر تشتتًا نسبًا؟

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؟

جدول 41: بيانات الأوزان والأطوال

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

الحل: لحساب معامل الإختلاف نستخدم الصيغة

$=STDEV(x_1,x_2,\ldots,x_n)/AVERAGE(x_1,x_2,\ldots,x_n)$

جدول 42: حساب معاملات الأطوال والأوزان

	A	B	C	D	E
1	رقم الشخص	الوزن	الطول		
2	1	69	164	معامل الإختلاف للوزن =	=STDEV(B2:B6)/AVERAGE(B2:B6)
3	2	59	162	معامل الإختلاف للطول =	=STDEV(C2:C6)/AVERAGE(C2:C6)
4	3	65	155		
5	4	67	165		
6	5	65	158		

	A	B	C	D	E
1	رقم الشخص	الوزن	الطول		
2	1	69	164	معامل الإختلاف للوزن =	0.05756396
3	2	59	162	معامل الإختلاف للطول =	0.026163786
4	3	65	155		
5	4	67	165		
6	5	65	158		

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.

X. تمارين

1) بيانات ألوان عينة لنوع من الزهور

حمراء، بيضاء، صفراء، زرقاء، زرقاء، بيضاء، بيضاء، بيضاء، خضراء، حمراء، صفراء، بيضاء، زرقاء، حمراء، صفراء، خضراء، زرقاء، خضراء، خضراء، بيضاء، زرقاء، صفراء، زرقاء، بيضاء، حمراء، خضراء، خضراء، صفراء، حمراء، خضراء، حمراء، حمراء، حمراء، زرقاء، زرقاء، صفراء، زرقاء، صفراء، صفراء، خضراء، صفراء، حمراء، حمراء، خضراء، بيضاء، خضراء، بيضاء، بيضاء صفراء، خضراء

أ. أي من مقاييس التشتت يستخدم مع هذه البيانات؟

ب. أوجد مقياس أو مقاييس تشتت للبيانات إذا وجد.

ج. هل يمكن حساب مدى لهذه البيانات.

2) بيانات أطوال طلاب أحد شعب 101 إحص (لأقرب سم):

168	169	170	167	168	166	170	167	169	166	166	168	168	170	166	167
167	167	167	169	167	170	169	169	167	166	168	170	168	166	169	169
166	166	169	166	169	167	167	167	170	169	170	170	168	170	168	168
169	170														

أ. أي من مقاييس التشتت يستخدم مع هذه البيانات؟

ب. أوجد مقياس أو مقاييس تشتت للبيانات.

ج. أوجد معامل الإختلاف والدرجات المعيارية للبيانات وأوجد فترة يقع فيها ما لا يقل عن 75% من البيانات.

د. أستخدم إكسل لحل الفقرات (ب و ج).

3) تصنيف نباتات عينة من خمسة أنواع من النباتات في رحلة برية

حين فرزها وجد التالي:

الصنف	الأول	الثاني	الثالث	الرابع	الخامس
العدد (نبتة)	20	35	15	25	10

أ. أي من مقاييس التشتت يستخدم مع هذه البيانات؟

ب. أوجد مقياس أو مقاييس تشتت للبيانات.

ج. أحسب معامل الاختلاف إذا وجد.

4) **تجارب طبيب صيدلي لأحد العقاقير على عينة من الفئران**

سجل النتائج التالية:

الشكل 16: تجربة عقاقير على عيّنة من الفئران

مجال الوزن (جرام)	100 - 51	150 -101	200 - 151	250 - 201
التكرار (عدد الفئران)	20	25	15	10

أ. أي من مقاييس التشتت يستخدم مع هذه البيانات؟

ب. أوجد مقياس أو مقاييس تشتت للبيانات.

ج. أحسب معامل الاختلاف والدرجات المعيارية إذا وجد أي منها.

د. أستخدم إكسل لحل الفقرات (ب و ج).

5) **تقديرات 100 طالب في شعبتين من مادة 101 إحص.**

B B D B B E E C A F D F C B F C A B B E

B B D D B D B D A D D E D D D C A B F C

F B B A B B E E D E B B D B B E E C A F

D F C B F C A B B E B B D D B D B D A D

D E D D D C A B F C F B B A B B E E D E

أ. أي من مقاييس التشتت يستخدم مع هذه البيانات؟

ب. أوجد مقياس أو مقاييس تشتت للبيانات.

