جامعة الشام الخاصة
Al-Sham Private University

كليّة الصيدلة
Faculty of Pharmacy

الفصل الدراسي الثاني
2024

مقرر الرياضيات قسم الإحصاء الحيوي

مدرس المقرر: د. سلطان الصلخدي

الفصل الأول

مقدمة في الإحصاء الوصفي

Introduction to descriptive statistics

1-1-مقدمة:

ظهر الإحصاء، شأنه شأن العلوم الأخرى، من الحاجات العملية للمجتمع، فقد برز وتطور بفعل حاجات الحياة. في البدء استعمل الإحصاء فقط لتعداد السكان إنما فيما بعد اتسعت دائرة نشاطه وأصبح أكثر التصاقاً بالحياة اليومية. فعلى إثر تزايد تعقد حياة الناس وتعمق العلاقات في المجتمع، تصبح التجربة الابتدائية غير كافية. وهذا يدفع بالتالي إلى ضرورة العمل للاستنتاج من تجارب وملاحظات الأفراد. وبالتالي فتاريخ الإحصاء طويل وكبير. سوف نكتفي هنا باستعراض الأحداث التي شكلت المنعطفات الرئيسية في تاريخ هذا العلم، وهي كافية لرسم خط تطوّره والأخذ بغرضه الحالي. على إثر ظهور المجتمع الطبقي والدولة برزت الحاجة الملحّة لمختلف المعلومات المتعلقة بالبلاد، كالسكان ومقومات ثروة الدولة، الخ .... وقد اقتضى الأمر القيام بأعمال التعداد الإحصائي للحصول على هذه المعلومات المختلفة. وإلى مثل هذه الأعمال، يمكن أن نرد نشأة أعمال الإحصاء القديمة، أو بالأحرى تعداد السكان، سيما الذكور القادرين على حمل السلاح منهم. وهناك بالنسبة لهذا الموضوع، معلومات عن تعداد السكان في مصر القديمة يعود للسنة 3500 ق. م. وفي غيرها من مختلف دول عصر الرق. ومع ظهور المجتمع الطبقي والسلطة السياسية نتيجة لذلك، برزت فكرة الجردة شبه الدائمة للسكان والخيرات المتوافرة في البلاد.

وقد تجسّدت آنذاك فكرة الجردة هذه بشكل رئيس في تعداد السكان تلبية للحاجات العملية، الحربية منها والضريبية، لدول عصر الرق. فقد كان لدى الصينيين، فيما بين الألفين الرابع والثاني ق. م، معلومات عن السكان، وكانوا يستعملون جداول إحصائية تتعلق بالزراعة والشيء نفسه بالنسبة للمصريين، الذين عرفوا التعداد الدائم.

وقد أماطت قراءة الكتابة الهيروغليفية اللثام عن النفوذ الضخم لهذا الإحصاء وأثره في حياة البلاد ولغتها. فلم يكن للسنين في الواقع تعبير رقمي شبيه بالذي نعرفه لها اليوم فقد كانت السنون تعرف وتميز، ومنذ عصور السلالات الأولى، بالأحداث الهامة التي وقعت فيها ودمغتها، كالسنة التي انتصر فيها مثلاً الفرعون على الآسيويين أو سنة تدشين ذاك الهيكل لذلك الإله.

فعبارة سنة كان يرمز لها بعلامة ملحقة بالحدث الهام. فيما بعد أصبحت السنوات تسجل بالنسبة للسنة التي جرى فيها إحصاء الضرائب، فيقال مع بدء زمن الإحصاء الثاني للماشية أو الحقول أو الذهب. في البداية كانت هذه العملية

الضريبية تجري مرة كل سنتين. إنما فيما بعد وعلى أثر القيام بها سنوياً، بدءاً من السلالة السادسة، أصبحت كلمة إحصاء مرادفة لكلمة سنة.

هذا ولا بد من الإشارة إلى أن المصريين وضعوا أقدم ميزان اقتصادي عرف في أيامهم وهو ميزان النيل فمستوى ارتفاع فيضان النيل كان مؤشراً ممتازاً للخصب، ويستعمل لتقدير حجم الضرائب. ومن المرجح أنه كان لدى الحضارات العائدة لهذه العصور القديمة، كالآمور والهندوس مثيل هذه الوسائل الإحصائية، التي من المحتمل أن تكشف عنها أعمال الحفريات القائمة.

أما بالنسبة لليونان والرومان فقد تخطوا في أعمالهم التعددية الأغراض الحربية والضريبية ليشملوا غيرها، كتوزيع الأراضي، وتعيين الوضع الاجتماعي للسكان والتموين (مثل خزن بومب يوس للقمح لأجل إطعام 486.000 شخص)، وتوزيع المعونات (مثلاً أعطى أغس طوس في السنة الخامسة ق. م 60 درهماً لكل من المواطنين من العامة والبالغ عددهم 32.000 شخص).

وفي روما القديمة كان الغرض من التعدادات هو تعيين حجم الضريبة المتوجب دفعها للدولة الرومانية. فأثناء القيام بعمليات تشكيل ونقل لوائح التعداد، كان المواطنون الرومان يصرحون، بعد حلفهم اليمين بقولهم الحق، عن قيم ممتلكاتهم وعن أعمالهم ووضعهم العائلي. وقد كانت هذه البيانات كافية لتقسيم السكان حسب الثروة والرتبة العسكرية.

وننهي هذه الفقرة مشيرين إلى أن تطور علم الإحصاء، المرتبط بتطور باقي العلوم، وعبر تداخل مختلف الأحداث الاجتماعية التي تعكس الازدواجية والتناقض لوحدة المجتمع الطبقي الجدلية، وأصبح يعرَّف **علم الإحصاء بأنه العلم الذي يختص بالطرائق العلمية لجمع البيانات وتنظيمها وتلخيصها، وذلك عن طريق التعبير عنها أو عرضها بصورة علمية وتحليلها بغرض الوصول إلى استنتاج النتائج والقوانين التي تحكمها، واتخاذ قرارات سديدة ملائمة لذلك.** وينبغي الإشارة إلى وجود قسمين رئيسيين للإحصاء هما:

1 – الإحصاء الوصفي: ويشمل الطرائق الخاصة بتنظيم البيانات وتلخيصها وعرضها في صورة جداول إحصائية أو رسوم بيانية، أو أشكال هندسية أو تلخيصها، أو حساب مقاييس النزعة المركزية، ومقاييس التشتت وغيرها من المقاييس الأخرى.

2 – الإحصاء الاستنتاجي (الاستدلالي): وهو عبارة عن مجموعة الطرائق العلمية التي تُطبق الاستدلال على المجتمع بناءً على البيانات الإحصائية التي جُمعت وفق طرائق إحصائية محددة، وتشتمل على عدد من المفاهيم والنظريات، مثل نظرية التقدير، واختبار الفرضيات، واختبار جودة الإنتاج.

1 – 2- بعض المفاهيم والمصطلحات الأساسية في الإحصاء:

نعرض في هذا الفصل بعض المفاهيم الأساسية في الإحصاء. يهتم هذا الفرع من الإحصاء باستخدام البيانات المتوفرة لدينا في إصدار أحكام أو تعميمات إحصائية مبنية على الاحتمال حول بيانات غير متوفرة لدينا في المجتمع أو المجتمعات الإحصائية. وقبل البدء بالاستنتاج الإحصائي، لابد وأن نذكر بعض المفاهيم الإحصائية التالية:

1 ــــ المجتمع الإحصائي والعينة: يستند الاستنتاج الإحصائي بصورة جوهرية على البيانات التي يتم الحصول عليها من العناصر ومن خلال هذه البيانات تصاغ التعميمات والاستنتاجات الإحصائية حول جميع العناصر الذين تمثلهم هذه البيانات.

المجتمع الإحصائي: هو أي مجموعة من الأشياء أو الأشخاص التي تشترك فيما بينها بصفة أو عدة صفات، مثل سكان مدينة دمشق، أطباء الجراحة العامة، صيادلة سوريا، مرضى السرطان.

العينة الإحصائية: هي مجموعة جزئية محدودة من المجتمع الإحصائي يتم جمع البيانات من خلالها ويتحكم في طريقة اختيارها وفي حجمها بعض المبادئ الإحصائية والتي سنتعرض لها فيما بعد.

2 ــــــــ الوحدة الإحصائية: هي كل ظاهرة بسيطة أو كائن أو شيء يشترك في صفة أو أكثر تكون موضوع الدراسة الإحصائية على أن تعرف تعريفاً واضحاً وهي أصغر جزء مستقل تجري عليه الدراسة وتجمع البيانات على أساسه. فيمكن أن تكون الوحدة الإحصائية كائناً حياً مثل إنسان في دراسة تعداد السكان أو طائر أو شجرة، وقد تكون شيئاً مثل السيارة في دراسة وسائط النقل أو بنك أو آلة حاسبة، وقد تكون وحدات قياس كالمسافات والمساحات والحجوم والأوزان والقيم.

وللحصول على الوحدة الإحصائية المثلى، ينبغي أن تتصف بالصفات التالية:

1 – أن تكون الوحدة الإحصائية ملائمة لغرض الدراسة.

2 – أن تكون الوحدة الإحصائية ثابتة ومستقرة.

3 – أن يكون مجموع الوحدات الإحصائية يشكل المجتمع الإحصائي.

وتقسم الوحدات الإحصائية بحسب أنواعها إلى قسمين:

آ – **الوحدات البسيطة:** وهي التي تستعمل في عمليات الجمع والتحليل والعرض الإحصائي، حيث تقسم إلى ثلاثة أنواع هي:

- **وحدات العد:** وهي التي تستعمل في عمليات العد، وتشمل الأشياء المادية وصفاتها.
- **وحدات القياس:** وهي التي يمكن قياسها بواسطة الوزن أو الطول أو المقاييس المختلفة.
- **الوحدات النقدية:** كالليرة السورية والدينار والدولار واليورو ....

ب – **الوحدات المركبة:** وهي التي تتكون من مقياس أو أكثر مثل ضغط الدم/ ساعة، الميلي غرام/ السنتيمتر مكعب، الكيلوغرام/المتري، الشخص/الساعي، الكيلو واط/الساعي.

3 – البيانات الإحصائية: تقسم البيانات عادة إلى قسمين أساسيين هما:

1– البيانات الوصفية أو الكيفية Qualitative Date: وهي البيانات الإحصائية التي تصف عناصر الظاهرة المدروسة في صورة غير رقمية، مثل الحالة التعليمية لموظفي الجامعة، لون الشعر أو لون العينين أو تقديرات النجاح في إحدى المواد.

2– البيانات الكمية Quantitative Date: وهي البيانات الإحصائية التي تقاس فيها عناصر الظاهرة المدروسة بالمقاييس الكمية المعروفة، مثل عدد مرضى مشفى المواساة خلال سنة، الأسهم المباعة في سوق الأوراق المالية، أطوال مجموعة طلاب الكليات الطبية، وغيرها من الظواهر الأخرى. وهذه البيانات الكمية يمكن تصنيفها وفق طبيعة الأرقام المستعملة إلى:

آ– **بيانات مستمرة** وهي متغيرات مستمرة يمكن أن تأخذ أية قيمة ضمن فترة محددة ومن ثُمَّ يمكن التعبير عنها بأعداد حقيقية، كالتي تعبر عن الوزن أو الطول أو الحجم أو النسبة.

ب– **بيانات منفصلة** وهي متغيرات منفصلة تأخذ قيماً محددة تماماً، مثل تعداد الكريات الحمر في واحد ميليمتر مكعب من الدم، وعدد المرضى في مشفى المواساة .

4 – الوسيط أو المعلمة: الوسيط هو شيء يميز المجتمع الإحصائي كله مثل متوسط الدخل الشهري للأسر في دولة معينة، أو نسبة الأشخاص المدخنين بصفة دائمة في مجتمع معين، أو نسبة الأدوية الفاسدة في شركة طبية وهكذا ...

5 – الإحصاء: الإحصاء هو شيء يميز العينة الإحصائية مثل متوسط الدخل الشهري لعينة مؤلفة من 200 عامل في شركة أدوية في دولة ما أو متوسط الذكاء لعينة مؤلفة من 100 طفل أعمارهم بين الثامنة والعاشرة وهكذا .

6 – المتغيرات: المتغيرات هي مقادير لها خصائص كمية أو وصفية تتغير من عنصر إلى آخر من عناصر المجتمع الإحصائي أو من العينة الإحصائية.

يتكون المجتمع الإحصائي عادة من مجموعة من الأفراد أو العناصر التي تشترك فيما بينها بخاصة أو أكثر مثل سكان مدينة ما، طلاب جامعة الشام الخاصة، مجموعة الفئران قيد التجارب، مجموعة الأدوية السرطانية ... .

وعند دراسة أحد هذه المجتمعات يجب علينا تحدد الهدف من الدراسة، فقد يكون هدفنا دراسة ظاهرة ما مثل الطول أو الوزن أو الذكاء أو الجنس أو لون العيون أو لون البشرة أو لون الشعر أو قياس مستوى الهيموغلوبين في الدم لدى مجموعة من الأشخاص، قياس السكر في الدم عند مرضى السكري، فإن مفردات الظاهرة لمتغير الطول أو الوزن أو الذكاء أو مستوى الهيموغلوبين في الدم أو كمية السكر في الدم تكون بيانات كمية بينما تكون بيانات وصفية لمتغير الجنس أو لون العيون أو لون البشرة أو لون الشعر. فبعد تحديد المجتمع والظاهرة أو الحالة التي نرغب في دراستها نعرف المتغير العشوائي، وهو الوسيلة الرياضية التي نقرن بها كل فرد من أفراد المجتمع (أحد طلاب الجامعة، مريض السكري... إلى آخره) بقياس عددي (مثل طول شخص، قياس مستوى الهيموغلوبين، قياس السكر في الدم).

بعد تحديد المتغير العشوائي ينصب جهدنا في دراسة وتحليل القيم العددية له أو مجموعة كل الحالات الممكنة وللمتغير العشوائي مجموعة من الصفات أهمها:

1 – العشوائية: تتغير قيمه بشكل عشوائي من عنصر لآخر من عناصر العينة.

2 – لكل فرد من أفراد العينة قيمة وحيدة.

3 – يعرَف على أفراد العينة من المجتمع، ويأخذ قيمة في R مجموعة الأعداد الحقيقية.

4 – لكل متغير عشوائي مدى، وهو مجموعة كل القيم التي يمكن أن يأخذها من R.

فإذا كان الهدف دراسة عدد أفراد أسر طلاب كلية الصيدلة فيكون مدى متغير عدد أفراد الأسرة هو مجموعة الأعداد {1،2، ... ،15} . وإذا كان المتغير يمثل أطوال الأشخاص البالغين في مدينة فإن مداه هو المجال التالي [150، 210] من R. ويكون مدى متغير عيار السكر في الدم هو كذلك مجال، وليكن [50،400] على أن يحوي كل القيم الممكنة للمتغير. والمتغيرات العشوائية نوعان هما:

- المتغيرات العشوائية المنفصلة: هي المتغيرات التي تكون قيمها أعداداً صحيحةً مثل عدد طلاب كلية العلوم، أعداد الكريات الحمراء في حجم محدد من سائل دموي.
- المتغيرات المستمرة: هي التي قيمها قد تكون أي عدد حقيقي من مجال محدد (مدى المتغير) مثل طول شخص بالغ، قياس السكر في الدم ... إلى آخره.

سيكون لقيم المتغيرات العشوائية دلالات مختلفة تعود لطبيعة العينة وللمتغير نفسه الذي يعبر عن الظاهرة المدروسة ، ولهذا سننظر إلى مدى البيانات أو القياسات بمقاييس مختلفة متدرجة في الدقة من مقاييس بسيطة إلى مقاييس عددية ، فالأعداد التي تعبر عن حالات معينة في المجتمع المدروس والمقيسة بسلم بسيط تستخدم فقط للتمييز بين أفراد العينة بينما الأعداد التي تعبر عن كميات تحتاج لمقياس ذو مستوى أعلى نستطيع عندها إجراء جميع العمليات الحسابية عليها ومن ثُمَّ نستطيع استخدام تقنيات إحصائية أكثر تنوعاً في دراسة العينة واستقراء المجتمع.

1 – 3- أنواع سلالم القياس:

من العوامل التي تحدد طريقة تلخيص البيانات وتحليلها نوعية المقياس المستخدم لتلك البيانات، فالمقياس هو استخدام الأرقام في وصف الأحداث، وذلك بناءً على قواعد معينة، وعند تغيير هذه القواعد سوف نحصل على أنواع مختلفة من المقاييس، وعليه فإنه ينبغي مراعاة ما يلي:

1 – القواعد المختلفة التي تستخدم الأرقام بناءً عليها. فمثلاً عند استخدام الأرقام تحت قاعدة التمييز فإن المقياس المستخدم يساعدنا فقط على أن نميز بين شيء وآخر دون تحديد كميته.

2 – الخواص الرياضية للمقياس الناتج عن استخدام الأرقام تحت هذه القواعد.

3 – العمليات الإحصائية المستخدمة في معالجة المقياس الناتج سواءً من حيث بناؤه وتكوينه أم من حيث تحليل نتائج تطبيقاته المختلفة.

يوجد أربعة مستويات لمقاييس القياس في البحث: اسمية، ونسبية، وترتيبية، والفترات .وهي تعرف أيضا باسم المستويات الأربع للقياس، حيث يعتبر كل مقياس منهم ترقية للمقياس الذي يسبقه كما أنه يشمله، وقبل أن نناقش المستويات الأربع من مقاييس القياس بالتفصيل، مع الأمثلة، دعنا نلقي نظرة سريعة ومختصرة عما تمثله هذه المستويات الأربع. المقياس الاسمي هو مقياس مسمِّي، تكون فيه المتغيرات "مسماة" أو "معنونة" ببساطة، بلا ترتيب معين. أما المقياس الترتيبي فتكون كافة متغيراته مرتبة ترتيب معين، يتعدى مجرد تسميتهم. بينما تقدم مقاييس الفترات العناوين، الترتيبات، الى جانب فترة محددة بين كل من خيارات متغيراتها. ويحمل المقياس النسبي جميع خصائص مقياس الفترات، بالإضافة الى أن بإمكانه استيعاب قيمة "الصفر" في أي من متغيراته.

وبناءً عليه سنميز بين أربعة أنواع من المقاييس:

1 – المقياس الاسمي (Nominal Scale):

وهو أدنى مستويات القياس، وفيه تستخدم الأعداد فقط للتمييز بين الأشياء، فالهدف من هذا النوع هو التصنيف والعمل على تجميع الأشياء التي تشترك في خاصة ما تميزها من غيرها من الفئات. فنحصل على بالتكرار، وأحياناً نصنف البيانات بالنسبة لخاصتين مختلفتين بنفس الوقت، فهنا كل مجموعة ليست متميزة من حيث الأهمية أو الترتيب كما أنه ليس للعمليات الحسابية على الأرقام أي معنى، لأن الأرقام هنا لا تعبر عن كميات ولا يمثل الرقم كمية ما يحتويه الشيء المصنف من تلك الخاصة وإنما، يدل الرقم على معنى كيفي لمجرد التصنيف فقط.

مثال (1): عند دراسة تأثير التدخين على الإصابة بمرض ما وبعد جمع البيانات سنحصل على عينة، كل فرد من أفرادها مدخن أو غير مدخن وكل فرد مصاب أو غير مصاب بذلك المرض، حينئذٍ نعرف متغيرين نعتبر الأول يأخذ القيمة 1 إذا كان الشخص مدخناً والقيمة 0 إن لم يكن مدخناً، والمتغير الثاني يأخذ القيمة 1 إن كان مصاباً والقيمة 0 إن كان سليماً فالرّقْمان صفر وواحد لا يمثلان كمية، بل استخدما لتصنيف العينة مرة إلى مدخنين وغير مدخنين، ومرة حسب الحالة الصحية إلى مصابين أو أصحاء.

ويمكننا أيضاً النظر للعينة على أنها مكونة من أربع فئات:

الفئة الأولى المدخّنون المرضى، فنعطيها مثلاً الرقم 1.

الفئة الثانية المدخّنون غير المرضى فنعطيها الرقم 2.

الفئة الثالثة غير المدخنين المرضى فنعطيها الرقم 3.

الفئة الرابعة غير المدخنين غير المرضى فنعطيها الرقم 4.

فالأرقام 1 ، 2 ، 3 ، 4 لا تعبر عن كمية، وليس للعمليات الحسابية معنى عليها، ولا تعط الرقم الأكبر أيّ أهمية لمجموعته.

مثال (2): لإيجاد علاقة افتراضية بين الزمرة الدموية وإحدى الصفات الجسدية كالوزن مثلاً، نسحب عينة من البالغين الذكور، ونسجل لكل شخص زمرته الدموية ووزنه، ثُمَّ نستخدم الأرقام لتصنيف العينة حسب الزمرة الدموية، ونجعل الرقم 1 مثلاً يعبر عن الزمرة A^+ والرقم 2 للزمرة A^- وهكذا... فليس للأرقام أي دلالة كمية، بل استخدمت لتصنيف العينة حسب الزمرة الدموية، ولا يكسب الرقم الأكبر للمجموعة الممثلة أيّ أهمية تميزها من الزمرة الدموية الممثلة برقم أصغر.

2 – المقياس الرتبي Ordinal Scale :

يأتي هذا المستوى بعد المستوى الاسمي من حيث التعقيد، فهو يسمح بترتيب السمات دون اعتبار لتساوي الفروق بين أي رتبتين، فالشخص الذي يمتاز بسمة أكبر من غيره يكون ترتيبه الأول وهكذا .... ولا يشترط أن تكون الفروق بين درجات الصفة المدروسة متناسبة أو مساوية للفروق بين رتبها، فرتبة السمة تعبر عن أن الشخص يمتلك من السمة المقيسة أكثر أو أقل مما يمتلكه آخر، ولكن تلك الرتبة لا تدل على مقدار ما يمتلكه كل منهم. لذلك لا نستطيع أن نجري أيّ عمليات حسابية على تلك القياسات. لكننا نستطيع أن نعد تكرارات العينة عند كل رتبة وحساب الوسيط وبعض اختبارات الدلالة الإحصائية مثل اختبار الوسيط وغير ذلك.

مثال (3): بفرض أننا نريد إجراء دراسة متعلقة بشدة الإصابة بمرض ارتفاع السكر في الدم لسكان مدينة دمشق. أخذت عينة من سكان المدينة بشكل عشوائي، وأجريت لها مجموعة من الاختبارات والتحاليل الطبية خلال فترات زمنية، وسجلنا النتائج لكل شخص، وحسب معايير معينة واعتماداً على خبرة لجنة من الأطباء صنفت العينة حسب شدة الإصابة بالمرض إلى الفئات الآتية (سليم، مصاب، مصاب إصابة خفيفة، مصاب بشدة، إصابة شديدة جداً) نرفق هذه الصفات بالأرقام الآتية على الترتيب (4، 3، 2، 1، 0)، وهي مجموعة قيم المتغير العشوائي.

إن هذه الأرقام التي تمثل شدة الإصابة تسمح لنا فقط بالتمييز والمقارنة بين شخص وآخر. لكن الفرق بين رقمين لا يقارن مع الفرق بين رقمين آخرين. أي لا معنى لعملية الطرح بين هذه الأرقام، وإذا كانت أداة القياس متدرجة كما في هذا المثال فيمكننا معاملة البيانات على أنها بيانات فترية واستخدام الإحصاءات الوسيطية في معالجتها.

3 – المقياس المجالي Interval Scale:

يعد هذا النوع من القياس أدق من القياسين السابقين ؛ إذ أنّه يتصف بكل ما سبق، إضافة إلى أنه يتمتع بوحدات متساوية تمكننا من أن نحدد إذا كان شَيءٌ يساوي شيئاً آخر أو أكبر أو أصغر وقيمة الفرق بين الكبير والصغير، لذلك نستطيع جمع هذه المسافات أو طرحها، ولكن لا يمكن إجراء عملية القسمة في هذا النوع من القياس وذلك لعدم وجود الصفر المطلق (أي إن قيمة "صفر" تكون نسبية وليست مطلقة، ولكننا لا نستطيع اعتبار أن هذه الدرجة تناظر مقدار السمة التي يفترض أن الاختيار قد صمم لقياسها وإلا كان معنى ذلك أن مقدار السمة المقيسة عند الطرف هي صفر. وفي هذا المستوى من القياس يمكن حساب المتوسطات والانحرافات المعيارية ومقاييس العلاقة الخطية.

مثال (4): بفرض أننا نرغب في دراسة العلاقة بين مستوى التحصيل العلمي لخريجي كلية العلوم ومستوى المعيشة الأسرية أو المنطقة التي يسكنها الطالب. وصنفنا الطلاب المتخرجين كما يأتي المجموعة الأولى من كان معدله أقل من 70% ، والثانية من كان معدله بين 70% و 80% ، والثالثة بين 80% و90%، والرابعة من كان معدله 90% أو أعلى.

نعرف الآن المتغير العشوائي بالشكل التالي: يأخذ القيمة صفراً إذا كان الطالب من المجموعة الأولى، والقيمة 1 إذا كان من المجموعة الثانية، وهكذا ... فتكون مجموعة قيمه هي المجموعة { 0 ،1 ،2 ،3 } ، يمكننا ترتيب هذه القيم حسب دلالاتها من الأدنى إلى الأعلى: الدرجة صفر مقبول، ثُمَّ الدرجة 1 نصنفها بالجيد ،والدرجة 2 جيد جداً ، والدرجة 3 ممتاز ومِنْ ثُمَّ يمكن مقارنة طالب درجته 1 عن آخر درجته 2 مثلاً. إضافة إلى ذلك نستطيع المقارنة بين الفرق الأول ما بين الدرجة 1 والدرجة صفر مع الفرق بين الدرجة 2 والدرجة 1، أي الفروق بين الدرجات تتناسب مع الفروق بين الرتب، ودرجة الصفر هنا لها معنى نسبي وليس معنى مطلقاً.

4 – المقياس النسبي Ratio Scale:

وهذا النوع من المقاييس هو أعلى مستويات القياس، حيث يمكن استخدام جميع العمليات الحسابية، إذ أن له صفراً مطلقاً يعني انعدام الصفة التي نقيسها. وتتوفي في هذا المستوى جميع خصائص مقاييس المسافة بالإضافة إلى الصفر المطلق وهذا النوع من المقاييس مألوف لنا أكثر من غيره من المقاييس، وذلك لأن جميع أبعاد الأجسام كالطول والوزن والحجم يمكن قياسها بهذه الطريقة، ولهذا يمكن القول إن الشخص الذي يبلغ طوله 180 سم له ضعف طول الشخص الذي طوله 90

سنتيمتر. وتسمية هذا النوع باسم مقاييس النسبة جاءت من قابليته لاستخراج النسبة بين الأعداد والتعبير عن القياس في صورة نسبة. ويستخدم هذا النوع من القياس لتمثيل صفات يمكن قياسها أو قياس كميتها مثل تركيز السكر في الدم، الكوليسترول، عدد الكريات البيضاء في حجم معين ... إلى آخره.

وهذا النوع من المقاييس غير معروف في المقاييس النفسية والتربوية إلا في حالات قليلة جداً مثل بعض الصفات النفسية الجسمية مثل زمن الرجع. ويجب أن نميز بين البيانات الكمية والبيانات الوصفية. إذ تنتج البيانات الكمية عن استخدام مقاييس الرتبة أو مقاييس المسافة أو المقاييس الاسمية، أما المتغيرات الوصفية فتتضمن تقسيم البيانات أو توزيعها في فئات من المستوى الاسمي، وعندما نستخدم هذه الفئات فإننا نهتم بعدد الأفراد الذين يحتلون كل فئة أي تكرارات الفئات. وكثير من العمليات الإحصائية التي نستخدمها بالنسبة للمتغيرات الكمية، لا يمكن استخدامها استخداماً ذا معنى مع البيانات التي تتكون من تكرارات. وذلك يجب على الباحث أن يعرف نوع البيانات التي لديه قبل أن يتخذ قراراً بنوع التحليل الإحصائي الذي يستخدمه.

نلاحظ أن القياسات التي تقيس مستويات تلك الصفات يمكن إجراء جميع العمليات الحسابية عليها، واستخدام الطرق الإحصائية الوسيطية.

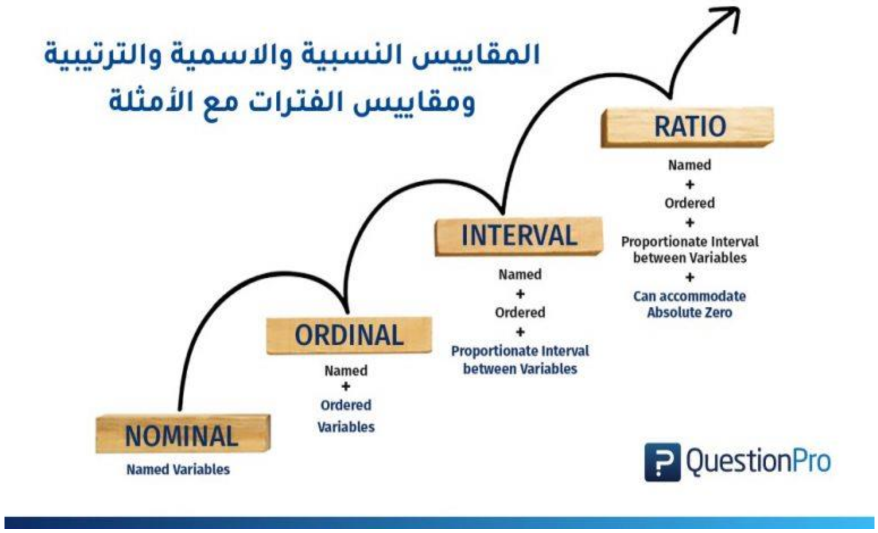
مثال (5): قياس متغير تركيز السكر في الدم هو من المستوى النسبي.

مثال: المخطط ادناه يوضح مخطط الساق والأوراق لعلامات 25 طالب) العلامة من 100 (. حيث أن الساق يمثل خانة العشرات، والأوراق تمثل خانة الآحاد. فمثلاً العلامات 55،55،56،59) هي الأقل في مجموعة العلامات(تم تمثيلها ضمن الساق الأول وهو العدد) 5(، والأوراق تضمنت الأرقام 5، 5، 6، 9 .بينما مثا الأرقام 100 ، 100 هما العلامتين الاعلى في المجموعة. والمكونة من ثلاث خانات، نخصص الخانتين الأوليتين 10 كساق، ونمثل الخانة الأخيرة بالورقة.

أمثلة المقياس النسبي: الأسئلة التالية تندرج ضمن فئة المقياس النسبي:

- كم يبلغ طول إبنتك حاليا؟
- أقل من 5 قدم
- 5قدم و1 إنش – 5 قدم و5 إنش
- 5قدم و6 إنش – 6 قدم
- أكثر من 6 قدم
- ما هو وزنك بالكيلوغرام؟
- أقل من 50 كيلوغرام
- 51 – 70كيلوغرام
- 71 – 90كيلوغرام
- 91 – 110كيلوغرام
- أكثر من 110 كيلوغرام

اعرف عن: مقياس الفترات مقارنة بالمقياس النسبي



1 – 4- بعض التطبيقات الهامة في الإحصاء:

تعد طرائق التحليل الإحصائية جزءاً هاماً من طرائق البحث العلمي مما أدى إلى استخدامها على نطاق واسع وبخاصة طرائق تحليل المتغيرات المتعددة التي يكثر تطبيقها عند دراسة مشاكل العلوم الطبيعية والطبية ... . سوف نستعرض بعض المشاكل الواقعية التي تستخدم فيها طرائق تحليل المتغيرات المتعددة بنجاح مما يسمح لنا أن نتبين مدى تنوع هذه المشاكل وكثرتها.

1 – علم الأحياء:

مثال (6): من المهم عند تربية وتحسين النباتات اختيار أنواع من النبات تستخدم في استنباط أنواع جديدة ذات مزايا وخواص لم تكن تتمتع بها الأنواع السابقة. ويهدف عالم النبات هنا إلى التوصل إلى أكبر عدد من المكاسب الوراثية في أقل وقت ممكن. ففي برنامج لتحسين نبات الفاصوليا حُوِّلت المتغيرات الخاصة بالبروتين الذي يحتويه النبات وكمية المحصول إلى مؤشر كمي يستخدم في اختيار أنواع معروفة من النبات بهدف تحسينها واستنباط أنواع جديدة منها. ولقد تم التوصل إلى هذا المؤشر الكمي باستخدام طرائق تحليل المتغيرات المتعددة. (الهدف هنا هو التوصل إلى مؤشر كمي ليحل محل متغيرات عديدة ، أي تخفيض البيانات) .

2 – دراسات بيئية:

مثال (7): أجريت دراسات موسعة على درجة تركز ملوثات الهواء في مدينة حمص. في إحدى هذه الدراسات، أخذت قياسات يومية عن سبع متغيرات مرتبطة بتلوث الهواء خلال فترة زمنية ممتدة. ولقد تركز الاهتمام على معرفة ما إذا كانت

مستويات تلوث الهواء ثابتة تقريباً خلال أيام الأسبوع أو ما إذا كان هناك فرق واضح بين مستويات تلوث الهواء خلال أيام الأسبوع ومستوياته خلال أيام العطلة الأسبوعية. كذلك كان من بين أهداف الدراسة معرفة ما إذا كان من الممكن تلخيص البيانات الضخمة التي تم جمعها بطريقة تمكن من فهمها وتفسيرها. (الهدف هنا هو إجراء اختبارات فروض وتخفيض البيانات) .

3 – الطب:

مثال (8): أجريت دراسة لمعرفة مدى استجابة مرضى السرطان للعلاج بالأشعة حيث أخذت بيانات عن 6 متغيرات في عينة من 98 مريضاً. ونتيجة لصعوبة فهم وتفسير البيانات التي جمعت عن المتغيرات الست في وقت واحد فإنه أصبح من الضروري الحصول على مقياس أبسط لقياس مدى استجابة المرضى لهذا النوع من العلاج.

ولقد استخدمت طرائق تحليل المتغيرات المتعددة للوصول إلى مثل هذا المقياس وذلك باستخدام أغلب البيانات المتاحة من العينة. (الهدف هو تخفيض البيانات) .

مثال (9): يمكن تسجيل رد الفعل المعاكس للمنبهات البصرية (مثل الضوء المبهر) من جمجمة الإنسان مباشرة وذلك باستخدام الحاسب الآلي. ويشار إلى هذا الأسلوب بالأحرف VECA وفي دراسة طبية لمعرفة تأثير تصلب الأنسجة المتعدد على الجهاز البصري استخدم تحليل المتغيرات المتعددة لمعرفة ما إذا كان استخدام أسلوب VECA يعتبر عملياً ويمكن الاعتماد عليه في تشخيص أمراض الجهاز البصري. (الهدف هنا هو التصنيف، أي الوصول إلى مقياس كمي يمكن بواسطته فصل الأشخاص الذين يعانون من تصلب الأنسجة المتعدد الذي يؤدي إلى أمراض في الجهاز البصري عن هؤلاء الذين لا يعانون من هذا المرض) .

4 – علم النفس:

مثال (10): تقترح نظريات حديثة في علم النفس أن التركيب الوظيفي للمجتمع الأمريكي يتحدد بناء على بعد اجتماعي اقتصادي هام بالإضافة إلى أبعاد أخرى تأثيرها ضئيل وغير معروفة بينما تقترح نظريات أخرى أنه يتحدد بناء على ثلاثة أبعاد معرفة تماماً هي (1) احتياجات العمل (2) نظام العمل (3) العائد . ولمعرفة أي من الاتجاهين السابقين يعتبر ملائماً، أخذت بيانات عن 25 متغيراً من 583 وظيفة وتم تحليلها باستخدام طرائق تحليل المتغيرات المتعددة.(لاحظ أن الهدف هنا هو اختبار الفرض القائل بأن التركيب الوظيفي للمجتمع الأمريكي يتحدد بناء على بعد واحد في مقابل الفرض القائل أنه يتحدد بناء على ثلاثة أبعاد) . تقترح نظريات حديثة في علم النفس أن التركيب الوظيفي للمجتمع الأمريكي يتحدد بناء على بعد اجتماعي اقتصادي هام بالإضافة إلى أبعاد أخرى تأثيرها ضئيل وغير معروفة بينما تقترح نظريات أخرى أنه يتحدد بناء على ثلاثة أبعاد معرفة تماماً هي (1) احتياجات العمل (2) نظام العمل (3) العائد . ولمعرفة أي من الاتجاهين

السابقين يعتبر ملائماً، أخذت بيانات عن 25 متغيراً من 583 وظيفة وتم تحليلها باستخدام طرائق تحليل المتغيرات المتعددة.(لاحظ أن الهدف هنا هو اختبار الفرض القائل بأن التركيب الوظيفي للمجتمع الأمريكي يتحدد بناء على بعد واحد في مقابل الفرض القائل أنه يتحدد بناء على ثلاثة أبعاد) .

5 – الاقتصاد والتجارة:

مثال (11): استخدمت البيانات المتاحة عن 6 من المتغيرات المحاسبية والمالية للوصول إلى نموذج إحصائي لمساعدة المهتمين بالنشاط التأميني للتعرف على شركات التأمين التي تعاني من مشاكل السيولة المالية، وباستخدام هذا النموذج يمكن تحديد ما إذا كانت الشركة تعاني من مشاكل السيولة أو لا حتى يمكن مساعدة الشركات التي تواجه مثل هذه المشاكل واتخاذ الإجراءات المناسبة لمنع إفلاسها. (الهدف هنا هو التوصل إلى قاعدة للتصنيف يمكن بها التمييز بين الشركات التي تعاني من مشاكل السيولة وتلك التي لا تعاني منها) .

6 – الرياضة والتعليم:

مثال (12): يهتم المختصون بالألعاب الأولمبية بتحليل مسابقات المضمار على أمل التعرف

على المهارات الأساسية اللازمة لكل مسابقة. جُمعت بيانات عن ثُمَّان مباريات عُشارية أولمبية مختلفة واستخدمت طرائق تحليل المتغيرات المتعددة بهدف التعرف على العوامل الجسمانية التي تفسر نتائج هذه المباريات العُشارية. ولقد بينت نتائج الدراسة أنه يمكن تفسير نتائج المباريات بناء على العوامل الجسمانية التالية: سرعة الجري، قوة الذراعين، قوة الرجلين، والقدرة على التحمل. (الهدف هنا هو تحديد مدى اعتماد متغيرات مشاهدة [نتائج مسابقات المضمار] على عدد أقل من المتغيرات [العوامل الجسمانية]) .

مثال (13): قام أحد المحللين الرياضيين بإجراء دراسة على نتائج ألعاب الفرق الرياضية السورية بكرة القدم، وذلك من خلال نتائج ألعاب الفرق خلال خمسة مواسم مختلفة في مباريات الدوري التي تقام سنوياً وأخذ المباريات التي أقيمت خارج أرضها. وقد بينت نتائج الدراسة أنه يمكن تبيان فيما إذا كان هناك اختلاف بين أداء الفرق السورية الكبرى عندما تلعب خارج أرضها، وذلك على أساس التصميم تام العشوائية. كما ويمكننا ترتيب هذه الفرق السابقة حسب الأداء الأفضل، ثُمَّ تبيان أياً من الفرق السابقة التي تختلف عن بعضها بناءً على الاختبارات التي تعرفها (سرعة الجري، قوة الرجلين، والقدرة على التحمل مثلاً).

مثال (14): غالباً ما تستخدم نتائج اختبار القدرات المدرسية (SAT) ونتائج الدراسة في المرحلة الثانوية كمؤشر لمدى النجاح في المرحلة الجامعية. جمعت بيانات عن خمسة متغيرات عن المرحلة الثانوية (من بينها نتائج اختبار القدرات المدرسية الشفوي والكمي ونتائج الدراسة في السنتين الأخيرتين من المرحلة الثانوية) وأربعة متغيرات عن المرحلة الجامعية (درجات دراسة أربع مواد مختلفة) وذلك لمعرفة مدى فائدة استخدام متغيرات المرحلة الثانوية للتنبؤ بالأداء في المرحلة

الجامعية . (إن هدفنا هو التنبؤ بالأداء في المرحلة الجامعية باستخدام متغيرات المرحلة الثانوية . ويمكننا أيضاً استخدام نتائج هذه الدراسة لتصنيف الطلبة إلى طلبة يتوقع نجاحهم في الدراسة الجامعية وطلبة لا يتوقع نجاحهم فيها) .

1-5- مصادر جمع البيانات Collection of Data:

يتم جمع البيانات من مصدرين أساسيين، وهما:

1-المصدر الرسمي والتاريخي: وهو أن تؤخذ البيانات الإحصائية من السجلات المحفوظة في الهيئات والمؤسسات والوزارات المختلفة، ويمكن معرفة البيانات الإحصائية المختلفة لدولة ما في مجالات الصحة والتعليم والاقتصاد والنشاطات الأخرى من سجلات هيئة الأمم المتحدة، أو المؤسسات الحكومية، أو الهيئات الدولية الأخرى.

2-المصدر الميداني: يتم جمع البيانات الإحصائية بطريقتين رئيسيتين هما:

آ -**طريقة البحث الشامل:** وتقتضي هذه الطريقة إجراء القياسات أو أخذ البيانات من جميع الوحدات أو الأفراد الذين لهم علاقة بموضوع الدراسة.