ج. هل يمكن حساب درجات معيارية لهذه البيانات.

6) **أوزان 50 سمكة (بالكيلوجرام)**

2.5 4.0 1.5 1.5 2.5 3.0 4.0 2.5 2.5 2.0

2.5 1.5 1.5 2.5 4.0 1.5 3.0 2.0 3.5 1.5

2.5 1.0 3.0 2.0 1.5 2.5 3.5 4.0 2.5 3.0

3.5 1.5 2.5 3.5 2.0 1.5 3.5 2.5 2.5 2.5

2.0 2.5 3.5 3.5 2.0 2.5 2.0 2.0 1.5 1.5

المطلوب: بإستخدام حزمة "إكسل" أوجد:

أ. المدى.

ب. نصف المدى الربيعي.

ج. التباين.

د. الإنحراف المعياري.

هـ. معامل الإختلاف.

و. الدرجات المعيارية.

7) سجل أحد الطلاب في 6 مواد لفصل دراسي

الشكل 17: عدد الساعات لكل مادة وتقدير الطالب في كل مادة

المادة	1	2	3	4	5	6
عدد الساعات	2	3	3	5	1	2
الدرجات	95	83	75	80	70	90

أ. أي من مقاييس التشتت يمكن حسابه من هذه البيانات؟

ب. أحسب على الأقل ثلاثة من مقاييس التشتت إذا وجد.

ج. احسب معمل الاختلاف إذا وجد.

د. هل يمكن حساب درجات معيارية للبيانات؟ وأحسبها إذا أمكن ذلك.

هـ. حل الفقرات (ب) و (ج) و (د) بإكسل إذا أمكن ذلك.

8) البيانات التالية عبارة عن الرواتب الشهرية لمنسوبي أحد الشركات بآلاف الريال:

14 11 1 6 12 7 3 8 11 6 12 7 2 13 8 2 15 9 3 7

المطلوب: أي مقاييس تشتت تمثل هذه البيانات بشكل عادل؟

الفصل الخامس:

مقاييس الالتواء والتفرطح

Flattening and kutorsis measurements

مقدمة

رأينا في الفصل الرابع كيف أن مقاييس النزعة المركزية تعطينا نظرة عامة عن شكل توزيع مجموعة من البيانات، كما أن المدرج التكراري الذي نحصل منه على المنحنى التكراري يعتبر وسيلة الاستدلال السريع على شكل توزيع المعطيات، إذ يُمكن لتوزيع متغير إحصائي أن يكون متماثل أو ملتويا إما إلى ناحية اليمين أو إلى ناحية اليسار كما يمكنه أن يكون أكثر أو أقل تفرطحا ٍ
للتعبير عن مختلف هذه الوضعيات و لدراسة شكل توزيع ما من إلتواء أو تفلطح نستعين إذن بمعاملات و مقاييس تزيد من دقة معرفتنا بذلك.

فبهذا يمكننا القول أن مقاييس الشكل هي تلك المقاييس التي تبين شكل التوزيع الاحصائي من إلتواء أو تفلطح مقارنة بتوزيع مرجعي[1]؛ و تعتمد هذه المقاييس على فكرة العزوم.

I. العزوم: Moments

1) تعريف العزوم

العزوم هي مصطلح فيزيائي يُستعمل في علم الميكانيك و قد إستعمله الاحصائيون بمعنى قريب من المعنى الفيزيائي؛ و العزم يقيس تأثير القيم)أو مراكز الفئات(أو قواها موزونة بتكراراتها النسبية في الانسحاب عن نقطة الاصل أو مركز معين، و يُرمز له بالرمز m.