ب -**طريقة المعاينة:** يتم أخذ البيانات من المجتمع الإحصائي وذلك وفق طريقة سحب العينات ويتم ذلك عن طريقة تصميم استبيان إحصائي لموضوع الدراسة، وتؤخذ المعلومات عن طريقة:

1 - **المقابلة الشخصية:** يقوم جامع البيانات بإجراء مقابلة مع الأفراد المراد جمع بيانات منهم وتسجيل إجاباتهم على استمارة خاصة ندعوها استبانة تحتوي على مجموعة من الأسئلة المطلوب الإجابة عنها. وهذه الطريقة تتيح لجامع البيانات الحصول على إجابات دقيقة، وذلك لإمكانية توضيح الأسئلة للفرد موضع الدراسة، ومراقبة رد فعله على بعض الأسئلة، وتمتاز هذه الطرق بارتفاع نسبة المستجيبين وذلك لإمكانية متابعتهم. أما عيوب هذه الطريقة فهي أن جامع البيانات قد يكون مصدراً من مصادر التحيز بسبب توجيه الفرد إلى إجابات معينة سواءً عن طريق الإعجاب أم الاستنكار لبعض الأمور بحيث يؤثر في إجاباته. كما أن بعضهم قد يسجل بيانات (إجابات) خاطئة أو يقوم بتزوير الإجابات ، وذلك دون مقابلة أي فرد فعلياً، فيملأ الاستمارة بنفسه كما يرغب أو يقوم هو بالإجابة عن الأسئلة وملء الاستمارة بغية تسجيل عدد أكبر من الاستمارات وبأقصر وقت ممكن، لهذا لا بد من تدريب العاملين بشكل جيد والتدقيق عليهم ومراقبتهم. كما أن تكلفة هذه الطريقة عالية نسبةً لغيرها من الطرق. وهذه الطريقة تستخدم في المجتمعات التي ترتفع فيها نسبة الأمية.

2 - **طريقة الهاتف:** يتم الحصول على البيانات بطرح الأسئلة عن طريق الهاتف للتقليل من التكلفة وتوفير الجهد والوقت. ومن مساوئها الطريقة أنّها لا تشمل جميع أفراد المجتمع؛ لأنها تستثني جميع الأفراد الذين لا يملكون هواتف، وبذلك يكون الإطار ناقصاً. وهذه الطريقة قد تكون جيدة إذا كان جميع أفراد المجتمع موضع الدراسة مشتركين بالهاتف..

3 - **طريقة البريد أو البريد الإلكتروني:** ترسل الاستمارات بالبريد للأفراد موضع الدراسة لتعبئتها وإعادتها. وهذه الطريقة تعتبر من أقل الطرق كلفة في الحصول على المعلومات، ولكن من عيوبها أن استجابات الأفراد تكون بطيئة وقليلة وغير كاملة أو غير مفهومة بل غير موثوقة، لذلك تكون البيانات غير دقيقة. ولتشجيع الأفراد على الاستجابة يجب تصميم الاستمارة على شكل دقيق وبسيط وجذاب يشجع على الاستجابة، كما يجب دفع أجور البريد أو تقديم حوافز معينة تدفع

الفرد وتحفزه لملء الاستمارة وإعادتها. ومن أهم عيوب هذه الطريقة قلة الردود، كما أنه من الممكن أن يتطفل أشخاص غير معنيين ولا يمثلون مجتمع الدراسة فيقوموا بالإجابة عن الأسئلة، وهذا مصدر تحيز آخر للبيانات..

4 – **طريقة الإنترنت وطرائق التواصل الاجتماعي**: يمكن طرح بعض الأسئلة أو بعض المواضيع للمناقشة على صفحات الويب ونطلب من الراغبين في المشاركة والإجابة أو إبداء الرأي في الموضوع المدروس. فيقوم متصفح الإنترنت بالإجابة أو إبداء الرأي، وتصلنا الإجابات خلال وقت قصير تمتاز هذه الطريقة ببساطتها وقلة تكاليفها وسرعتها، لكن من أهم مساوئها أن العينة متحيزة ولا تشمل المجتمع، بل هي مقتصرة على مستخدمي الإنترنت والمشاركين في المنتديات الاجتماعية..

5 – **الهواتف المحمولة**: يمكن أيضاً طرح الأسئلة وابداء رأي الأفراد عن طريق الهواتف المحمولة لنصل إلى مشتركي خدمة الهواتف المحمولة. فهذه الطريقة قد تكون أفضل من سابقتها؛ لأنها تشمل شريحة أوسع من المجتمع، فهي تغطي كل الأشخاص المشتركين في هذه الخدمة. كذلك لا بد من تقديم بعض الحوافز التي تدفع الأفراد وتحفزهم للرد على الأسئلة، وذلك بصراحة وصدق.

6 – **الملاحظة أو المشاهدة**: من أقدم الطرق المستخدمة لجمع البيانات لمراقبة كثير من الظواهر ولجمع بيانات عن سلوك الأفراد وعن تفاعل الأفراد وكذلك عن البيئة المحيطة، ولاسيّما تلك التي يصعب سؤال الأفراد موضع الدراسة عنها مثل الأطفال أو الحيوانات ... إلى آخره أو عند مراقبة ظاهرة معينة خلال فترات زمنية مثل كميات الأمطار، سرعة الرياح ــ أعداد حوادث السير على طريق... إلى آخره. وتتم الملاحظة بتعيين أشخاص مدربين للمراقبة وتسجيل ملاحظاتهم. ومن مساوئها أيضاً الوقت الطويل والكلفة العالية وتحيز المراقب أو الملاحظ الذي قد يوجه النظر لسلوكيات بعينها دون غيرها.

7 – **التجربة**: نقوم بإجراء التجربة وبتسجيل نتائجها، وذلك للحصول على معلومات مفيدة. فمثلاً لقياس تأثير بعض العوامل في منتج زراعي نقوم بتصميم التجربة بطرق إحصائية مثلاً لقياس شدة تأثير نوع البذار ونوع التربة وطريقة الري في كمية المحصول الزراعي. أو نقوم بمراقبة نتيجة إعطاء لقاح معين لمعالجة أحد الأمراض، وقد تكون التجربة هي قياس شدة تأثير نوع من الهرمونات في مجموعة من الفئران......

8 – **التسجيل**: تستخدم بعض التقنيات الإلكترونية للمراقبة والعد وتسجيل الملاحظات. فنقوم بتركيب كاميرات مراقبة للداخلين والخارجين من مكان ما أو نقوم بتركيب جهاز يقوم بعد المركبات على طريق محددة، ويمكن أن نطلب من الأفراد تسجيل المعلومات في بعض السجلات المخصصة لذلك، وتستخدم اليوم وعلى نطاق واسع الفاكس والبريد الإلكتروني والإنترنت في جمع المعلومات وتسجيلها.

9 – **المجموعات البؤرية focus groups discussion** هي تقنية غير رسمية إلى حد ما يمكن أن تساعد في تقويم احتياجات المستخدمين ومشاعرهم على حد سواء. يدعى إلى المجموعة عادة 6–9 مشاركين لمناقشة الموضوع قيد الدراسة، وتستغرق المجموعة عادة ساعتين من النقاش، ويحكم نجاحها مدير للحوار مدرب بالقدر الكافي وبرتوكول نقاش مصوغ بشكل جيد وديناميكية مجموعة جيدة يعززها المحاور وطبيعة المشاركين.

10 – المقابلات المعمقة: تجرى المقابلات المعمقة in–depth interviews بأيد خبيرة باستخدام دليل للمقابلة، وعادة ما توفر بيانات غنية ومعمقة حول المواضيع قيد الاستكشاف، وتسمح في توضيح المسائل فتزيد في احتمال حدوث استجابات مفيدة. ومن مساوئها أنها تستغرق وقتاً طويلاً وتحتاج إلى تدريب جيد للمنفذين كما أن حجم المعلومات عنها يكون عادة كبيراً بسبب الحوار الطويل.

1–6– عرض البيانات Presentation of Data:

بعد الانتهاء من جمع البيانات الإحصائية بطريقة أو أكثر من الطرائق السابقة، فإنها تكون في صورة غير معبرة ومبعثرة، وقد يصعب استنتاج أي معلومات مفيدة منها. وقد تكون عبارة عن مجموعة أرقام غير مرتبة، أو مجموعة صفات لبعض الخصائص الموجودة في الاستبيان الإحصائي. ولتوضيح ذلك نعرض المثالين التاليين:

مثال (15): البيانات الآتية تبين التقديرات التي حصل عليها ثلاثون طالباً من طلاب جامعة البعث في إحدى السنوات:

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

مثال (16): البيانات الآتية تمثل كمية سكر الدم لعينة مؤلفة من خمسين مريضاً راجعو مشفى المواساة خلال شهر كانون الأول الماضي يعانون من احتشاءات دماغية مختلفة:

180	90	205	110	250	280	210	200	290	275
62	80	195	110	245	190	200	215	285	270
222	225	185	120	330	298	190	205	150	150
240	250	338	130	160	140	180	210	160	160
230	240	175	125	170	175	195	215	130	110

البيانات الواردة في المثالين السابقين لا يمكن الاستفادة منهما في أية دراسة بهذا الشكل وذلك لعدم وضوحهما، وصعوبة الحصول على أي معالم من التقديرات في المثال (15)، وكمية سكر الدم في المثال (16)، فمثلاً البيانات السابقة بوضعها الحالي تجعل من الصعب التعرف على الطلاب الحاصلين على تقدير مشترك ممتاز (A) أو جيد جداً (B) ...، وكذلك الحال لا يمكننا معرفة متوسط كمية سكر الدم التي أخذت من عينة مؤلفة من خمسين مريضاً راجعو مشفى المواساة خلال شهر كانون الأول الماضي والذين يعانون من احتشاءات دماغية مختلفة بسهولة من بيانات المثال (16) بوضعها الحالي.
لذلك أصبحت الحاجة إلى إيجاد طريقة لتنظيم وتلخيص مثل هذه البيانات في صورة سهلة، ضرورية جداً، حتى يمكننا دراستها، واستنتاج كل المعلومات المطلوبة بسهولة، ومن الطرائق المستخدمة لتلخيص البيانات التوزيعات التكرارية.

أولاً –جداول التوزيع التكرارية:

لتلخيص وتنظيم البيانات الوصفية نكون جدولاً مؤلفاً من ثلاثة أعمدة رأسية يكتب في بداية كل عمود عنوانه المناسب، يسمى مثل هذا الجدول جدول تفريغ البيانات ومنه نستنتج جدولاً آخراً يسمى جدول التوزيع التكراري، فمثلاً إذا كانت الدراسة هي التقديرات التي حصل عليها ثلاثون طالباً من طلاب جامعة البعث في إحدى السنوات فإننا نكتب كلمة التقدير في العمود الأول ثُمَّ يكتب تحت العنوان في العمود الأول كل الصفات، وهذه الصفات في هذه الحالة هي: A، B، C، D. أما في العمود الثاني فيكون العنوان هو إفراغ البيانات وفيه تسجيل القراءات على شكل خطوط رأسية مثل " | " فإذا ما وصل عدد الخطوط إلى أربع مثل " |||| " فإن الخط الخامس يكتب بشكل أفقي ليكوّن حزمة مثل "卌" عدد عناصرها خمسة. وبعد تفريغ البيانات نقوم بعدّ جميع عناصر الحزم أمام كل صفة ونكتب العدد الناتج في العمود الثالث الذي يدل على التكرار، ويقصد بالتكرار عدد عناصر الظاهرة أمام كل صفة من الصفات الموجودة في العمود الأول.

ومن هذا الجدول يصاغ جدول التوزيع التكراري المكون من عمودين الأول يشتمل على أسماء الصفات، والثاني التكرارات.

ففي المثال الأول يكون جدول تفريغ البيانات بالشكل الآتي:

الصفة	**إفراغ البيانات**	التكرار
A	卌 ||	7
B	卌 卌	10
C	卌 |||	8
D	卌	5
المجموع		30

لنحذف الآن العمود الثاني من الجدول السابق فإننا نحصل على جدول مكوّن من عمودين فقط يسمى جدول التوزيع التكراري كما هو موضح بالجدول الآتي:

الصفة	التكرار
A	7
B	10
C	8
D	5
المجموع	30

نلاحظ كذلك، أن أي جدول إحصائي يحتوي على عنوان يوضح نوعية الجدول، وطبيعة البيانات المعروضة فيه، كما هو موضح في الجدولين السابقين. ولتلخيص وتنظيم البيانات الكمية فإننا نكوّن جدولاً مؤلفاً من ثلاثة أعمدة، حيث نستبدل الصفة في العمود الأول بما يسمى حدود الفئات ولإنشاء هذا الجدول نتبع الخطوات الآتية:

1 – نحدد مدى البيانات، ومن المثال (16) يكون المدى كالآتي:

$$R = \max x - \min x = 338 - 62 = 276$$

2 – نقسم المدى إلى عدد اختياري مناسب من الفئات (فئات نصف مفتوحة)، وعادة يتراوح عدد الفئات الاختياري من 5 إلى 15 فئة تقريباً، ففي المثال (16) نختار عدد الفئات، يساوي 7 فئات مثلاً.

3 – نحسب طول الفئة l ، وهو يساوي المدى مقسوماً على عدد الفئات المقترح، بحيث يقرّب الناتج إلى أقرب عدد صحيح، ففي المثال (16) السابق يكون طول الفئة معرفاً بالعلاقة:

$$l = \frac{\text{المدى}}{\text{عدد الفئات المقترح}} = \frac{276}{7} = 39.4 \cong 40$$

وهناك عدة طرق لحساب عدد الفئات نذكر منها:

1– اقترح العالم ستورجز تطبيق المعادلة التالية في تحديد عدد الفئات:

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

حيث k هو عدد الفئات، $\operatorname{Log} n$ هو اللوغاريتم العشري و n هو عدد التكرارات.

2– كما واقترح العالم يول yule تطبيق المعادلة التالية في تحديد عدد الفئات:

$$k = 2.5 \times \sqrt[4]{n}$$

حيث k هو عدد الفئات، $\sqrt[4]{n}$ هو الجذر الرابع لعدد التكرارات.

ففي هذه الحالة نجد عدد الفئات من العلاقة التالية:

$$k = 1 + 3.322 \log n = 1 + 3.322\, log(50)$$

$$= 1 + 3.322(1.69897) = 1 + 5.64 = 6.64 \cong 7$$

وفي الحالة الثانية نجد عدد الفئات يعطى بالعلاقة:

$$k = 2.5 \times \sqrt[4]{n} = 2.5 \times \sqrt[4]{50} = 2.5 \times 2.659 = 6.6475 \cong 7$$

وهو نفس الجواب الذي حصلنا عليه من العلاقة السابقة.

هناك حالات قد يكون فيها عدد الفئات مختلفاً باستخدام تلك العلاقتين.

4 – نحدد بداية الفئة الأولى، وذلك بأخذ أصغر رقم في البيانات أو أي رقم يكون قريباً منه، ونسميه الحد الأدنى للفئة الأولى، وكذلك نحدد نهاية الفئة الأولى بإضافة طول الفئة إلى الحد الأدنى للفئة الأولى، مع الأخذ بعين الاعتبار أن الحد الأدنى ينتمي إلى الفئة بينما لا ينتمي الحد الأعلى إلى الفئة ذاتها. وهكذا بالنسبة لباقي الفئات.

5 – نستخدم الخطوات السابقة، ونكوِّن جدولاً مؤلفاً من ثلاثة أعمدة، ثُمَّ نستبدل الصفة بحدود الفئات في العمود الأول.

ففي المثال الثاني نحصل على جدول البيانات بالشكل الآتي:

التكرار	إفراغ البيانات	حدود الفئات
3	𝍷𝍷𝍷	[60 – 100[
7	𝍸𝍷𝍷	[100 – 140[
9	𝍸 𝍷𝍷𝍷𝍷	[140 – 180[
15	𝍸 𝍸 𝍸	[180 – 220[
8	𝍸 𝍷𝍷𝍷	[220 – 260[
6	𝍸 𝍷	[260 – 300[
2	𝍷𝍷	[300 – 340[
50	**–**	**المجموع**

وبحذف العمود الثاني من الجدول السابق نحصل على جدول التوزيع التكراري للبيانات الكمية:

التكرار	حدود الفئات
3	[60 – 100[
7	[100 – 140[
9	[140 – 180[
15	[180 – 220[
8	[220 – 260[
6	[260 – 300[
2	[300 – 340[
50	**المجموع**

6 - يمكننا إضافة عمود جديد وهو مركز الفئة، وهو يساوي نصف مجموع الحدين الأدنى والأعلى للفئة. كما هو مبين في الجدول الآتي:

مراكز الفئات	التكرار	حدود الفئات
80	3	[60 – 100[
120	7	[100 – 140[
160	9	[140 – 180[
200	15	[180 – 220[
240	8	[220 – 260[
280	6	[260 – 300[
320	2	[300 – 340[
	50	**المجموع**

7 - وبإضافة عمودين جديدين لجدول التوزيع التكراري السابق وهما التكرار النسبي والمئوي، وذلك لأنه يطلب منّا أحياناً معرفة الفرق بين ظاهرتين أو أكثر بنفس الخاصية في فترات مختلفة، أو مقارنة الظواهر المختلفة لنفس الخاصية في نظامين مختلفين. ويعرَّف التكرار النسبي لفئة ما بأنه نسبة تكرار هذه الفئة إلى مجموع التكرارات، وأن التكرار المئوي هو عبارة عن حاصل ضرب التكرار النسبي بـ 100.

فلو عدنا إلى المثال (16) لوجدنا التكرار النسبي والمئوي مبينين في الجدول الآتي:

حدود الفئات	التكرار	التكرار النسبي	التكرار المئوي
$[60-100[$	3	0.06	6
$[100-140[$	7	0.14	14
$[140-180[$	9	0.18	18
$[180-220[$	15	0.30	30
$[220-260[$	8	0.16	16
$[260-300[$	6	0.12	12
$[300-340[$	2	0.04	4
المجموع	**50**	**1**	**100**

8 - يمكننا إنشاء جدولين جديدين آخرين هما جدول التوزيع التكراري الصاعد وجدول التوزيع التكراري الهابط، وذلك لأنه قد يكون المطلوب معرفة عدد التكرارات للظاهرة المدروسة التي تزيد أو تقل عن قيمة معينة.

ولإنشاء جدول التوزيع التكراري الصاعد نأخذ تكرار الفئة الأول وهو التكرار الصاعد الأول ثُمَّ مجموع تكراري الفئة الأولى وتكرار الفئة الثانية وهو التكرار الصاعد الثاني وهكذا ... كما ويمكن إنشاء جدول التوزيع الهابط وذلك بأخذ تكرار الفئة الأخيرة وهو التكرار الهابط الأول ثُمَّ حاصل طرح هذا التكرار من تكرار الفئة الأولى وهو التكرار الهابط الثاني وهكذا ... .

وسوف نوضح ذلك بالشكل التالي:

حدود الفئات	التكرار	التكرار الصاعد $f_i \uparrow$	التكرار الهابط $f_i \downarrow$
$[60-100[$	3	3	50
$[100-140[$	7	10	47
$[140-180[$	9	19	40
$[180-220[$	15	34	31
$[220-260[$	8	42	16
$[260-300[$	6	48	8
$[300-340[$	2	50	2
المجموع	**50**	—	—

ملاحظات:

أ – عند كتابة الفئات فإنه يذكر الحد الأدنى والأعلى لكل فئة إذا كان المتغير منفصـــل، ويحدد أحد الحدين ويحدد الثاني ضمنياً إذا كان المتغير مستمر.

ب – عند تفريغ البيانات فإنه يجب أن تنتمي كل مفردة إلى فئة واحدة فقط.

ج – يفضل استخدام الفئات المتساوية الطول، إلا أنه في بعض الحالات يمكن استخدام الفئات غير المتساوية، من هذه الحالات ما يلي:

1- إذا كان الغرض من الدراسة هو الاهتمام ببعض الفئات والتركيز عليها وإهمال باقي الفئات، فيمكن عندها دمج الفئات التي لا تهم الباحث في فئة واحدة.

2- إذا كان تكرار بعض الفئات صغيراً مقارنة بباقي الفئات، فيمكن دمج هذه الفئات معاً.

ثانياً -التمثيل البياني للتوزيعات التكرارية:

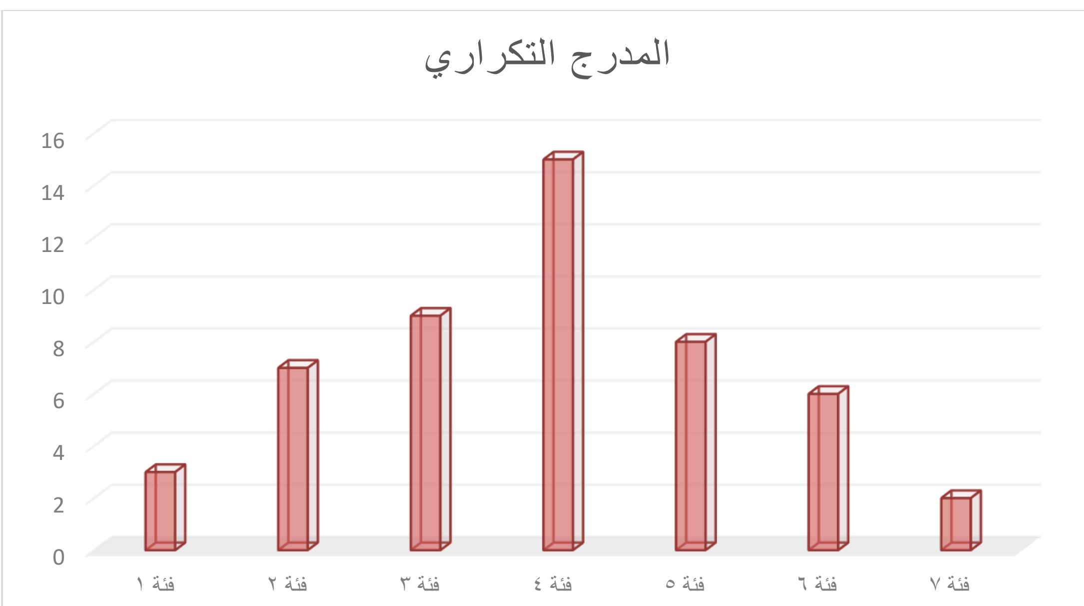
تحدثنا في الفقرات السابقة عن طرائق تنظيم البيانات وتلخيصها بواسطة جداول التوزيع التكرارية، ورأينا أن من مساوئ الجداول التكرارية هو أن البيانات تفقد هويتها نظراً لدمج عدة بيانات معاً في فئة واحدة، وأن هناك صعوبة في حساب أحد المقاييس الإحصائية عندما تكون الفئة الأخيرة مفتوحة، أما الآن فسوف نستعرض بعض أهم طرائق التمثيل البياني وهي:

1 -المدرج التكراري: يتألف المدرج التكراري من أعمدة بيانية (مستطيلات متلاصقة أو متباعدة) لها العرض نفسه مرسومة في مستوى للإحداثيات الديكارتية، حيث نمثل حدود الفئات على المحور الأفقي وتكرارات هذه الفئات على المحور العمودي.