2) مقاييس التشتت النسبي

إنّ مقاييس التشتت النسبي هي مقاييس خالية من وحدات القياس لذلك يتم استعمالها للمقارنة بين تشتت مجموعتين او اكثر تختلف في وحدات قياسها. ومن اهم مقاييس التشتت النسبي هو : coefficient of variationمعامل الاختلاف يعبر عن معامل الاختلاف كالاتي:

$$c.v = \frac{s}{\bar{x}} \times 100$$

حيث ان:

S: الانحراف المعياري للعينة

$\bar{x}$: الوسط الحسابي للعينة

تمرين 53: جد نتائج الامتحانات النهائية لمقياس الإحصاء و المنهجية لطلبة الماستر تنظيم والعمل؟

جدول 43: نتائج الامتحانات النهائية لمقياس الإحصاء و المنهجية لطلبة الماستر تنظيم والعمل

	الإحصاء	المنهجية
الوسط الحسابي	16	12
الانحراف المقياسي	8	4

المطلوب: في أي الموضوعين كانت تشتت الدرجات ؟

الحل:

$$c.v = \frac{s}{\bar{x}} \times 100$$

$$c.v = \frac{8}{16} \times 100 = 50.00\%$$

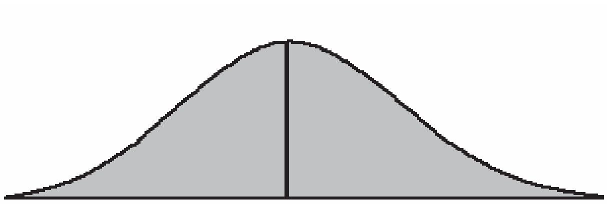
$$c.v = \frac{4}{12} \times 100 = 33.33\ \%$$

نلاحظ ان التشتت لدرجات الاحصاء كان اقل لذلك فان اداء الطلبة في امتحان الاحصاء كانت اكثر تقاربا.

II. مقاييس الالتواء والتفرطح

1) التوزيع الطبيعي

عند تمثيل بيانات الظاهرة في شكل منحني تكراري، فان المنحني يأخذ اشكالا مختلفة ، فقد يكون هذا المنحني متماثل بمعنى ان له قمة في المنتصف ، ولو اسقطنا عمودا من قمته على المحور الافقي لشطره نصفين متماثلين ، مثل منحني التوزيع الطبيعي ، كما مبين بالشكل التالي:



منحنى 1: التوزيع الطبيعي (منحني متماثل)

وعندما يكون الشكل متماثل ، فان الوسط والوسيط والمنوال كلهم يقعون على نقطة واحدة ، ولكن في كثير من الحالات يكون هناك قيم كبيرة في البيانات تجذب اليها الوسط الحسابي ، وهذا معناه ان المنحني التكراري سوف يكون له ذيل جهة اليمين ، مشيرا بوجود التواء جهة اليمين ، وكذلك العكس لو ان البيانات بها قيم صغيرة ، فإنها تجذب الوسط اليها ، ويدل المنحني التكراري على وجود التواء جهة اليسار ، كما يمكن من خلال الشكل البياني معرفة ما اذا كان توزيع البيانات منبسط ، او مدبب ، وهذا من الناحية البيانية ، الا ان هناك مقاييس كثيرة لوصف البيانات تعتمد في حسابها على مقاييس النزعة المركزية والتشتت معا ، ومنها مقاييس الالتواء والتفلطح.

2) مقاييس الالتواء

طريقة "بيرسون Person " في قياس الالتواء

توجد طرائق لقياس الالتواء ومنها ما يلي:

تأخذ هذه الطريقة في الاعتبار العلاقة بين الوسط والوسيط والمنوال ، في حالة ما اذا كان التوزيع قريب من التماثل وليس شديد الالتواء ، وهذه العلاقة هي:

المنوال = الوسط الحسابي -3 (الوسط الحسابي – الوسيط)

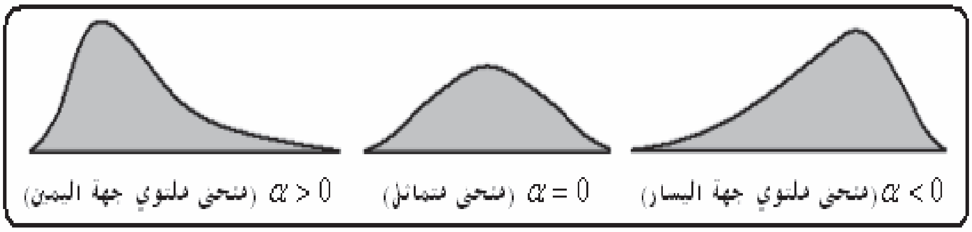
ومن ثم فان طريقة "بيرسون" في قياس الالتواء ، تتحدد بالمعادلة التالية.

$$\alpha = \frac{3(Mean - Median)}{Stan\ dard\ Deviation} = \frac{3(\bar{x} - Med)}{s}$$

حيث ان α (الفا) هو معامل الالتواء "لبيرسون " ، $\bar{x}$ *الوسط الحسابي* ، *Med* هو الوسيط ، S هو الانحراف المعياري ، ويمكن من خلال الإشارة التي يأخذها هذا المعامل الحكم على شكل الالتواء ، كما يلي :

- اذا كان (الوسط الحسابي = الوسيط) كان قيمة المعامل ($\alpha=0$) ويدل ذلك على منحني التوزيع التكراري متماثل.
- اذا كان (الوسط الحسابي > الوسيط) كان قيمة المعامل ($\alpha > 0$) ،ويدل ذلك على ان منحني التوزيع التكراري ملتوي جهة اليمين.
- اذا كان (الوسط الحسابي < الوسيط) كان قيمة المعامل ($\alpha < 0$) ،ويدل ذلك على ان منحني التوزيع التكراري ملتوي جهة اليسار.

اشكال التواء البيانات



تمرين 54: جد مربعات علامات الطلاب في الاختبار النهائي في مادة الاحصاء؟

الشكل 18: مربعات علامات الطلاب في الاختبار النهائي في مادة الاحصاء

x^2	الدرجة x
4356	66
7225	85
2704	52
6084	78
6400	80
8281	91
5476	74
3364	58
43890	84

كانت درجات 8 طلاب في الاختبار النهائي في مادة الاحصاء

66 85 52 78 80 91 74 58

المطلوب: حساب معامل الالتواء بطريقة "بيرسون"

الحل : حساب معامل الالتواء بطريقة "بيرسون"

- حساب الوسط الحسابي ، والانحراف المعياري:

$$\sum x = 584 \quad , \sum x^2 = 43890$$

$$\bar{x} = \frac{\sum x}{n} = \frac{584}{8} = 74$$

$$s = \sqrt{\frac{\sum x^2 - (\sum x)^2/n}{n-1}} = \sqrt{\frac{43890 - (584)^2 /8}{8-1}}$$

$$\sqrt{\frac{1258}{7}} = \sqrt{179.71428} = 13.406$$

- **حساب الوسيط** : $\frac{n}{2}, \frac{n}{2} + 1$

الوسيط = الوسط الحسابي للقيمتين 74 , 78

$$\frac{74+78}{2} = \frac{152}{2} = 76$$

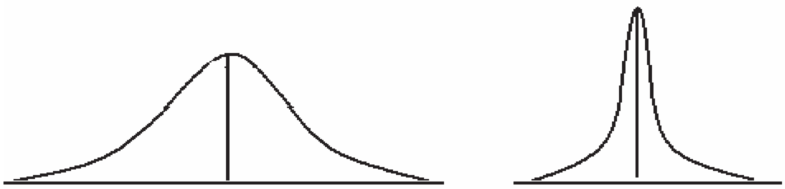
- **معامل الالتواء "بيرسون"**

$$\alpha = \frac{3(\bar{x} - med)}{s} = \frac{3(73-76)}{13.406} = -0.67$$

اذا منحني توزيع درجات الطلاب ملتوي جهة اليسار.

3) مقاييس التفلطح

عند تمثيل التوزيع التكراري في شكل منحني تكراري ، قد يكون المنحني منبسط ، او مدبب ، فعندما يتركز عدد اكبر من القيم بالقرب من منتصف المنحني ، ويقل في طرفية ، يكون المنحني مدببا ، وعندما يتركز عدد اكبر على طرفي المنحني ، ويقل بالقرب من المنتصف يكون المنحني مفرطحا ، او منبسطا، ويظهر ذلك من الشكل التالي:



منحني مفرطح   منحني مدبب

ويمكن قياس التفلطح باستخدام عدد من الطرق ، ومنها طريقة العزوم ،حيث يحسب معامل التفلطح (k) بتطبيق المعادلة التالية :

$$k = \frac{\frac{1}{n}\sum (x-\bar{x})^4}{s^4}$$

حيث ان مقدار $\sum (x-\bar{x})^4/n$ هو العزم الرابع حول الوسط ، s هو الانحراف المعياري . ومعامل التفلطح في التوزيع الطبيعي يساوي 3 ، ومن ثم يمكن وصف منحني التوزيع من حيث التفلطح ، والتدبب كما يلي :