ملاحظة: يمكن أن تكون الأعمدة البيانية عبارة عن أعمدة متلاصقة أو متباعدة، وفي حال تمثيل مجموعتين من البيانات أو أكثر نقوم بتلوين هذه الأعمدة البيانية بألوان مختلفة وذلك لكي نستطيع تمييز هذه البيانات ومن ثُمَّ المقارنة بينهما.

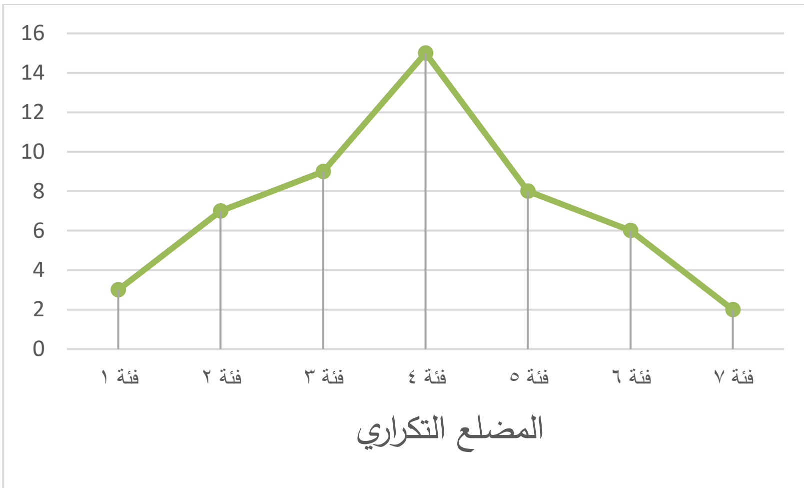
وسوف نوضح ذلك في المثال الآتي:

مثال (17): ارسم المدرج التكراري للبيانات الواردة في المثال رقم (16):



2 –المضلع التكراري: المضلع التكراري هو عبارة عن خطٍ منكسرٍ مؤلفٍ من مجموعة من القطع المستقيمة والمتلاصقة التي تصل بين منتصفات القواعد العليا للمستطيلات المرسومة في المدرج التكراري، وسوف نوضح ذلك في المثال الآتي:

مثال (18): ارسم المضلع التكراري للبيانات الواردة في المثال رقم (16):



3 –المنحني التكراري: المنحني التكراري هو مضلع تكراري يكون فيه الخط المنكسر قريباً من خطاً منحنياً، أي أنه إذا صغرت مسافة الفئة، وفي الوقت نفسه زاد عدد المتغيرات، فعندئذٍ يتقارب المضلع التكراري من خط منحن يدعى المنحني التكراري. ويعتبر المنحني التكراري من أهم الخطوط البيانية المتعلقة بالتوزيعات التكرارية، ويمكن رسمه باستبدال الخط المنكسر في المضلع التكراري بمنحنٍ يصل أغلب النقاط كما هو موضح في المثال الآتي:

مثال (19): أرسم المنحني التكراري للبيانات الواردة في المثال رقم (16):

ملاحظة (1): يمكن رسم المدرج، المضلع، المنحني التكراري الصاعد والهابط للبيانات الواردة في جدول التوزيع، ويسمى مثل هذا المدرج، المضلع، المنحني الصاعد أو الهابط.

ملاحظة (2): يمكن تمثيل ظاهرتين مختلفتين ودراسة الفرق بينهما، من خلال رسم المدرج أو المضلع أو المنحني التكراري لكل من الظاهرتين في شكل واحد.

4 – مخطط الساق والورقة:

هناك أسلوب آخر لتنظيم البيانات أعده (Tukey) يشبه أسلوب الجداول التكرارية والأعمدة، وهو مخطط الساق والورقة، فقد استبدل الأعمدة بالأعداد نفسها، فالساق هو القسم الصحيح من العدد والورقة القسم العشري.

مثال (20): يبين الجدول التالي قيم المتغير الكمي الدال على حجم الزفير القسري بالثانية لخمسين طالباً من طلاب كلية الطب بجامعة القلمون الخاصة:

2.85	2.98	3.04	3.10	3.10	3.19	3.30	3.39	3.42	3.48
3.50	3.54	3.52	3.54	3.57	3.60	3.69	3.75	3.78	3.83
3.90	3.96	4.05	4.08	4.10	4.14	4.14	4.16	4.20	4.20
4.30	4.30	4.32	4.44	4.47	4.47	4.50	4.56	4.56	4.68
4.70	4.78	4.80	4.80	4.90	5	5.1	5.1	5.2	5.3

والترتيب الآتي هو مخطط الساق والورقة لتلك البيانات

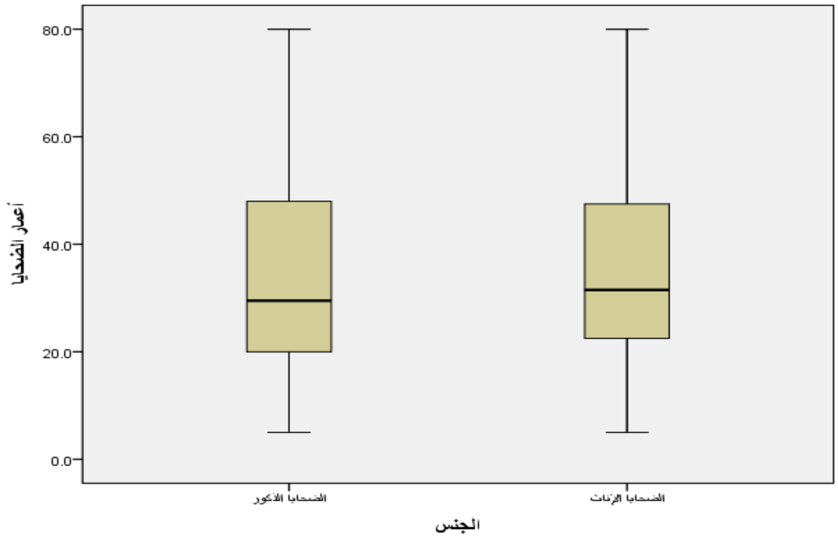
2	**8 9**
3	**0 1 1 1 3 3 4 4 5 5 5 5 5 6 6 7 7 8 9 9**
4	**0 0 1 1 1 1 2 2 3 3 3 4 4 4 5 5 5 6 7 7 8 8 9**
5	**0 1 1 2 3**

ففي الساق الأولى 2 مثلاً ورقتان 8 و9 تمثلان العددين 2.8 و2.9 للدلالة على القياسين 2.85 و2.98 من العينة ومخطط الساق والورقة شكل بسيط يوضح لنا كيفية توزع القياسات.

5 – المخطط الصندوقي: نرغب في كثير من التطبيقات بعرض توزيع البيانات بشكل مبسط لمقارنة عينتين أو أكثر، حينئذٍ قد يكون من المفيد استخدام الربيعيات. وكما نعلم لكل مجموعة من القياسات الترتيبية (على الأقل) ثلاث ربيعيات، الربيعي الأول والثاني والثالث، ويعرَف كل ربيعي بالقياس الذي تسبقه على الترتيب 25% و 50% و75% من القياسات بعد ترتيبها تصاعدياً إذ تستخدم هذه المقاييس لوصف توزيع وتشتت البيانات.

فالمخطط الصندوقي عبارة عن صندوق ذي قرنين ممتدين بشكل موازٍ للمحور الشاقولي الذي تتوزع عليه القياسات، وتشير حافتا الصندوق السفلية والعلوية للربيعيين الأول والثالث على الترتيب، وهذا يعني أن حافتي الصندوق تتضمنان نصف القياسات متوسطة القيم، ويفصل بينها خط أفقي يمثل الوسيط (الربيعي الثاني). وتقع ربع القياسات ذات القيم الأصغر تحت الصندوق وربع القياسات ذات القيم الأعلى فوقه.

مثال (21): بهدف مقارنة تشتت أعمار ضحايا حوادث الطرق للذكور والإناث رسمنا المخطط الصندوقي لعينتي الذكور والإناث انظر الشكل التالي:



6 –الرسـوم الدائرية: هي عبارة عن دائرة تقسـم إلى قطاعات زاوية أو زوايا مركزية بحيث إن مسـاحة كل قطاع زاوي أو قيمة كل زاوية مركزية تتناسب مع عدد التكرارات وتحسب مساحة القطاع الزاوي المقابل لفئة ما أو الزاوية المركزية المقابلة لفئة ما من العلاقة الآتية:

$$\text{الزاوية المركزية المقابلة لفئة ما} = \frac{\text{التكرار المقابل لهذه الفئة}}{\text{مجموع التكرارات}} \times 360$$

مثال (22): يمثل الجدول التالي مساحات القارات الست مأخوذة بالمليون كيلو متر مربع والمطلوب استخدم الرسوم الدائرية للبيانات الواردة في هذا الجدول:

القارة	المساحة بالمليون كم2
أستراليا ونيوزيلندا	8.5
أفريقيا	30.3
أسيا	47.4
أوربا	4.9
أمريكا الشمالية	24.3
أمريكا الجنوبية	17.9

الحل:

لتمثيل هذه البيانات، نقوم بإيجاد الزاوية المركزية للقطاع المقابل لكل قارة من العلاقة التالية:

$$\text{الزاوية المركزية المقابلة لفئة ما} = \frac{\text{التكرار المقابل لهذه الفئة}}{\text{مجموع التكرارات}} \times 360$$

لنحسب مجموع التكرارات بالشكل:

$$\text{مجموع التكرارات} = 30.3 + 47.4 + 4.9 + 24.3 + 17.9 + 8.5 = 133.3$$

وبالتالي فإن الزوايا المركزية المقابلة للقطاعات تصبح بالشكل:

الزاوية المركزية للقطاع المقابل لقارة أفريقيا $= \frac{30.3}{133.3} \times 360 = 82$

الزاوية المركزية للقطاع المقابل لقارة أسيا $= \frac{47.4}{133.3} \times 360 = 128$

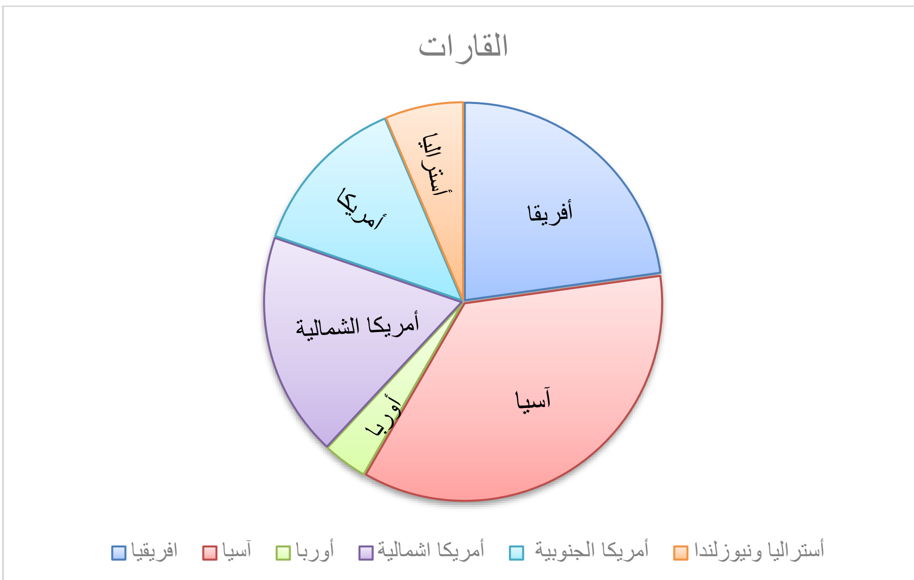
الزاوية المركزية للقطاع المقابل لقارة أوروبا $= \frac{4.9}{133.3} \times 360 = 13$

الزاوية المركزية للقطاع المقابل لقارة أمريكا الشمالية $= \frac{24.3}{133.3} \times 360 = 66$

الزاوية المركزية للقطاع المقابل لقارة أمريكا الجنوبية $= \frac{17.9}{133.3} \times 360 = 48$

الزاوية المركزية للقطاع المقابل لقارة أستراليا ونيوزيلندا $= \frac{8.5}{133.3} \times 360 = 23$

نقوم الآن برسم دائرة بحيث يمكن رسم الزوايا المركزية المقابلة لكل قارة بالشكل:



مثال (23): استخدم الرسوم الدائرية للبيانات الواردة في تمثيل بيانات هذا الجدول:

حدود الفئات	التكرار
$[60 - 100[$	3
$[100 - 140[$	7
$[140 - 180[$	9
$[180 - 220[$	15
$[220 - 260[$	8
$[260 - 300[$	6
$[300 - 340[$	2
المجموع	**50**

الحل: لتمثيل هذه البيانات، نقوم بإيجاد الزاوية المركزية للقطاع المقابل لكل فئة من العلاقة التالية:

$$\text{الزاوية المركزية المقابلة لفئة ما} = \frac{\text{التكرار المقابل لهذه الفئة}}{\text{مجموع التكرارات}} \times \mathbf{360}$$

لنحسب مجموع التكرارات بالشكل:

$$\text{مجموع التكرارات} = 3 + 7 + 9 + 15 + 8 + 6 + 2 = 50$$

وبالتالي فإن الزوايا المركزية المقابلة للقطاعات تصبح بالشكل:

$$\text{الزاوية المركزية للقطاع المقابلة للفئة الأولى} = \frac{3}{50} \times 360 = 21.6$$

$$\text{الزاوية المركزية للقطاع المقابلة للفئة الثانية} = \frac{7}{50} \times 360 = 50.4$$

$$\text{الزاوية المركزية للقطاع المقابلة للفئة الثالثة} = \frac{9}{50} \times 360 = 64.8$$

$$\text{الزاوية المركزية للقطاع المقابلة للفئة الرابعة} = \frac{15}{50} \times 360 = 108$$

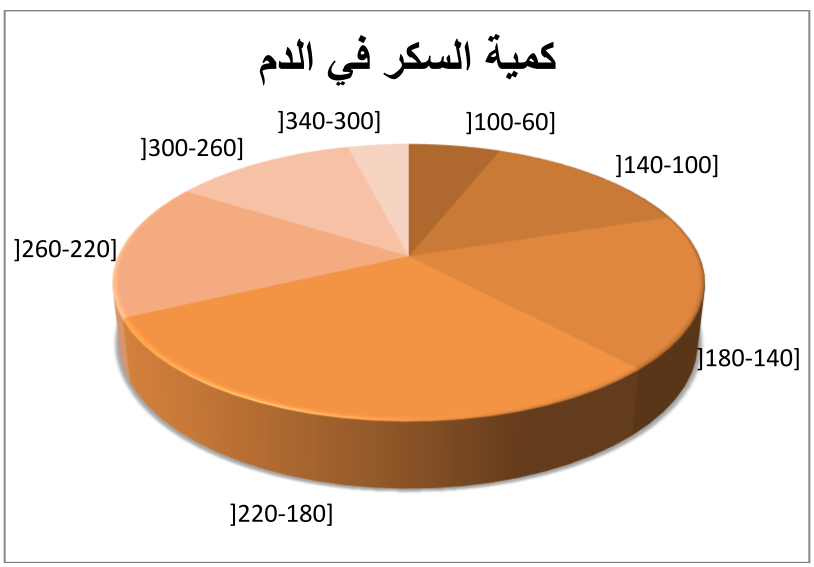
$$\text{الزاوية المركزية للقطاع المقابلة للفئة آخرهامسة إلى} = \frac{8}{50} \times 360 = 57.6$$

$$\text{الزاوية المركزية للقطاع المقابلة للفئة السادسة} = \frac{6}{50} \times 360 = 43.2$$

$$\text{الزاوية المركزية للقطاع المقابل للفئة السابعة} = \frac{2}{50} \times 360 = 14.4$$

نلاحظ أن مجموع هذه الزوايا المركزية يساوي 360 درجة

نقوم الآن برسم دائرة بحيث يمكن رسم الزوايا المركزية المقابلة لكل قارة بالشكل:



ثالثاً –محاسن الرسوم البيانية ومساوئها:

ترغب كثير من الهيئات والمؤسسات العامة في توضيح مظاهر التطور الذي تقوم به في كافة المجالات في صورة رسوم بيانية يمكن للشخص العادي استيعابها وفهمها بسهولة، إلا أن هذه الطريقة لها محاسن ومساوئ نذكر منها:

آ –**محاسن الرسوم البيانية:**

1–البساطة في قراءة البيانات وخاصة إذا كان عدد التكرارات كبيراً.

2–سهولة تذكر النتائج، حيث أن الرسوم البيانية تعطي فكرة أكثر ثباتاً من الأرقام أو الكلمات.

3–جذب الانتباه، وذلك عن طريق استخدام الألوان فإذا عني برسم الشكل البياني فمن السهل أن يجذب الانتباه إليه، ويبقى في الذاكرة لمدة أطول، بينما فإن الكثيرين لا يهتمون بعرض الجداول كثيراً.

ب –مساوئ الرسوم البيانية:

1– التضحية في دقة البيانات، إذ إن الأشكال توضح فقط التغيرات العامة، ولا تبين التفاصيل الدقيقة، لذا يفضل دوماً إرفاق الجداول مع الرسوم البيانية.

2 – تكون الرسوم البيانية أحياناً معقدة، إذا احتوت على مجموعات من البيانات المختلفة، أو تكون كثيرة التكاليف إذا كانت تحتوي على بيانات تحتاج إلى مقاييس رسم كبيرة.

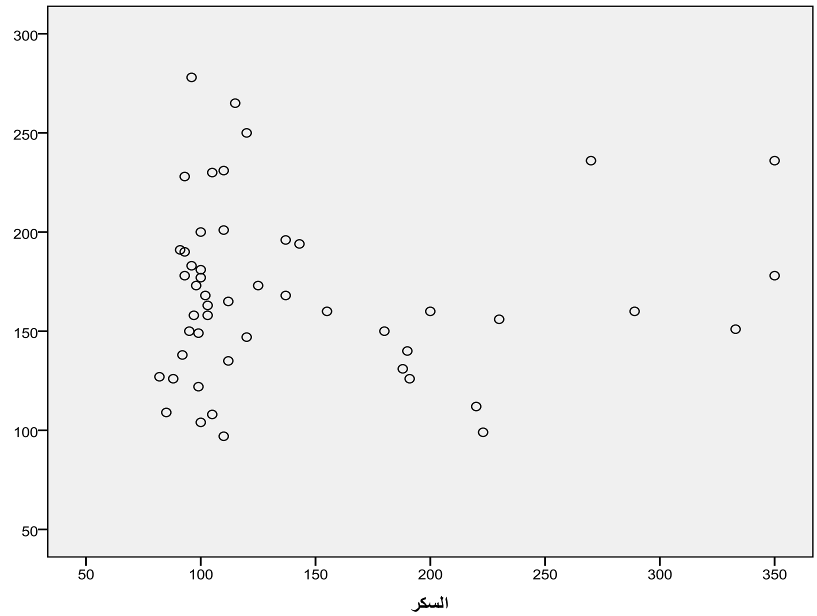
7 – مخطط الانتشار :

لكل نوع من البيانات الأسلوب الأفضل الذي يناسبه ، فعند دراسة العلاقة بين متغيرين نرى من الأجدى أن نجعل المستوي ساحة لتمثيل العينة حيث تمثل قياسات المتغير الأول على المحور الأفقي والثاني على المحور الشاقولي، ومِنْ ثَمَّ تمثل العينة بنقاط منتشرة في المستوي مبعثرة أحياناً ومتجمعة أحياناً أخرى، يدلنا شكل توزعها على طبيعة العلاقة بين المتغيرين.

مثال (24): تمثل البيانات في الجدول الآتي قياسات معدل السكر في الدم والكوليسترول لخمسين مريضاً راجعوا مشفى الأسد الجامعي يعانون احتشاءاتٍ دماغيةً مختلفة. لأخذ تصور أولي إذا كانت هناك علاقة ظاهرة بين السكر والكولسترول لمجتمع الاحتشاءات ،نرسم شكل الانتشار ، ندخل قياسات أحد المتغيرين وليكن الأول (السكر) ونمثله على المحور الأفقي وقياسات المتغير الثاني ، ونمثله على المحور العمودي ، و نظهر شكل الانتشار للقياسات .