- *اذا كان $k=3$ كان منحني توزيعا معتدلا .*
- *اذا كان $k>3$ كان منحني التوزيع مدببا.*
- *اذا كان $k<3$ كان منحني التوزيع منبسطا (مفرطحا).*

وبالتطبيق على بيانات المثال نجد ان $\bar{x}=73$

جدول 44: قيم التوزيع لمثال التفلطح

x	66	85	52	78	80	91	74	58	584
$(x-\bar{x})$	-7	12	-21	5	7	18	1	-15	0
$(x-\bar{x})^2$	49	144	441	25	49	324	1	225	1258
$(x-\bar{x})^4$	2401	20736	194481	625	2401	104976	1	50625	376246

ومن البيانات أعلاه نجد ان :

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}} = \sqrt{\frac{1258}{7}} = 13.406$$

$$\frac{1}{n}\sum (x - \bar{x})^4 = \frac{1}{8}(376246) = 47030.75$$

اذا كان معامل التفلطح هو

$$K = \frac{47030.75}{(13.406)^4} = \frac{47030.75}{(32299.58)} = 1.456$$

إذا كان شكل توزيع بيانات الدرجات مفرطح.

قائمة المراجع

1. سالم عيسى بدر، عماد غصاب عبابنة، مبادئ الإحصاء الوصفي والاستدلالي، دار المسيرة، الأردن 2007
2. عبد القادر حليمي مدخل إلى الإحصاء، ديوان المطبوعات الجامعية، ط 6 الجزائر، 2009
3. مصطفى خلف عبد الجواد، الاحصاء الاجتماعي المبادئ والتطبيقات، دار المسيرة، الأردن 2013
4. موريس أنجرس منهجية البحث العلمي في العلوم الانسانية، تر مجموعة من الاساتذة، دار القصبة الجزائر 2004
5. موساوي عبد النور، بركان يوسف، الإحصاء ،1 دار العلوم، الجزائر 2009

مخطط الدرس لمقياس الاحصاء الوصفي والاستدلالي للسداسي1

السداسي : الأوّل

اسم الوحدة :وحدة التعليم المنهجية

إسم المـادة :إحصاء وصفي واستدلي (1)

الرصيد: 3

المعامل: 2

أهداف التعليم

ذكر ما يفترض على الطالب اكتسابه من مؤهلات بعد نجاحه في هذه المادة، في (ثلاثة أسطر على اكثر) تزويد الطالب بعموميات حول ا-حصاء الوصفي والاستدلالي واستخداماته في البحوث الاجتماعية. المعارف المسبقة المطلوبة :وصف تفصيلي للمعرف المطلوبة والتي تمكن الطالب من مواصلة هذا التعليم،(سطرين على الاكثر*).*

طريقة التقييم :مراقبة مستمرة، امتحان ...إلخ

- امتحان كتابي + متواصل.

محتوى المادة :إجبارية تحديد المحتوى المفصل لكل مادة (*مع الاشارة إلى العمل الشخصي للطالب*)

- **مدخل مفاهيمي**: التعريف بعلم الإحصاء، التعريف بالإحصاء الوصفي، أهميته، علاقته بالعلوم الأخرى الاجتماعية، أهم المفاهيم الاحصائية.
- **تقنيات عرض البيانات الاحصائية**
- **المصادروتقنيات العرض الجدولي والبياني**
- **مقاييس النزعة المركزية الوسط الحسابي، الوسيط، المنوال،**
- **مقاييس التشتت: المدى، التباين، الانحراف: الربيعي، المعياري، المتوسط**
- **مقاييس الشكل الالتواء، التفلطح**

المراجع: (:كتب، ومطبوعات، مواقع انترنت، إلخ)

- إبراهيم على عبد ربه :الاحصاء الوصفي، الدار الجامعية، اسكندرية، 2005.
- أحمد اشقر:مقدمة في الاحصاء مفاهيم وطرائق، مكتبة دار الثقافة، عمان، 1999.
- أحمد عبادة سرحان :مقدمة في الاحصاء الاجتماعي، دار المعارف، مصر.
- موساوي عبد النور :الاحصاء الوصفي، منشورات جامعة منتوري، الجزائر، .2004