الكولسترول	سكر الدم	الكولسترول	سكر الدم	الكولسترول	سكر الدم	الكولسترول	سكر الدم	الكولسترول	سكر الدم
194	143	181	100	122	99	236	270	156	230
160	289	196	137	151	333	158	97	191	91
104	100	236	350	183	96	108	105	150	95
99	223	201	110	147	120	140	190	150	180
168	137	200	100	135	112	163	103	131	188
126	88	250	120	160	155	138	92	160	200
127	82	97	110	178	93	173	98	158	103
228	93	230	105	173	125	149	99	265	115
126	191	109	85	165	112	168	102	178	350
278	96	190	93	231	110	112	220	177	100

جدول نسبة السكر والكوليسترول في الدم لخمسين مريضاً يعانون نشباتٍ دماغيةً مختلفةً



شكل الانتشار لمعدل الكوليسترول مع السكر في الدم لخمسين مريضاً

1 – 7 – مقاييس النزعة المركزية (Measures of central tendency):

سبق وأن تحدثنا في الفقرات السابقة عن طرائق تنظيم البيانات الإحصائية وتلخيصها في جداول توزيع تكرارية، وتمثيلها بيانياً، ومع أن هذه الطرائق كانت مفيدة جداً في توضيح شكل التوزيعات التكرارية، إلا أنه لا يمكن استخدامها دوماً وخاصةً عند المقارنة بين ظاهرتين أو أكثر، لذا دعت الحاجة إلى إيجاد مقاييس عددية تقيس لنا الظاهرة المدروسة نستطيع من خلالها من المقارنة بين ظاهرتين أو أكثر. وهذه المقاييس التي سوف ندرسها في هذه الفقرة هي مقاييس النزعة المركزية وهي عبارة عن قيم مثلى تقترب منها معظم البيانات الإحصائية، أو تتركز حولها، أو تتوزع بالقرب منها.

أولاً –الوسط الحسابي: يُعد الوسط الحسابي من أهم مقاييس النزعة المركزية على الإطلاق وأبسطها لأنه يدخل في جميع عمليات التحليل الإحصائي.

1–الوسط الحسابي لبيانات غير مبوبة:

لتكن لدينا مجموعة من البيانات $x_1, x_2, \ldots, x_n$ مأخوذة من مجتمع إحصائي ما، فإن الوسط الحسابي لهذه البيانات والذي نرمز له بالرمز $\overline{x}$ يعطى بالعلاقة الآتية:

$$\overline{x} = \frac{x_1+x_2+\cdots+x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (1)$$

مثال (25): إذا علمت أن عدد الإجازات السنوية بالأيام التي حصل عليها بعض موظفو جامعة دمشق هو: 14 ، 15 ، 18 ، 15 ، 14 ، 7 ، 18 ، 17 ، 18 ، 4 ، فأوجد الوسط الحسابي.

إن الوسط الحسابي لمجموعة من البيانات يعطى بالعلاقة:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{14+15+18+15+14+7+18+17+18+4}{10} = \frac{140}{10} = 14 \quad \text{يوم}$$

2 – الوسط الحسابي من بيانات مبوبة:

لتكن لدينا مجموعة من البيانات موزعة وفق جدول توزيع تكراري مؤلفاً من r فئة. عندئذٍ يعرَّف الوسط الحسابي لهذه البيانات بالعلاقة التالية:

$$\overline{x} = \frac{f_1x'_1+f_2x'_2+\cdots+f_rx'_r}{f_1+f_2+\cdots+f_r} = \frac{\sum_{i=1}^{r} f_i x'_i}{\sum_{i=1}^{r} f_i} = \frac{1}{n}\sum_{i=1}^{r} f_i x'_i \qquad (2)$$

حيث $f_1، f_2، \dots ، f_r$ تكرارات لفئات عددها r ومراكزها $x'_1، x'_2، \dots ، x'_r$ على الترتيب.

مثال (26): أوجد الوسط الحسابي لمجموعة البيانات الواردة في المثال رقم (16) وذلك في حالة البيانات غير المبوبة وفي حالة البيانات المبوبة ثُمَّ قارن بين النتيجتين.

1– الوسط الحسابي في حالة البيانات غير المبوبة:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{180+90+\cdots+110}{50} = \frac{9750}{50} = 195$$

2–الوسط الحسابي في حالة البيانات المبوبة:

لتسهيل الحسابات نشكل الجدول الآتي ثُمَّ نعوض في القانون فنجد:

حدود الفئات	التكرار f_i	مراكز الفئات x'_i	$f_i x'_i$
$[60-100[$	3	80	240
$[100-140[$	7	120	840
$[140-180[$	9	160	1440
$[180-220[$	15	200	3000
$[220-260[$	8	240	1920
$[260-300[$	6	280	1680
$[300-340[$	2	320	640
المجموع	50	------	9760

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{r} f_i x'_i = \frac{9760}{50} = 195.2$$

نلاحظ أن الوسط الحسابي المحسوب بطريقة البيانات غير المبوبة يختلف عنه بطريقة البيانات المبوبة، وهذا الخلاف ناتج عن أن بعض البيانات تكون بعيدة بعض الشيء عن مراكز الفئات.

3 – الوسط الحسابي الموزون:

لتكن لدينا مجموعة من البيانات $x_1، x_2، \dots ، x_n$ مأخوذة من مجتمع إحصـــائي ما، وبفرض أن تكراراتها هي على الترتيب $f_1، f_2، \dots ، f_n$ فإن الوسط الحسابي لهذه البيانات والذي نرمز له بالرمز $\overline{x}$ يعطى بالعلاقة الآتية:

$$\overline{x} = \frac{f_1x_1+f_2x_2+\cdots+f_nx_n}{f_1+f_2+\cdots+f_n} = \frac{1}{n}\sum_{i=1}^{n} f_i x_i \qquad (3)$$

مثال (27): أوجد الوسط الحسابي المرجح لدرجات طالب في أربع مواد درجاتهم معطاة بالقيم 95، 80، 45، 60 وكانت ساعات الدراسة الأسبوعية لهذه المواد على الترتيب 4، 5، 2، 3 .

الحل: إن الوسط الحسابي يعطى بالعلاقة:

$$\overline{x} = \frac{f_1x_1+f_2x_2+\cdots+f_nx_n}{f_1+f_2+\cdots+f_n}$$

$$= \frac{4\times95+5\times80+2\times45+3\times60}{4+5+2+3} = \frac{1050}{14} = 75 \quad \text{درجة}$$

ملاحظة (3): يمكن اعتبار الوسط الحسابي لبيانات مبوبة وسطاً مرجحاً وذلك باستبدال مراكز الفئات $x'_1، x'_2، \dots، x'_r$ بالبيانات $x_1، x_2، \dots، x_n$ وتكرارات مراكز الفئات $f_1، f_2، \dots، f_r$ بتكرارات البيانات $f_1، f_2، \dots، f_n$ فنجد:

$$\overline{x} = \frac{f_1x_1+f_2x_2+\cdots+f_nx_n}{f_1+f_2+\cdots+f_n} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{1}{n}\sum_{i=1}^{n} f_i x_i$$

3–محاسن الوسط الحسابي ومساوئه:

آ – **محاسن الوسط الحسابي:**

1– سهولة حسابه في أية مجموعة من البيانات، لذا يعتبر من أشهر المتوسطات.

2– يأخذ جميع القيم في الحسبان، ولا يحتاج في حسابه إلى ترتيب البيانات.

3– إن وسط مجموعة من البيانات لا يتغير، وهذه ميزة هامة في أسلوب المعاينات.

4– مجموع انحرافات مجموعة من البيانات عن وسطها الحسابي يساوي الصفر دوماً، ومجموع مربعات انحرافاتها عن وسطها الحسابي أقل من مجموع مربعات الانحرافات عن أي عدد آخر.

ب – مساوئ الوسط الحسابي:

1– يتأثر بالقيم المتطرفة للبيانات مما يفقده معناه وأهميته.

2– لا يمكن حسابه في حالة البيانات الوصفية.

3– لا يساوي أياً من القيم الداخلة في حسابه، فقد يحتوي على عدد كسري لبيانات مكونة من أعداد صحيحة، وذلك في حالة البيانات المنفصلة، مثل عدد المواليد في مجتمع ما.

ثانياً – الوسيط: يعرَّف الوسيط لمجموعة من البيانات الإحصائية بأنه القيمة العددية التي تقسم تلك البيانات إلى مجموعتين متساويتين بعد أن نقوم بترتيب البيانات ترتيباً تصاعدياً أو تنازلياً.

1 –الوسيط للبيانات غير المبوبة:

لإيجاد الوسيط للبيانات غير المبوبة، نقوم بترتيب البيانات ترتيباً تصاعدياً أو تنازلياً، ويكون الوسيط هو القيمة التي يكون ترتيبها $\frac{n+1}{2}$ عندما يكون عدد البيانات فردياً.

أي أن الوسيط يعطى بالشكل: الوسيط = القيمة التي ترتيبها $\left(\frac{n+1}{2}\right)$

ويكون الوسيط هو القيمة الناتجة عن متوسط القراءتين اللتين ترتيبهما $\frac{n}{2}$، $\frac{n}{2}+1$ عندما يكون عدد البيانات زوجياً. أي أن الوسيط يعطى بالشكل:

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

مثال (28): أوجد الوسيط لعدد الإجازات المرضية السنوية بالأيام التي حصل عليها سبعة موظفين من موظفي جامعة القلمون الخاصة هي: 25، 15 ، 16 ، 20 ، 20 ، 24 ، 25.

الحل: نرتب البيانات ترتيباً تصاعدياً كما يأتي: 15 ، 16 ، 20 ، 20 ، 24 ، 25 ، 25.

لدينا عدد البيانات n = 7 يوماً وهو عدد فردي، ومن ثُمَّ فإن الوسيط هو القيمة التي ترتيبها:

$$\frac{n+1}{2} = \frac{7+1}{2} = 4$$

أي أن الوسيط هو القيمة الرابعة وهي القيمة 20 يوماً.

مثال (29): إذا كانت أطوال ثُمَّانية من طلاب جامعة القلمون الخاصة هي:

180، 172، 175، 179، 185، 170، 177، 182 بالسنتمترات، فأوجد الوسيط.

الحل: نقوم بترتيب هذه البيانات ترتيباً تصاعدياً كما يأتي:

185 ، 182 ، 180 ، 179 ، 177 ، 175 ، 172 ، 170

نلاحظ أن عدد البيانات هو 8 وهو عدد زوجي، لذلك فإن الوسيط هو القيمة الناتجة عن متوسط القراءتين:

$$\frac{n}{2} = \frac{8}{2} = 4 \quad , \quad \frac{n}{2}+1 = \frac{8}{2}+1 = 5$$

أي هو متوسط القيمتين الرابعة والخامسة ويساوي:

$$\frac{1}{2}(177+179) = 178 \quad \text{سم}$$

2 – الوسيط للبيانات المبوبة:

لإيجاد الوسيط لمجموعة من البيانات المبوبة نتبع الخطوات الآتية:

1– نوجد جدول التوزيع التكراري المتجمع الصاعد.

2– نوجد ترتيب الوسيط وهو عبارة عن نصف مجموع القراءات أي $\frac{n}{2}$.

3– نحدد الفئة الوسيطية من جدول التوزيع التكراري المتجمع الصاعد وهي الفئة التي تحوي على ترتيب الوسيط، وذلك بأن نبحث في عمود التكرار الصاعد عن القيمة $\frac{n}{2}$ ونضع خطاً أفقياً بين التكرارين اللذين تقع بينهما القيمة $\frac{n}{2}$.

4– نحدد التكرار المتجمع الصاعد للفئة السابقة للفئة الوسيطية $\uparrow f_{i-1}$.

5– نحدد التكرار المطلق أو التكرار الأساسي للفئة الوسيطية.

6– نوجد الوسيط من العلاقة:

$$Median = A + \left(\frac{\frac{n}{2} - f_{i-1}\uparrow}{f_i}\right) \times l \qquad (4)$$

حيث: A : هي الحد الأدنى للفئة الوسيطية.

$f_{i-1}\uparrow$: التكرار المتجمع الصاعد السابق للفئة الوسيطية.

f_i : التكرار المطلق أو التكرار الأساسي للفئة الوسيطية.

l : طول الفئة.

مثال (30): أوجد الوسيط للبيانات الواردة في المثال (16) .

الحل: نوجد جدول التوزيع التكراري المتجمع الصاعد بالشكل التالي:

حدود الفئات	التكرار	التكرار الصاعد $f_i\uparrow$
$[60-100[$	3	3
$[100-140[$	7	10
$[140-180[$	9	$f_{i-1}\uparrow$ 19
$[180-220[$	f_i 15	34
$[220-260[$	8	42
$[260-300[$	6	48
$[300-340[$	2	50
المجموع	50	—

ثُمَّ نعوض في القانون التالي فنجد:

$$Median = A + \left(\frac{\frac{n}{2} - f_{i-1}\uparrow}{f_i}\right) \times l$$

$$= 180 + \left(\frac{25-19}{15}\right) \times 40$$

$$= 180 + \frac{240}{15} = 180 + 16 = 196$$

ملاحظة (4): يمكننا حساب الوسيط للبيانات المبوبة من جدول التوزيع التكراري المتجمع الهابط.

ملاحظة (5): يمكن إيجاد الوسيط بيانياً من المنحنين المتجمع الصاعد والهابط أو من تقاطع المنحنيين في رسم واحد.

ففي حالة المنحني المتجمع الصاعد تحدد نقطة $\frac{n}{2}$ على المحور الشاقولي ونرسم منها خطاً أفقياً موازياً للمحور الأفقي وهو محور الفئات إلى أن يلتقي بالمنحني في نقطة. نسقط من تلك النقطة عموداً يلاقي المحور الأفقي في نقطة تكون قيمتها هي قيمة الوسيط بيانياً.

وفي حالة المنحني المتجمع الهابط نتبع الخطوات السابقة نفسها التي اتبعناها في حالة المنحني المتجمع الصاعد بيانياً.

أما في حالة تقاطع المنحنيين الصاعد والهابط في نقطة، نسقط من هذه النقطة عموداً على المحور الأفقي فتكون نقطة تقاطعه هي قيمة الوسيط المطلوب.

4 –محاسن الوسيط ومساوئه:

آ – محاسن الوسيط:

1- تعريفه واضح وسهل ولا يتأثر بالقيم المتطرفة أو الشاذة.

2- لا تتغير قيمة الوسيط إذا غيرنا جميع القيم التي قبله أو بعده.

3- يمكن إيجاد الوسيط للبيانات الوصفية التي لها صفة الترتيب.

ب – مساوئ الوسيط:

1- لا يأخذ جميع القيم في الحسبان.

2- عدم اتصافه بميزات جبرية وإن استعماله محدود في الأساليب الإحصائية.

3- لا يصلح لقياس النزعة المركزية عندما يكون عدد البيانات صغيراً ومن ثَمَّ فإن الوسيط المحسوب هو قيمة غير ثابتة تختلف اختلافاً كبيراً عند إضافة بيانات جديدة إلى القيمة.

ثالثاً –المنوال: يعرَف المنوال لمجموعة من البيانات بأنه القيمة الأكثر تكراراً وقد يكون المنوال موجوداً أو لا، وإن وجد فهو ليس وحيداً، أي قد يكون لمجموعة من البيانات أكثر من منوال.

1 – المنوال للبيانات غير المبوبة:

لإيجاد المنوال، نوجد القيمة الأكثر تكراراً مباشرة من تلك البيانات، دون ترتيب هذه البيانات.

مثال (31): أوجد المنوال للبيانات الآتية التي تدل على أطوال عشرة أشخاص:

175 ، 180 ، 175 ، 172 ، 175 ، 180 ، 181 ، 170 ، 168 ، 165 .

الحل: نلاحظ أن الطول 175 هو أكبر تكرار، ومن ثُمَّ فإن المنوال في هذه الحالة هو 175 سم.

2 – المنوال للبيانات المبوبة:

لإيجاد المنوال للبيانات المبوبة نتبع الخطوات الآتية:

1- نحدد الفئة المنوالية وهي الفئة التي يقابلها أكبر تكرار، ونرمز لتكرارها بالرمز f.

2- نوجد الحد الأدنى للفئة المنوالية ونرمز له بالرمز A.

3- نوجد التكرار السابق واللاحق للفئة المنوالية f_2 ، f_1 على الترتيب، ثُمَّ نوجد طول الفئة المنوالية وهو طول الفئة المحسوب سابقاً l .

4- نوجد المنوال من العلاقة:

$$Mod = A + \left(\frac{f-f_1}{2f-f_1-f_2}\right) \times l \qquad (5)$$

حيث: A : هي الحد الأدنى للفئة المنوالية.

f : تكرار الفئة المنوالية.

f_1 : التكرار السابق للفئة المنوالية.

f_2 : التكرار اللاحق للفئة المنوالية.

l : طول الفئة.

مثال (32): أوجد المنوال لمجموعة البيانات الواردة في المثال (16) من الفصل الحالي.

الحل: نوجد جدول التوزيع التكراي لفئات القيم كما في الشكل:

التكرار	حدود الفئات
3	$[60-100[$
7	$[100-140[$
f_1 9	$[140-180[$
f 15	$[180-220[$
f_2 8	$[220-260[$
6	$[260-300[$
2	$[300-340[$
50	المجموع

نلاحظ من جدول التوزيع التكراري أن الفئة المنوالية هي الفئة الرابعة وأن تكرارها $f=15$ ، كما وأن التكرار السابق هو $f_1=9$ والتكرار اللاحق هو $f_2=8$ وطول الفئة هو $l=40$ وأخيراً فإن الحد الأدنى للفئة المنوالية هو $A=180$ ، ثُمَّ نوجد المنوال من العلاقة:

$$Mod = A + \left(\frac{f-f_1}{2f-f_1-f_2}\right)\times l$$

$$= 180 + \left(\frac{15-9}{30-9-8}\right)\times 40$$

$$= 180 + \frac{240}{13} = 180 + 18.46 = 196.46$$

ملاحظة (6): يمكن إيجاد المنوال بيانياً وذلك برسم ثلاث مستطيلات من المدرج التكراري فقط وهي المستطيل الممثل لأكبر تكرار ثُمَّ المستطيل السابق والمستطيل اللاحق. نصل الزاوية العليا اليمنى للمستطيل السابق بالزاوية العليا اليمنى للمستطيل الممثل لأكبر تكرار ثُمَّ نصل الزاوية العليا اليسرى للمستطيل اللاحق بالزاوية العليا اليسرى للمستطيل الممثل لأكبر تكرار، وأخيراً ننزل عمود من نقطة تقاطعهما على المحور الأفقي فتكون نقطة التقاطع هي قيمة المنوال.

3 – محاسن المنوال ومساوئه:

آ: **محاسن المنوال:**

1- لا يتأثر بالقيم المتطرفة أو الشاذة.

2- سهولة فهمه وتعينه بحيث لا يتطلب أكثر من تعريفه لجعل معناه واضحاً.

3- يمكن إيجاده من بيانات وصفية وكذلك من توزيعات تكرارية مفتوحة.

ب: مساوئ المنوال:

1- لا يأخذ جميع القيم في الحسبان.

2- قد يكون المنوال غير موجودٍ.

3- قد يكون لمجموعة من البيانات أكثر من منوال، وبذلك يكون المنوال متعدد القيم، ومن ثُمَّ يصعب التعامل معه في التحليل الإحصائي.

4 –العلاقة بين الوسط الحسابي والوسيط والمنوال:

نلاحظ أنه في حالة التوزيعات التكرارية المتماثلة وحيدة المنوال نجد أن المقاييس الثلاثة تكون متطابقة، أي أن قيم الوسط الحسابي والوسيط والمنوال تكون متساوية. ولكن في حالة عدم التماثل، أي عند وجود التواء نحو اليمين أو نحو اليسار فإن قيم هذه المقاييس تختلف عن بعضها البعض، ويكون الوسط الحسابي أكبر المقاييس السابقة في حالة الالتواء نحو اليمين، يليه الوسيط، ثُمَّ المنوال وهو أصغر المقاييس الثلاثة في هذه الحالة. وإذا كان الالتواء نحو اليسار فإننا نجد أن الوسط الحسابي يكون أصغر المقاييس السابقة ويليه الوسيط ثُمَّ المنوال وهو أكبر المقاييس الثلاثة في هذه الحالة.

أما في حالة الالتواء البسيط نحو اليمين أو اليسار فإن هناك علاقة تجريبية بين هذه المقاييس:

$$\left(\text{الوسط الحسابي} - \text{المنوال}\right) = 3\left(\text{الوسط الحسابي} - \text{الوسيط}\right)$$

وتصبح هذه العلاقة غير صحيحة في حالة الالتواء الحاد.

وهذه العلاقة تعطينا طريقة أخرى تقريبية لحساب المنوال لا تقل دقة عن الطرائق السابقة، كما تساعدنا على حساب الوسط الحسابي في حالة التوزيعات التكرارية المفتوحة.

رابعاً – الوسط الهندسي: لاحظنا سابقاً أن الوسط الحسابي يتأثر بالقيم الشاذة، أي القيم الصغيرة جداً، أو القيم الكبيرة جداً مقارنة ببقية البيانات، لذلك دعت الحاجة إلى إيجاد مقاييس عددية تكون أقل تأثيراً بالقيم الشاذة وخاصة القيم الكبيرة جداً مقارنة ببقية البيانات، ومن هذه المقاييس الوسط الهندسي الذي يعطي قيماً أدق من الوسط الحسابي في دراسة بعض الظواهر الطبيعية أو الاقتصادية التي تزيد مفرداتها بمعدلات ثابتة مثل ظاهرتي النمو السكاني والاقتصادي وغيرها.

ويمتاز الوسط الهندسي عن الوسط الحسابي بأنه أقل تأثيراً بالقيم المتطرفة في البيانات، وأن الوسط الهندسي لمجموعة من البيانات دوماً أقل من وسطها الحسابي.

1 –الوسط الهندسي لبيانات غير مبوبة:

يعرَّف الوسط الهندسي لقيمتين x_1،x_2 بأنه الجذر التربيعي الموجب لحاصل جدائهما، فإذا رمزنا للوسط الهندسي بالرمز G الوسط الهندسي للقيمتين x_1،x_2 يعطى بالعلاقة:

$$G = \sqrt[2]{x_1 . x_2}$$

وكذلك فإن الوسط الهندسي لثلاثة قيم x_1،x_2،x_3 يعطى بالعلاقة:

$$G = \sqrt[3]{x_1 . x_2 . x_3}$$

وبشكل عام إذا كانت لدينا مجموعة من البيانات $x_1, x_2, \dots, x_n$ مأخوذة من مجتمع إحصائي ما، فإن الوسط الهندسي G لهذه البيانات يعطى بالعلاقة:

$$G = \sqrt[n]{x_1 . x_2 \dots x_n}$$

أي أن الوسط الهندسي لهذه البيانات والتي عددها n هو الجذر النوني لحاصل جدائهما، وإذا أخذنا اللوغاريتم العشري للطرفين في العلاقة الأخيرة فإننا نجد:

$$LogG = \frac{1}{n} Log(x_1 . x_2 \dots x_n) = \frac{1}{n} \sum_{i=1}^{n} Logx_i \qquad (6)$$

2 –الوسط الهندسي لبيانات مبوبة:

يعطى الوسط الهندسي في حالة البيانات المبوبة بالعلاقة الآتية:

$$LogG = \frac{1}{n} \sum_{i=1}^{r} f_i Logx'_i$$

حيث f_1،f_2، ...،f_r تكرارات لفئات عددها r ومراكزها x'_1،x'_2، ...،x'_r على الترتيب.

ولإيجاد الوسط الهندسي من بيانات مبوبة نتبع الخطوات الآتية:

1- نحسب اللوغاريتم العشري لمراكز الفئات.

2- نضرب لوغاريتم مركز الفئة بتكرار تلك الفئة ثُمَّ نحسب المجموع.

3- نقسم مجموع جداء لوغاريتم مركز الفئة بتكرار تلك الفئة على مجموع التكرارات، فنحصل على لوغاريتم الوسط الهندسي.

4- نبحث عن مقابل لوغاريتم G فنحصل على قيمة الوسط الهندسي.

مثال (33): أوجد الوسط الهندسي للبيانات التالية: 12 ، 10 ، 7 ، 6 ، 6 ، 5 ، 3 .

الحل:

نلاحظ أن هذه البيانات غير مبوبة ومن ثُمَّ فإن الوسط الهندسي يعطى بالعلاقة:

$$LogG = \frac{1}{7}(Log3 + Log5 + 2Log6 + Log7 + Log10 + Log12)$$
$$= \frac{1}{7}(0.477 + 0.699 + 1.556 + 0.845 + 1 + 1.079) = 0.808$$

ومن جدول اللوغاريتم نجد أن الوسط الهندسي هو: $G = 6.43$.

- وبحساب الوسط الحسابي نجد:

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{3+5+6+6+7+10+12}{7} = \frac{49}{7} = 7$$

نلاحظ أن الوسط الهندسي أصغر من الوسط الحسابي وهذا يوضح حقيقة أن الوسط الهندسي لمجموعة أرقام موجبة غير متساوية، أصغر من وسطها الحسابي.

مثال (34): أوجد الوسط الهندسي للبيانات المبوبة في الجدول الآتي:

الفئـات	f_i	x'_i	$Logx'_i$	$f_iLogx'_i$
$[8-12[$	10	10	1	10
$[12-16[$	25	14	1.146	28.65
$[16-20[$	50	18	1.255	62.75
$[20-24[$	10	22	1.342	13.42
$[24-28[$	5	26	1.414	7.07
المجموع	100			121.89

نلاحظ أن هذه البيانات مبوبة ومن ثُمَّ فإن الوسط الهندسي يعطى من العلاقة:

$$LogG = \frac{1}{n}\sum_{i=1}^{r} f_iLogx'_i$$

$$= \frac{1}{100}(121.89) = 1.2189$$

ومن جدول اللوغاريتم نجد قيمة الوسط الهندسي G فنجد:

$$G = 16.56 \quad \text{وحدة}$$

3 –محاسن الوسط الهندسي ومساوئه:

آ –محاسن الوسط الهندسي:

1– يُعدُّ المتوسط الوحيد الذي يعطي نتائج سليمة رياضياً عند حساب متوسط المعدلات الزمنية.

2– هو متوسط محدد رياضياً بدقة، حيث يتصف بميزات جبرية معينة.

3– من فوائده الهامة قياس متوسط نسب التغير كما هو الحال في تغيرات النمو.

ب –مساوئ الوسط الهندسي:

1– صعوبة فهمه وعدم إمكانية تحديده في القيم السالبة.

2– يعتبر من أكثر المتوسطات تعقيداً لأنه يعتمد في حسابه على الآلات الحاسبة.

خامساً – الوسط التوافقي: يُعدُّ الوسط التوافقي من المقاييس التي تحد من تأثير القيم الشاذة وخاصة في حالة القيم الكبيرة مقارنة ببقية البيانات. ويلاحظ أن تأثير الوسط التوافقي أكبر من تأثير الوسط الهندسي في الحد من القيم الشاذة نحو الكبر. لأن قيمته تكون أصغر من قيمة الوسط الهندسي لنفس مجموعة البيانات.

1–الوسط التوافقي لبيانات غير مبوبة:

إذا كانت لدينا مجموعة من البيانات x_1،x_2، ... ،x_n مأخوذة من مجتمع إحصائي ما، فإن الوسط التوافقي والذي نرمز له بالرمز H يعطى بالعلاقة الآتية:

$$\frac{1}{H} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{x_i}\right) \qquad (7)$$

أو بالعلاقة:

$$H=\frac{n}{\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)}=\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}}$$

2–الوسط التوافقي لبيانات مبوبة:

يعرَف الوسط التوافقي في حالة البيانات المبوبة بالعلاقة الآتية:

$$\frac{1}{H}=\frac{1}{f_1+f_2+\cdots+f_r}\left(\frac{f_1}{x_1'}+\frac{f_2}{x_2'}+\cdots+\frac{f_r}{x_r'}\right)=\frac{1}{\sum_{i=1}^{r}f_i}\sum_{i=1}^{r}\left(\frac{f_i}{x_i'}\right)=\frac{1}{n}\sum_{i=1}^{r}\left(\frac{f_i}{x_i'}\right)$$

حيث $f_1، f_2، \dots، f_r$ تكرارات لفئات عددها r ومراكزها $x_1'، x_2'، \dots، x_r'$ على الترتيب.

أو بالعلاقة:

$$H=\frac{1}{\frac{\frac{f_1}{x_1'}+\frac{f_2}{x_2'}+\cdots+\frac{f_r}{x_r'}}{f_1+f_2+\cdots+f_r}}=\frac{\sum_{i=1}^{r}f_i}{\sum_{i=1}^{r}\left(\frac{f_i}{x_i'}\right)}$$

مثـــــال (35): لنفترض أن طبيب الأسنان يستطيع أن يكشف عن 6 أسنان في الساعة وأن يقلع 4 أسنان في الساعة وأن يداوي 3 أسنان في الساعة. فما هو متوسط عدد الأسنان التي يمكن معالجتها في الساعة.

الحل: طريقة أولى:

الزمن اللازم للكشف عن السن $=\frac{1}{6}$ الساعة $=10$ دقائق

الزمن اللازم لقلع السن $=\frac{1}{4}$ الساعة $=15$ دقيقة

الزمن اللازم لمداواة السن $=\frac{1}{3}$ الساعة $=20$ دقيقة

وبالتالي فإن متوسط الزمن اللازم لمعالجة أي سن هو:

$$\frac{10+15+20}{3}=\frac{45}{3}=15 \quad \text{دقيقة}$$

ومتوسط عدد الأسنان الممكن معالجتها بالساعة هو: $\frac{60}{15}=4$ أسنان

طريقة ثانية: باستخدام علاقة الوسط التوافقي التالية نجد:

$$H=\frac{n}{\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)}=\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}}=\frac{3}{\frac{1}{6}+\frac{1}{4}+\frac{1}{3}}=\frac{3\times 24}{18}=4 \quad \text{أسنان}$$

وهو عبارة عن متوسط عدد الأسنان التي يمكن مداواتها في الساعة.

مثال (36): احسب الوسط التوافقي H للبيانات المعرفة في المثال (33) السابق.

الحل:

نلاحظ أن هذه البيانات غير مبوبة ومن ثَمَّ فإن الوسط التوافقي يعطى من العلاقة:

$$\frac{1}{H}=\frac{1}{n}\sum_{i=1}^{r}\left(\frac{f_i}{x_i}\right)$$

$$=\frac{1}{7}\left(\frac{1}{3}+\frac{1}{5}+\frac{1}{6}+\frac{1}{6}+\frac{1}{7}+\frac{1}{10}+\frac{1}{12}\right)=\frac{501}{2940}=5.87$$

- لقد وجدنا سابقاً أن الوسط الحسابي لهذه البيانات هو $\overline{x} = 7$ وأن الوسط الهندسي هو $G = 6.43$. بالمقارنة نجد أن الوسط التوافقي للبيانات نفسها $H = 5.87$ وهو أصغر المتوسطات الثلاثة.

مثال (37): قامت إحدى المستشفيات بشراء 800 علبة شاش من ثلاث مخازن مختلفة كما يلي:

رقم المخزن	عدد العلب المشتراة بعشر ليرات x	الكمية f
1	10	110
2	12	240
3	15	450

والمطلوب أحسب الوسط التوافقي لعدد العلب الشاش الممكن شراؤها بعشر ليرات سورية.

الحل: إن الوسط التوافقي يعطى بالشكل:

$$H = \frac{1}{\frac{\frac{f_1}{x_1}+\frac{f_2}{x_2}+\cdots+\frac{f_r}{x_r}}{f_1+f_2+\cdots+f_r}} = \frac{\sum_{i=1}^{r} f_i}{\sum_{i=1}^{r}\left(\frac{f_i}{x_i}\right)} = \frac{1}{\frac{\left(\frac{110}{10}\right)+\left(\frac{240}{12}\right)+\left(\frac{450}{15}\right)}{110+240+450}} = \frac{800}{11+20+30} = 13.115 \quad \text{ل.س}$$

أي أن متوسط عدد العلب الممكن شراؤها بعش ليرات سورية هو 13 علبة تقريباً.

3 – محاسن الوسط التوافقي ومساوئه:

آ – محاسن الوسط التوافقي:

1- يأخذ جميع القيم بالحسبان. وهو أصغر من الوسط الهندسي.

2- يعد الوسط التوافقي أفضل المتوسطات في حالة إيجاد معدلات النسب ومعدلات الإنتاج.

3 – الوسط التوافقي أقل تأثيراً بالقيم الشاذة الكبيرة من الوسط الحسابي.

ب – مساوئ الوسط التوافقي:

1- يصعب حسابه في البيانات الوصفية والبيانات الكمية ذات التوزيعات التكرارية المفتوحة.

2- لا يساوي في الغالب أياً من القيم الداخلة في حسابه.

ملاحظة: عندما نريد أن نعطي فكرة واضحة ودقيقة عن ظاهرة ما، فإننا نسعى إلى حساب عدة أنواع من المتوسطات، لأنه في الحقيقة لكل نوع من المتوسطات استعمالاته وفوائده وذلك حسب الحاجة والهدف، والاكتفاء بنوع واحد من المتوسطات ربما لا يكون كافياً لوصف التوزيع التكراري الذي يمثل الظاهرة المدروسة بشكل دقيق، لذا نلجأ إلى حساب مقاييس أخرى.

1-7-مقاييس التشتت (Measures of dispersion):

لقد تحدثنا عن طرائق تلخيص وتنظيم البيانات الإحصائية وعرضها بصورها المختلفة الجدولية والبيانية ثُمَّ تناولنا بعد ذلك طريقة حساب مقاييس النزعة المركزية وذلك من أجل إيجاد قيم عددية محددة تصف هذه البيانات بأشكالها المختلفة ولكن هذه المقاييس تكون غير كافية من أجل توضيح مقدار التفاوت بين مفردات المشاهدة للظاهرة المدروسة كما في المثال الآتي:

مثال (38): في دراسة عن الأضرار التي لحقت بالمشافي العامة وذلك بسبب أعطال الأجهزة الطبية نتيجة انقطاع التيار الكهربائي المتكرر ونتيجة ضعف التيار الكهربائي، وقد تم الحصول على البيانات التالية من 25 مشفى في القطر العربي السوري التي تدل على الأعطال:

عدد الأعطال نتيجة انقطاع التيار الكهربائي				
1	7	7	6	1
2	2	1	7	2
1	3	2	7	5
6	1	7	4	1
5	7	6	3	6

وكذلك وجدنا:

عدد الأعطال نتيجة ضعف التيار الكهربائي				
1	2	4	4	7
3	3	2	4	5
2	4	3	5	3
4	4	3	6	5
5	6	4	6	5

لنوجد جدول التوزيع التكراري (التكرار النسبي) لهاتين المجموعتين من البيانات بالشكل التالي:

عدد الأعطال	التكرار	التكرار النسبي
1	6	0.24
2	4	0.16
3	2	0.08
4	1	0.04
5	2	0.08
6	4	0.16
7	6	0.24

يمثل هذا الجدول أعطال الأجهزة الطبية نتيجة انقطاع التيار الكهربائي، ويمثل الجدول التالي أعطال الأجهزة الطبية نتيجة ضعف التيار الكهربائي.

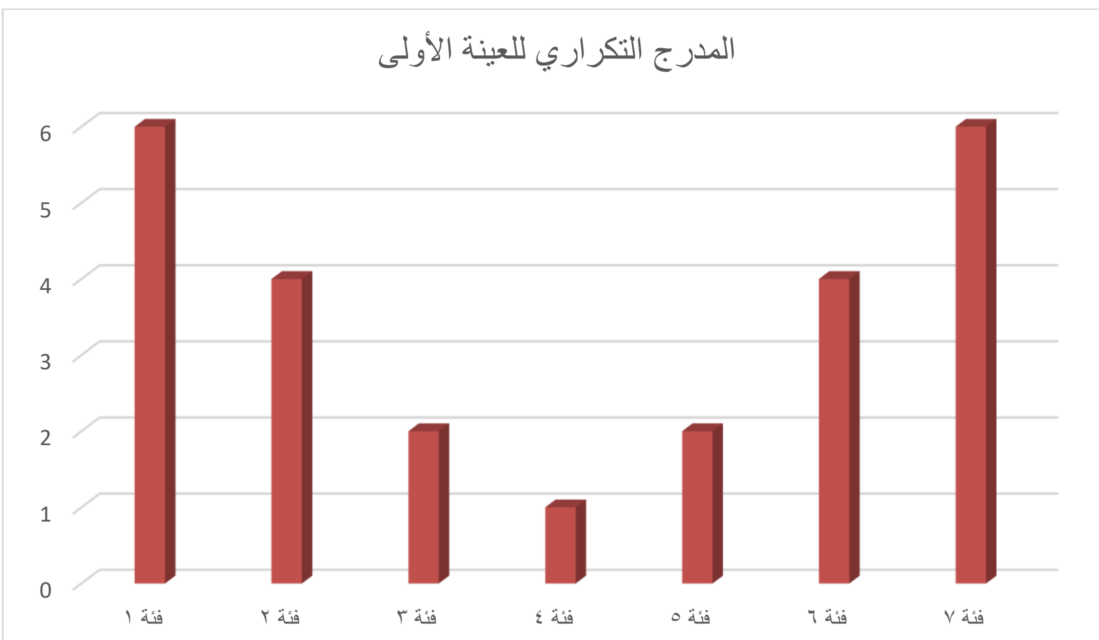
عدد الأعطال	التكرار	التكرار النسبي
1	1	0.04
2	3	0.12
3	5	0.20
4	7	0.28
5	5	0.20
6	3	0.12
7	1	0.04

لنوجد الآن الوسط الحسابي لهاتين العينتين. نرمز للوسط الحسابي للعينة الأولى بالرمز $\overline{x}$ وللوسط الحسابي للعينة الثانية بالرمز $\overline{y}$ ونكتب:

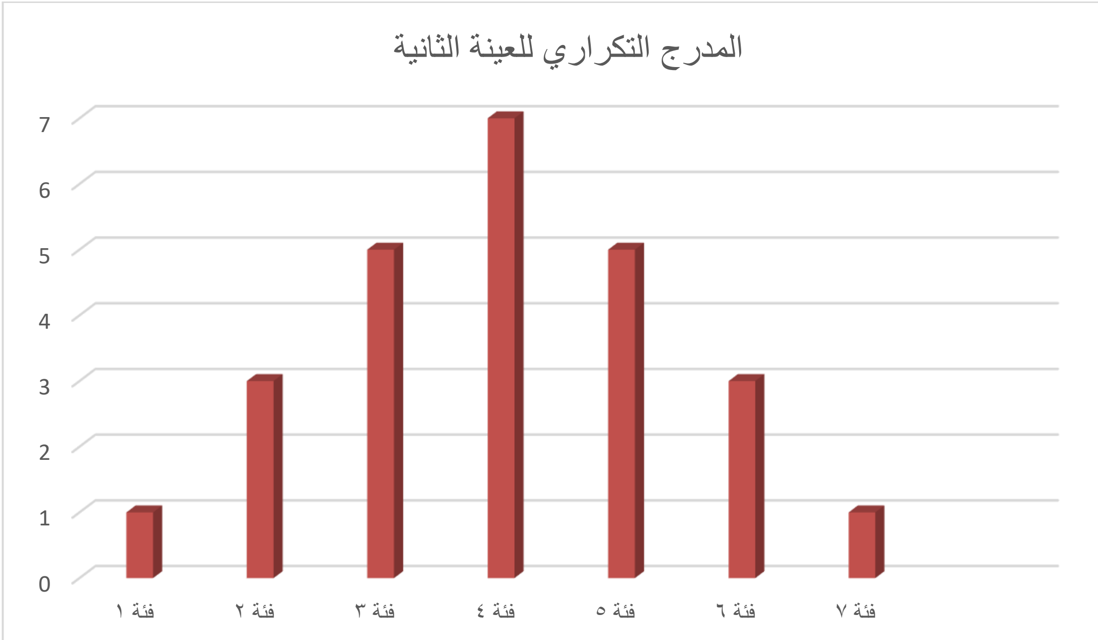
$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1+7+7+\cdots+3+6}{25} = 4 \quad \text{أعطال}$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1+2+4+\cdots+6+5}{25} = 4 \quad \text{أعطال}$$

لنوجد الآن المدرج التكراري لكل من الجدولين السابقين:



نلاحظ من هذا المثال أنه على الرغم من اختلاف عدد الأعطال الناتجة بسبب انقطاع وضعف التيار الكهربائي في هذه المشافي، إلا أن متوسط الأعطال هو نفسه. كما ونلاحظ أيضاً أن المدرجين التكراريين لهاتين العينتين مختلفان تماماً. أي أن لهاتين العينتين توزيعين مختلفين تماماً على الرغم من أن لهما الوسط الحسابي نفسه.



نستنتج من هذا كله أنه على الرغم من ميل البيانات الإحصائية لعينة ما للتجمع حول وسطها الحسابي، نجد في الوقت نفسه تقترب وتبتعد عن هذه القيمة بمقادير مختلفة، وبالتالي فإننا نلاحظ أن مثل هذه الخواص لمقاييس النزعة المركزية يجعلها غير كافية لوصف البيانات من حيث تباعد وتبعثر البيانات الإحصائية للمجموعة الواحدة بعضها عن بعض. لهذا السبب دعت الحاجة إلى إيجاد مقاييس عددية أخرى لقياس مقدار هذا التفاوت أو هذا التباعد بين البيانات. وتسمى مثل هذه المقاييس مقاييس التشتت، وسوف نقدم في هذه الفقرة كيفية حساب بعض أهم خصائص مقاييس التشتت، وبالتحديد مقاييس المدى ونصف المدى الربيعي، والتباين والانحراف المعياري، ومعاملي الاختلاف والالتواء. وسنتناول أيضاً بعض المقاييس الأخرى ذات العلاقة بمقاييس التشتت مثل مقاييس الالتواء، ومقاييس التفلطح في آخر هذا الفصل.

أولاً –المدى:

1–المدى في حالة البيانات غير المبوبة: وهو الفرق بين أكبر قيمة وأصغر قيمة. ونكتب:

المدى = أكبر قيمة – أصغر قيمة

مثال (39): أوجد المدى لعدد المخالفات الانضباطية لعينة من الجنود مكونة من عشرة جنود في كتيبة ما خلا ل عام وكانت: 65، 55، 77، 89، 90، 99، 80، 60، 88، 70.

نلاحظ أن أكبر عدد من المخالفات هو55 مخالفة وأن أصغر عدد من المخالفات هو 99 مخالفة. فيكون المدى هو 44 .

2–المدى في حالة البيانات المبوبة: فيوجد أكثر من تعريف نذكر منها التعريفين الآتيين:

التعريف الأول: المدى هو عبارة عن الفرق بين مركز الفئة العليا ومركز الفئة الدنيا. أي أن:

المدى = مركز الفئة العليا – مركز الفئة الدنيا

التعريف الثاني: المدى هو عبارة عن الفرق بين الحد الأدنى للفئة الدنيا والأعلى للفئة العليا أي:

المدى= الحد الأعلى للفئة العليا– الحد الأدنى للفئة الدنيا

مثال (40): أوجد المدى للبيانات المبينة بالجدول الآتي:

حدود الفئات	[10 – 20[	[20 – 30[	[30 – 40[	[40 – 50[	[50 – 60[	[60 – 70[
التكرار	2	4	6	10	6	4

الحل: نلاحظ من الجدول السابق أن مركز الفئة الدنيا يساوي 15 ومركز الفئة العليا يساوي 65.

الحد الأعلى للفئة العليا يساوي 69 والحد الأدنى للفئة الدنيا يساوي 10.

المدى باستخدام التعريف الأول يساوي 50 والمدى باستخدام التعريف الثاني يساوي 59.

3–محاسن المدى ومساوئه:

آ–محاسن المدى:

1– يعطي فكرة سريعة عن طبيعة البيانات الإحصائية.

2– سهولة حسابه كما ويستفاد منه في مراقبة جودة الإنتاج والأحوال الجوية.

ب–مساوئ المدى:

1– إن المدى مقياس تقريبي لا يُعتمد عليه، لأنه يعتمد في حسابه فقط على المفردتين الشاذتين.

2–يصعب حسابه في حالة التوزيعات التكرارية المفتوحة وفي حالة البيانات الوصفية.

ثانياً–نصف المدى الربيعي: نلاحظ مما سبق أن من أهم خصائص المدى غير المرغوب فيها تأثره بالقيم الشاذة الصغرى أو الكبرى.

لذلك دعت الحاجة إلى إيجاد مقاييس أخرى تستبعد هذه القيم الشاذة من الطرفين، ومن أهم هذه المقاييس هو نصف المدى الربيعي، والذي يمكن حسابه بعد ترتيب البيانات تصاعديا، وتقسيمها إلى أربعة أقسام يستبعد منها ربع المفردات الصغرى من ناحية، وكذلك ربع المفردات الكبرى من الناحية الأخرى.

بعد ذلك نسمي القيمة التي تكون دونها ربع المفردات بالربيع الأدنى ونرمز له بالرمز r_1 أما القيمة التي تحدد ثلاثة أرباع المفردات فتسمى بالربيع الأعلى، ونرمز له بالرمز r_3 والفرق بينهما هو عبارة عن المدى الربيعي. أما نصف المدى الربيعي، ونرمز له بالرمز r يعطى بالشكل:

$$r = \frac{r_3 - r_1}{2}$$

ويُعدُّ نصف المدى الربيعي مقياسا يستبعد القيم المتطرفة من الطرفين الأعلى والأدنى. وتسمى القراءة التي تكون دونها نصف البيانات بالربيع الثاني، ونرمز له بالرمز r_2.

1 – نصف المدى الربيعي في حالة البيانات غير المبوبة:

إذا كانت لدينا مجموعة من البيانات $x_1, x_2, \dots, x_n$ مأخوذة من مجتمع إحصائي ما، وسطها الحسابي $\bar{x}$ ، فإنه لإيجاد نصف المدى الربيعي لها نتبع الخطوات الآتية:

1– نرتب البيانات، وليكن عددها n ترتيبا تصاعديا مثلاً.

2- نوجد رتبة الربيع الأدنى r_1 وهي القراءة التي رتبتها $\frac{n}{4}$ في حالة كون n تقبل القسمة على 4، أما إذا كانت n لا تقبل القسمة على 4 فتكون قيمة الربيع الأدنى r_1 عبارة عن متوسط القراءتين اللتين يقع بينهما العدد الكسري $\frac{n}{4}$.

3- نحسب الربيع الأعلى r_3 وهي القراءة التي رتبتها $\frac{3n}{4}$ في حالة كون n تقبل القسمة على 4، أما إذا كانت n لا تقبل القسمة على 4 فتكون قيمة الربيع الأدنى r_1 عبارة عن متوسط القراءتين اللتين يقع بينهما العدد الكسري $\frac{3n}{4}$.

4- نحسب نصف المدى الربيعي r بتطبيق العلاقة السابقة ونوضح ذلك بالمثال الآتي:

مثال (41): أوجد نصف المدى الربيعي لأعمار عينة مكونة من 8 موظفين في كلية العلوم أعمارهم كانت كما يأتي:
35،39،44،15،21،22،30،50.

الحل: نرتب البيانات تصاعديا كالآتي: 50،44،39،35،30،22،21،15

رتبة الربيع الأدنى هي:

$$r_1 \text{ رتبة} = \frac{n}{4} = \frac{8}{4} = 2 \quad ، \; n = 8$$

أي أن الربيع الأدنى هو القراءة الثانية من جهة اليمين وهو: سنة $r_1 = 21$

أما رتبة الربيع الأعلى هي:

$$r_3 \text{ رتبة} = \frac{3n}{4} = \frac{24}{4} = 6$$

أي أن الربيع الأعلى هو الحد السادس من جهة اليمين، وهو: سنة $r_3 = 39$.

أما نصف المدى الربيعي فيكون:

$$r = \frac{r_3 - r_1}{2} = \frac{39-21}{2} = \frac{18}{2} = 9 \qquad \text{سنوات}$$

مثال (42): أوجد نصف المدى الربيعي لأعمار مفردات العينة المكونة من 10 موظفين في جامعة دمشق حيث إن البيانات كانت كالآتي: $18, 33, 23, 27, 50, 23, 36, 38, 45, 30$

الحل: نرتب البيانات تصاعدياً فتكون:

$$18, 23, 23, 27, 30, 33, 36, 38, 45, 50$$

رتبة الربيع الأدنى هي:

$$r_1 \text{ رتبة} = \frac{n}{4} = \frac{10}{4} = 2.5 \quad ، \; n = 10$$

وبذلك تكون قيمة الربيع الأدنى:

$$r_1 = \frac{23+23}{2} = \frac{46}{2} = 23 \qquad \text{سنة}$$

أما رتبة الربيع الأعلى هي:

$$r_3 \text{ رتبة} = \frac{3n}{4} = \frac{30}{4} = 7.5 \qquad \text{سنة}$$

أي قيمة الربيع الأعلى هي عبارة عن متوسط الحدين السابع والثامن، وتحسب بالشكل الآتي:

$$r_3 = \frac{36+38}{2} = 37 \qquad \text{سنة}$$

وعليه فإن قيمة نصف المدى الربيعي r هي:

$$r = \frac{r_3 - r_1}{2} = \frac{37-23}{2} = \frac{14}{2} = 7 \quad \text{سنوات}$$

2 – نصف المدى الربيعي في حالة البيانات المبوبة:

يتم حساب نصف المدى الربيعي

طنن

3 – محاسن نصف المدى الربيعي ومساوئه:

آ–محاسن نصف المدى الربيعي:

1– لا يتأثر بالقيم الشاذة.

2– يمكن حسابه في حالة البيانات الوصفية التي لها صفة الترتيب وفي حالة الجداول التكرارية المفتوحة.

ب –مساوئ نصف المدى الربيعي:

1–لا يأخذ جميع البيانات في الحسبان.

2–يصعب التعامل معه في التحليل الإحصائي.

ثالثاً–المئينات: وجدنا سابقاً أن الوسيط يقسم البيانات إلى جزأين متساويين، وأن المقياس التربيعي يقسم البيانات إلى أربعة أجزاء متساوية وذلك بعد ترتيب البيانات ترتيباً تنازلياً أو تصاعدياً. حيث يدعى الجزء الأول والذي يحوي ربع البيانات بالربيع الأول ونرمز له بالرمز r_1 وأما الربيع الثاني فهو الجزء الذي يحتوي على نصف البيانات ونرمز له بالرمز r_2 ، والربيع الثالث فهو الجزء الذي يحوي على ثلاثة أرباع البيانات ونرمز له بالرمز r_3 أما الربيع الرابع فهو الجزء الذي يحتوي على جميع البيانات ونلاحظ أن الربيع الثاني يقابل الوسيط.

وإذا قسمنا البيانات إلى مائة جزء متساوٍ، فإننا نسمي كل جزء بالمئين ونقول مثلاً المئين العاشر هو الجزء الذي يحتوي على عشر البيانات. وهكذا ونرمز للمئين العاشر بالرمز P_{10} ، وللمئين التسعين بالرمز P_{90} ، وهكذا......

رابعاً–التباين والانحراف المعياري: يُعدُّ التباين والانحراف المعياري من أهم مقاييس التشتت المستخدمة في أغلب المسائل الإحصائية.

تعريف: إذا كانت لدينا مجموعة من البيانات $x_1, x_2, \dots, x_n$ مأخوذة من مجتمع إحصائي ما، وسطها الحسابي $\overline{x}$ ، عندئذٍ يعرَّف التباين بالعلاقة:

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

أي أن التباين لمجموعة n من البيانات هو عبارة عن متوسط مجموع مربعات انحراف تلك البيانات عن وسطها الحسابي.

نرمز للتباين بالرمز S^2 وتتلخص فكرة حسابه في حساب الانحرافات عن أحد مقاييس النزعة المركزية، ويستخدم الوسط الحسابي فقط لهذا الغرض، كما وأن مكانته بين مقاييس التشتت كمكان الوسط الحسابي بين مقاييس النزعة المركزية.

ويســــمى الجذر التربيعي الموجب للتباين بالانحراف المعياري ونرمز له بالرمز S . حيث يُعدّ الانحراف المعياري من أهم وأدق وأفضل مقاييس التشتت على الإطلاق.

كما وأنه يمكننا تعريف التباين في حالة كون المشاهدات عبارة عن عينة عشوائية من العلاقة الآتية:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \qquad (9)$$

وبحساب الجذر التربيعي الموجب للتباين نحصل على الانحراف المعياري أي أن:

$$S = \sqrt{S^2} \qquad (10)$$

والجذر التربيعي يعطينا قياساً بنفس وحدات المتغير x .

سوف نتناول طريقة حساب التباين والانحراف المعياري للبيانات المبوبة وغير المبوبة:

1–التباين والانحراف المعياري للبيانات غير المبوبة:

إذا كانت لدينا مجموعة من البيانات x_1،x_2، ... ،x_n مأخوذة من مجتمع إحصــائي ما، وســطها الحســابي $\overline{x}$ ، فإن التباين يعطى بالعلاقة (9) والانحراف المعياري يعطى بالعلاقة (10)، وسوف نوضح طريقة الحساب في المثال الآتي:

مثال (43): أوجد التباين والانحراف المعياري لأعمار عينة من الموظفين في كلية العلوم بياناتها كالآتي: 15 ، 21 ، 22 ، 30 ، 35 ، 39 ، 44 ، 50 .

الحل: لقد وجدنا الوسط الحسابي لهذه البيانات فكان مساوياً 32, لنُكوِّن الآن الجدول الآتي:

x_i	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$
15	-17	189
21	-11	121
22	-10	100
30	-2	4
35	3	9
39	7	49
44	12	144
50	18	324
$\sum_{i=1}^{n} x_i = 256$	$\sum_{i=1}^{n}(x_i - \overline{x}) = 0$	$\sum_{i=1}^{n}(x_i - \overline{x})^2 = 1040$

لنحسب الآن التباين من العلاقة السابقة فنجد:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{7}(1040) = 148.57 \quad \text{سنة}$$

والانحراف المعياري هو:

$$S = \sqrt{148.57} = 12.19 \qquad \text{سنة}$$

نلاحظ عند حساب التباين باستخدام العلاقة (9) السابقة أنه لابد من حساب الوسط الحسابي $\overline{x}$ وطرحه من جميع القيم. وقد يكون الوسط الحسابي عدداً كسرياً مما يزيد من صعوبة الحسابات مما دعت الحاجة إلى إيجاد صيغ أخرى تكون أبسط في الحساب وذلك كما يأتي:

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right] = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n(\overline{x})^2\right) \qquad (11)$$

2–التباين والانحراف المعياري للبيانات المبوبة:

يعرَف التباين لمجموعة من البيانات x_1، x_2، ... ، x_n ، وسطها الحسابي $\overline{x}$ بالعلاقة التالية:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{r} f_i(x_i' - \overline{x})^2 \qquad (12)$$

حيث f_1، f_2، ... ، f_r تكرارات لفئات عددها r ومراكزها x_1'، x_2'، ... ، x_r' على الترتيب.

مثال (44): حل المثال رقم (43) السابق باستخدام العلاقة (11).

الحل: نُكوِّن جدولاً مؤلفاً من عمودين فقط كما يأتي، ثُمَّ نجد التباين من العلاقة (11):

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]$$

$$= \frac{1}{7}\left(9232 - \frac{65536}{8}\right) = 148.57 \text{ سنة}$$

x_i	x_i^2
15	225
21	441
22	484
30	900
35	1225
39	1521
44	1936
50	2500
$\sum_{i=1}^{n} x_i = 256$	$\sum_{i=1}^{n} x_i^2 = 9232$

ومنه نجد الانحراف المعياري:

$$S = \sqrt{148.57} = 12.19 \qquad \text{سنة}$$

مثال (45): احسب التباين والانحراف المعياري لبيانات المثال (16).

الحل: لدينا الوسط الحسابي $\overline{x} = 195$ ولتسهل الحسابات نُكوِّن جدول الحل كما يأتي:

حدود الفئات	التكرار f	مراكز الفئات x'	$x' - \overline{x}$	$(x' - \overline{x})^2$	$f(x' - \overline{x})^2$
$[60 - 100[$	3	80	-115	13225	39675
$[100 - 140[$	7	120	-75	5625	39375
$[140 - 180[$	9	160	-35	1225	11025
$[180 - 220[$	15	200	5	25	375
$[220 - 260[$	8	240	45	2025	16200
$[260 - 300[$	6	280	85	7225	43350
$[300 - 340[$	2	320	125	15625	31250
المجموع	50	----	----	----	181250

وباستخدام العلاقة (12) نجد أن:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{r} f_i (x_i' - \overline{x})^2$$

$$= \frac{181250}{49} = 3698.98 \Longrightarrow S = \sqrt{3698.98} = 60.82$$

3-محاسن الانحراف المعياري ومساوئه:

آ-محاسن الانحراف المعياري:

1- يأخذ جميع القيم بالحسبان.

2- يعتبر من أدق مقاييس التشتت.

3- يدخل في معظم التحاليل الإحصائية لسهولة التعامل معه رياضياً.

ب-مساوئ الانحراف المعياري:

1- يتأثر بالقيم الشاذة.

2- يصعب حسابه في البيانات الوصفية وفي حالة الجداول المفتوحة.

1 – 9 -مقاييس التشتت النسبية (Measures of relative dispersion):

لقد درسنا في الفقرات السابقة بعض أهم مقاييس التشتت والتي لها وحدات حسب طبيعة الظاهرة المدروسة، وتصلح للمقارنة بين الظواهر التي لها الوحدات نفسها، مثل المقارنة بين الانحراف المعياري لأطوال مجموعة من الطلاب مع الانحراف المعياري لأطوال مجموعة أخرى من الموظفين، وهكذا....

وإذا أردنا المقارنة بين ظاهرتين لكل منها وحدات تختلف عن الأخرى مثل مقارنة الانحراف المعياري لأوزان مجموعة من الأشخاص مع الانحراف المعياري لأطوالهم، فإن المقاييس السابقة للتشتت لا تصلح للمقارنة، وذلك لاختلاف الوحدات، لأن الانحراف المعياري للأوزان يقاس بالكيلوغرام والأطوال تقاس بالسنتمتر، لذلك دعت الحاجة إلى إيجاد مقياس نسبي لا يعتمد على وحدات القياس نذكر منها:

1 – معامل الاختلاف (Coefficient of Variation):

إن مقاييس التشتت التي عرفناها سابقاً تعتمد جميعها على وحدات القياس المستخدمة في البيانات ولكي نحصل على مقياس لا يعتمد على وحدات القياس المستخدمة نعرف مقياس جديد يسمى معامل الاختلاف أو معامل التغير ويعرَّف بالعلاقة التالية:

$$\text{معامل الاختلاف} = \mathrm{C.V.} = \frac{\text{الانحراف المعياري}}{\text{الوسط الحسابي}} = \frac{S}{\overline{X}} \qquad (13)$$

أم في حالة كون جداول التوزيعات التكرارية مفتوحة لمجموعة من البيانات بأنه للتغلب على ذلك يعرَّف معامل الاختلاف باستخدام الربيعات بالعلاقة التالية:

$$\text{معامل الاختلاف} = \mathrm{C.V.} = \frac{r_3 - r_1}{r_3 + r_1} \qquad (14)$$

حيث r_1 هو الربيع الأول أو الربيع الأدنى r_3 هو الربيع الثالث أو الربيع الأعلى.

- يستخدم معامل الاختلاف قياس درجة التفاوت بين المفردات أو البيانات ولا يعتمد على وحدات القياس المستعملة، أي يستخدم في المقارنة بين التغير أو الاختلاف في عدة مجموعات أو توزيعات تكرارية بغض النظر عن وحدات القياس المستعملة مختلفة أو نفسها.

مثال (46): احسب معامل الاختلاف لبيانات المجموعتين التاليتين ثُمَّ بيّن أيهما أكثر تغيراً:

بيانات المجموعة الأولى: $75, 80, 82, 87, 96$.

بيانات المجموعة الثانية: $35, 23, 27, 25, 21, 45, 34$

الحل:

- نحسب معامل الاختلاف للمجموعة الأولى بالشكل التالي:

$$\overline{x}_1 = \frac{1}{n}\sum_{i=1}^{n} x_{1i} = \frac{75+80+82+87+96}{5} = 84$$

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$= \frac{(75-84)^2+(80-84)^2+(82-84)^2+(87-84)^2+(96-84)^2}{4}$$

$$= \frac{81+16+4+9+144}{4} = 63.5 \quad \Rightarrow S_1 = 7.97$$

وباستخدام العلاقة (13) نجد أن:

$$\text{معامل الاختلاف للتوزيع الأول} = \mathrm{C.V.} = \frac{\text{الانحراف المعياري}}{\text{الوسط الحسابي}} = \frac{7.97}{84} = 0.09$$

- ثُمَّ نحسب معامل الاختلاف للمجموعة الثانية بالشكل التالي:

$$\overline{x}_2 = \frac{1}{n}\sum_{i=1}^{n} x_{2i} = \frac{35+23+27+25+21+45+34}{7} = 30$$

$$S_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$= \frac{(35-30)^2+(23-30)^2+(27-30)^2+(25-30)^2+(21-30)^2+(45-30)^2+(34-30)^2}{6}$$

$$= \frac{25+49+9+25+81+225+16}{6} = 71.67 \quad \Longrightarrow S_2 = 8.47$$

وباستخدام العلاقة (13) نجد أن:

$$\text{معامل الاختلاف للتوزيع الثاني} = \text{C. V.} = \frac{\text{الانحراف المعياري}}{\text{الوسط الحسابي}} = \frac{8.47}{30} = 0.28$$

نلاحظ أن معامل الاختلاف في بيانات المجموعة الأولى أصغر من معامل الاختلاف في بيانات المجموعة الثانية. أي أن مقدار التفاوت بين بيانات المجموعة الأولى أصغر من مقدار التفاوت بين بيانات المجموعة الثانية.

مثال (47): احسب معامل الاختلاف باستخدام الربيعات لبيانات التوزيعين التكراريين التاليين ثُمَّ بيّن أيهما أكثر تغيراً:

بيانات التوزيع التكراري الأول: $r_1 = 15.65\ , r_3 = 86.35$.

بيانات التوزيع التكراري الثاني: $r_1 = 11.02\ , r_3 = 35.98$.

الحل:

بحساب معامل الاختلاف لكل من التوزيعين فنجد معامل الاختلاف للتوزيع الأول:

$$\text{معامل الاختلاف للتوزيع الأول} = \text{C. V.} = \frac{r_3-r_1}{r_3+r_1} = \frac{86.35-15.65}{86.35+15.65} = \frac{70.7}{102} = 0.69$$

ويكون معامل الاختلاف للتوزيع الثاني:

$$\text{معامل الاختلاف للتوزيع الثاني} = \text{C. V.} = \frac{r_3-r_1}{r_3+r_1} = \frac{35.98-11.02}{35.98+11.02} = \frac{24.96}{47} = 0.53$$

نلاحظ أن معامل الاختلاف في التوزيع التكراري الأول أكبر من معامل الاختلاف في التوزيع التكراري الثاني. أي أن مقدار التفاوت بين بيانات التوزيع التكراري الأول أكبر من مقدار التفاوت بين بيانات التوزيع التكراري الثاني.

ملاحظة:

عند المقارنة بين قيمتي معامل الاختلاف لبيانات توزيع تكراري باستخدام التعريفين السابقين، فإننا نحصل على جوابين مختلفين، وذلك لاختلاف الأساس الرياضي في كل من التعريفين السابقين، وفي هذه الحالة نفضل استخدام التعريف الأول إذا كانت جداول التوزيعات التكرارية غير مفتوحة وذلك لدقته.

2 – العزوم (The Moments):

يعرف العزم الرائي أو العزم من المرتبة r لمجموعة من البيانات x_1،x_2، ...،x_n حول وسطها الحسابي $\overline{x}$ بالعلاقة التالية:

$$m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^r$$

ويعرف العزم الرائي أو العزم من المرتبة r لمجموعة من البيانات x_1،x_2، ...،x_n حول نقطة الأصـــل أو حول المبدأ بالعلاقة التالية:

$$m_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$$

يعرف العزم الرائي أو العزم من المرتبة r لمجموعة من البيانات x_1،x_2، ...،x_n موزعة على فئات مراكزها $x'_1, x'_2, \dots, x'_k$ وتكراراتها $f_1, f_2, \dots, f_k$ حول وسطه الحسابي $\overline{x}$ بالعلاقة التالية:

$$m_r = \frac{1}{n}\sum_{i=1}^{k} f_i (x'_i - \overline{x})^r$$

ويعرف العزم الرائي أو العزم من المرتبة r لمجموعة من البيانات x_1،x_2، ...،x_n موزعة على فئات مراكزها $x'_1, x'_2, \dots, x'_k$ وتكراراتها $f_1, f_2, \dots, f_k$ حول نقطة الأصل أو حول المبدأ بالعلاقة التالية:

$$m_r = \frac{1}{n}\sum_{i=1}^{k} f_i \, {x'_i}^r$$

3 – مقاييس الالتواء (Measures of Skewness):

لقد سبق أن ذكرنا أن أشكال المنحنيات

1 – مقياس الالتواء لبيرسون (*Pearson*):

يعرف معامل بيرسون للالتواء بالعلاقة التالية:

$$\mathrm{Sk} = \frac{(\overline{x} - Mod)}{s}$$

أو بالعلاقة:

$$\mathrm{Sk} = \frac{3\,(\overline{x} - Median)}{s}$$

نلاحظ أن معامل الالتواء لبيرسون يعطي نتائج مقبولة عندما يكون الالتواء بسيطاً، ويعطي نتائج غير مقبولة عندما يكون الالتواء شديداً أو في حالة التوزيعات التكرارية المفتوحة.

2 – مقياس الالتواء لباولي (*Bowley*):

يعرف معامل الالتواء لباولي بالعلاقة التالية:

$$\mathrm{Sk} = \frac{(r_3 - Median) - (Median - r_1)}{(r_3 - Median) + (Median - r_1)}$$

يستخدم معامل التواء باولي في حالة التوزيعات التكرارية المفتوحة والمغلقة.

3 – مقياس الالتواء بطريقة العزوم:

يعرف معامل الالتواء بطريقة العزوم بالعلاقة التالية:

$$\mathrm{Sk} = \frac{m_r}{S^3} = \frac{\text{العزم الثالث المركزي}}{\left(\text{الانحراف المعياري}\right)^3}$$

مثال (45):

4 – معامل التفلطح (Coefficient of Kurtosis):

الفصل الثاني

الارتباط والانحدار

Correlation and Regression

2-1-الانحدار (Regression):

الانحدار هو طريقة إحصائية حديثة كثير التطبيقات في الحواسيب وهو يحوي على قدرٍ كبير من الخلفية النظرية وتحتاج هذه الطريقة إلى قاعدة جيدة في معرفة اســـتخدام أحد البرامج الإحصــــائية المعروفة مثل SAS، S، Mathematica، SPSS، وترجع بداية هذا البحث إلى العالم الإنكليزي Sir Galton عام 1900 حيث إن ملكة إنكلترة قد طلبت منه أن يدرس لها تطور الطبقة الارســـتقراطية في إنكلترة وذلك من خلال دراســـة العلاقة الموجودة بين طول الوالد وطول الابن الأكبر في عائلة واحدة. وكان اعتقاده أنه كلما كان الوالد طويلاً كلما كان الابن أشـد طولاً ولكن فقد وجد خلاف ذلك تماماً حيث إنه يوجد انحدار نحو المتوسط.

تعريـــف: الانحدار هو مجموعة من العمليات الرياضية المستخدمة من أجل كشف اللثام عن علاقـــة بين المتغيرات ولدينا نوعين من المتغيرات، ندعو المتغير الأول y المتغير غير المستقل كما وندعو المتغير الثاني x المتغير المستقل.

لقد تعرفنا في الفقرات السـابقة إلى كيفية حسـاب مقاييس النزعة المركزية ومقاييس التشـتت لمجموعة واحدة من البيانات أو أكثر بغية المقارنة فيما بينها، علماً أن الصـــفة المشـــتركة التي كانت تمثل هذه المجموعات كانت تعتمد على متغير واحد فقط وهو المتغير المدروس.

ولكن في الحياة العملية قد تكون مفردات العينة عبارة عن أزواج من القيم لخاصـــتين مختلفتين، كما قد يكون المطلوب في مثل هذه الحالة دراســـة العلاقة بينهما ومعرفة ما إذا كان تغير إحدى الظاهرتين مرتبطاً بتغير الأخرى ومن ثُمَّ تحديد نوع العلاقة التي تربطهما، ومقياس قوة هذه العلاقة واتجاهها كأن تكون طردية أو عكسية أو غير ذلك.

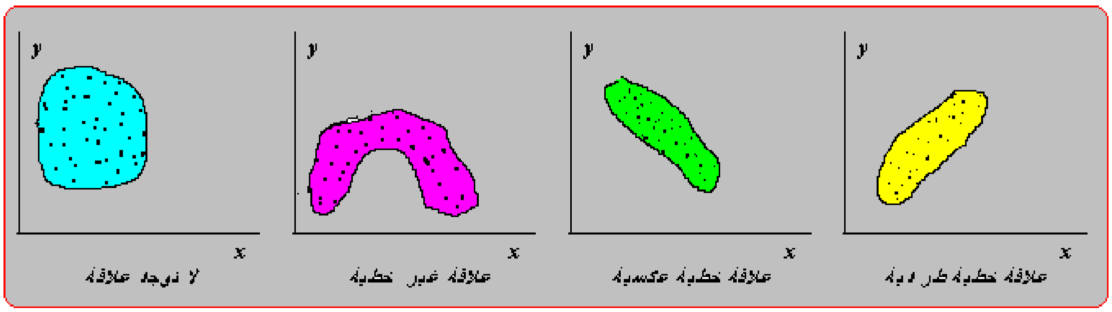
من هذه العلاقات على سبيل المثال دراسة العلاقة بين مستوى الذكاء والعامل الوراثي أو دراسـة العلاقة بين الطول والوزن لمجموعة من الأشخاص وهكذا ......

في هذه الفقرة سوف نقوم بدراسة طرائق قياس مقدار العلاقة بين متغيرين مفروضين، مثل قياس قوة الارتباط بينهما وإيجاد مقاييس عددية لقياس قوة الارتباط تربط بين المتغيرين بعضـــــهما ببعض، تمكننا من التنبؤ بقيمة أحد المتغيرين لقيمة محدودة للمتغير الآخر. وقبل البدء بهذه الدراسة يجب أن نتعرف على أشكال الانتشار والتي تصف العلاقة بين المتغيرات وهي:

1 – إذا كانت نقاط الانتشار منتشرة بشكل عشوائي ومبعثرة، بحيث أن نصيب كل وحدة مساحة بشكل وسطي لا يختلف من مكان لأخر في المستوى أي دون أي نظام فإن مثل هذا الشكل يدل على عدم وجود أية علاقة بين المتغيرين y ، x والشكل اليساري من الشكل التالي يبين ذلك.

2 – إذا كانت نقاط الانتشار في شكل حزمة ذات اتجاه ثابت كأن تكون من أعلى اليسار إلى أدنى اليمين، وفي هذه الحالة تكون العلاقة بين المتغيرين علاقة عكسية أو سالبة أي أن المتغير y ينقص بزيادة المتغير x والشكل الثالث من الشكل السابق يبين ذلك.

3 – أما إذا كانت نقاط الانتشار منتشرة في شكل حزمة ذات اتجاه ثابت كأن تكون من أعلى اليمين إلى أدنى اليسار، وفي هذه الحالة تكون العلاقة بين المتغيرين علاقة طردية كما في الشكل الرابع من الشكل السابق، أي أن المتغير y يزداد بزيادة المتغير x ونسمي هذا الارتباط بالارتباط الموجب.



يتضح مما تقدم أن إجراء تحليل الارتباط بين سلوك متغيرين أو ظاهرتين يعتمد في الدرجة الأولى على تصور العلاقة القائمة بينهما ومن أجل الحصول على فكرة أولية عن طبيعة العلاقة القائمة بين متغيرين ، نعمد كما رأينا إلى رسم محورين متعامدين ونحدد في الشكل نقاط الانتشار حيث إن الإحداثيات الأفقية تساوي قيم x وهو المتغير المستقل والإحداثيات العمودية التي تناظرها وتساوي قيم y وهو المتغير التابع ، أي أننا نقوم برسم نقاط في مستوى الإحداثيات الديكارتية وإحداثيات هذه النقاط هي (x_1, y_1)،$(x_2, y_2), \dots, (x_n, y_n)$. وكما رأينا فإن الشكل البياني الذي نحصل عليه يدعى شكل الانتشار، سوف نوضح ذلك في المثال الآتي:

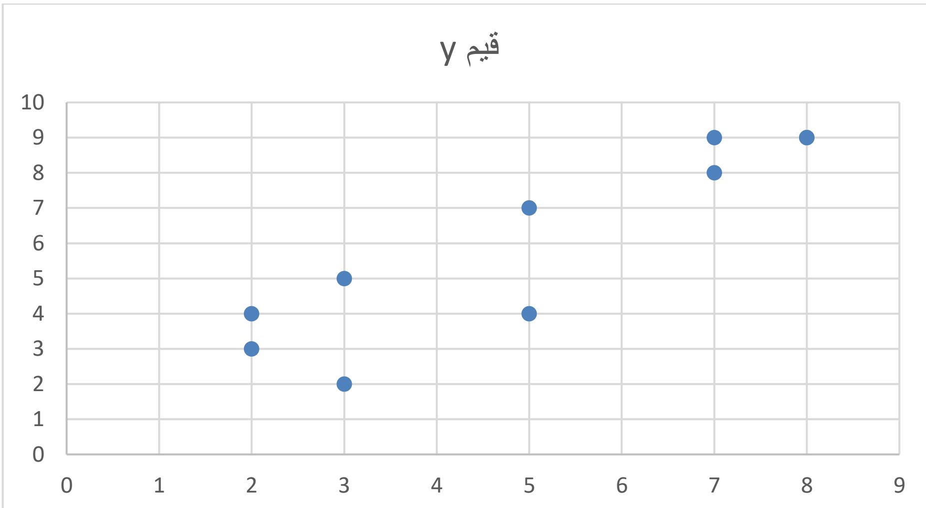
مثـــال (1): لنأخذ بعين الاعتبار سلوك كل من المتغيرين x و y الذين يرمزان إلى مدة الخدمة الفعلية بالسنوات في المطارات المدنية، وعدد ساعات الطيران بمئات الساعات وذلك لعشرة طيارين والمبينة في الجدول الآتي، ولندرس العلاقة بين الظاهرتين.

مدة الخدمة الفعلية x	8	5	2	7	7	8	2	5	3	3
عدد ساعات الطيران y	9	7	4	9	8	9	3	4	5	2

إذا مثلنا هذه البيانات في مستوي إحداثيات ديكارتية، حيث يدل المحور الأفقي x إلى مدة الخدمة الفعلية بالسنوات ويدل المحور العمودي y إلى عدد ساعات الطيران، ومثلنا كل زوج من القيم بنقطة فاصلتها وترتيبها على الترتيب مدة الخدمة وعدد ساعات الطيران لطيار ما، لحصلنا على مجموعة من النقاط تدعى شكل الانتشار.

من الممكن إيجاد المستقيم الذي يلاءم هذه النقاط على أحسن وجه ووضع معادلتين ويكون بمثابة المحور لهذه السحابة، ويعطينا فكرة عن العلاقة بين الظاهرتين، كما وأنه يمكننا هذا المستقيم عند إيجاده، ورسمه من تقدير قيمة معينة لظاهرة ما إذا عُلمت القيمة الموافقة للظاهرة الأخرى، أي تقدير عدد ساعات الطيران لطيار عُرفت مدة خدمته.

يدعى هذا المستقيم بمستقيم الانحدار كما وأن هناك مقياساً لشدة العلاقة بين الظاهرتين يدعى معامل الارتباط. سوف نكتفي بدراسة معامل الارتباط الخطي وكذلك معادلة الانحدار الخطي بأشكال الانتشار وذلك لمراعاة مستوى وطبيعة تخصصات الدارسين لهذا الكتاب. وسوف نقوم بدراسة معامل الارتباط الخطي لبيرسون بين (x, y). كما سنحاول تبسيط عرضنا للموضوع كلما أمكن، وذلك باستخدام الأمثلة.



1– معامل الارتباط الخطي لبيرسون:(Pearson's coefficient of correlation)

يستخدم معامل الارتباط الخطي لبيرسون لقياس قوة الارتباط بين متغيرين x , y عندما تكون أزواج القراءات كمية أي رقمية، وسوف نقدم معامل الارتباط الخطي لبيرسون في حالة البيانات غير المبوبة مباشرة، فإذا كان لدينا أزواج القيم للمتغيرين x, y من المجتمع محل الدراسة بالشكل (x_1, y_1)،(x_2, y_2)، … ،(x_n, y_n) فإننا نعرّف معامل الارتباط الخطي لبيرسون والذي نرمز له بالرمز R بالعلاقة الآتية:

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

وهناك علاقات أخرى عرفنا أسهلها.

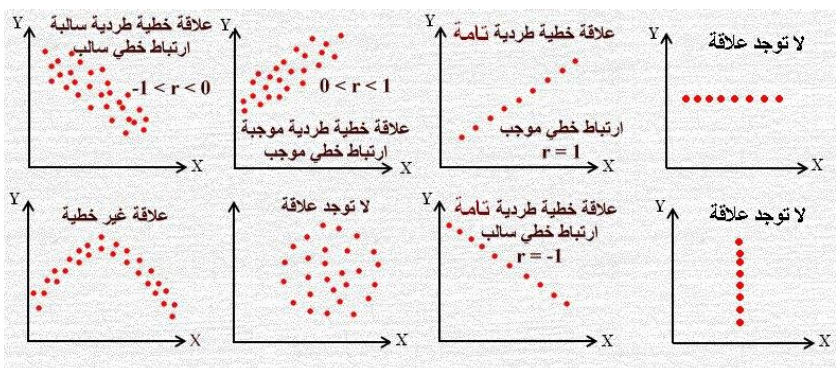
ملاحظة: تجدر الإشارة إلى أنه لا توجد حدود فاصلة تبين قوة وضعف الارتباط، ولكن يمكن وضع حدود تقريبية لقيم r مبنية على الخبرة السابقة، وسوف نذكر ذلك للقيم الموجبة وبالمثل يمكن تطبيقها عندما تكون r سالبة وذلك بتغير إشارة الحدود في الجدول الآتي:

قيم معامل الارتباط r	قوة الارتباط
صفر إلى 0.3	لا يوجد ارتباط
0.3 إلى 0.5	ارتباط ضعيف
0.5 إلى 0.7	ارتباط متوسط
0.7 إلى 0.9	ارتباط قوي
0.9 إلى 1	ارتباط قوي جداً

وبذلك يكون الارتباط في مثال السابق قوياً جداً.

كما وتجدر الإشارة إلى أن من أهم خصائصه أنه لا يعتمد على القيم نفسها وإنما يعتمد على مقدار تباعد هذه القيم عن بعضها، ولذلك إذا جمعنا أو طرحنا مقداراً ثابتاً من كل قراءات الظاهرتين x أو y فإن قيمة معامل الارتباط لا تتغير.

كما ويتمتع معامل الارتباط بهذه الخاصة بالنسبة للضرب والقسمة إلا إنه في حالة ضرب أو قسمة مقدار ثابت في كل من x،y فإن قيمة معامل الارتباط لا تتغير بمثل هذه العمليات البسيطة.



ملاحظات: نلاحظ أن قيمة معامل الارتباط هي عبارة عن عدد حقيقي محصور بين العددين الصحيحين $+1$, -1 .

ونقول عن الارتباط أنه طردي إذا كانت قيمة معامله r موجبة وتزداد قوة الارتباط كلما اقتربت قيمة معامل الارتباط r من الواحد الصحيح، وتضعف قيمته كلما اقتربت قيمة معامل الارتباط r من الصفر.

وأنه عكسي إذا كانت قيمة معامله r سالبة. كما وتزداد قوة الارتباط العكسي كلما اقتربت قيمة معامل الارتباط r من العدد -1 . وتضعف قيمته كلما اقتربت قيمة معامل الارتباط r من الصفر.

مثال (2): أوجد معامل الارتباط الخطي لبيرسون بين مدة الخدمة بالسنوات x وعدد ساعات الطيران بمئات الساعات y لمجموعة مكونة من عشرة طيارين مدنيين مبينة كما في الجدول التالي:

مدة الخدمة بالسنوات x	3	3	5	2	8	7	7	2	5	8
عدد ساعات الطيران y	2	5	4	3	9	8	9	4	7	9

الحل: لتسهيل الحسابات ننشئ الجدول التالي:

مدة الخدمة بالسنوات x	عدد ساعات الطيران y	xy	x^2	y^2
3	2	6	9	4
3	5	15	9	25
5	4	20	25	16
2	3	6	4	9
8	9	72	64	81
7	8	56	49	64
7	9	63	49	81
2	4	8	4	16
5	7	35	25	49
8	9	72	64	81
50	60	353	302	426

بالتعويض في قانون الارتباط الخطي نجد:

$$r=\frac{n\sum xy-\sum x\sum y}{\sqrt{[n\sum x^2-(\sum x)^2][n\sum y^2-(\sum y)^2]}}=\frac{10(353)-(50)(60)}{\sqrt{[10(302)-(50)^2][10(426)-(60)^2]}}$$

$$=\frac{3530-3000}{\sqrt{(3020-2500)(4260-3600)}}=\frac{530}{\sqrt{(520)(660)}}=\frac{530}{\sqrt{343200}}=\frac{530}{585.8}\cong 0.90$$

2− معامل ارتباط سبيرمان Spearman's Coefficient of Correlation:

نلاحظ من الفقرة الســـابقة أن الارتباط الخطي لبيرســـون يبين مدى قوة الارتباط بين المتغيرين (x,y) في حالة البيانات الكمية فقط، ولكن في كثير من الدراسات نصادف بيانات وصفيـــة يكون المطلوب فيها إيجاد قوة الارتباط بين المتغيرين الوصفيين.

لذلك دعت الحاجة إلى إيجاد مقياس يستخدم في حالة البيانات الوصـفية خاصـة إذا أمكن وضـعها في صـورة ترتيبية، مثل تقديرات الجنود، أو الرتب العسكرية لضباط الجيش، أو المراتب والدرجات لموظفين حسب السلم الوظيفي.

ويمكن ملاحظة أن استخدام معامل ارتباط الرتب لسبيرمان يفيد في مثل حالة هذه البيانات الوصفية السالفة الذكر أو البيانات الكمية كذلك مع مراعاة أن تكون عدد أزواج القيم أقل من 30 حتى يمكن أن يعطي معامل ارتباط الرتب في أغلب الأحيان قوة الارتباط بصورة أكثر دقة من معامل ارتباط بيرسون.

ويُعرّف معامل ارتباط الرتب لسبيرمان r بالعلاقة الآتية:

$$r = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

حيث إن n عدد المشاهدات، d فرق الرتبة بين المتغيرين.

ولتوضيح طريقة إيجاد رتب مجموعة من الأرقام نتصور أننا رتبنا الأرقام تصاعدياً أو تنازلياً فيكون الرقم الأول رتبته 1 والرقم الثاني رتبته 2 وهكذا ....

وإذا تساوى رقمان فإننا نأخذ متوسط مجموع الرتبتين لهما، ونوضح ذلك باستخدام الترتيب التصاعدي في الأمثلة الثلاثة الآتية، حيث نوضح في المثال (3) كيفية تحديد رتب القراءات، ومن ثُمَّ نطبق ذلك لإيجاد معامل ارتباط الراتب في المثالين الآتيين (4) ، (5) .

مثال (3): أوجد رتب الأعداد الآتية:

x	3	5	8	9	2	3	6

إذا تصورنا ترتيب البيانات تصاعدياً فإن الرقم 2 يحتل المرتبة الأولى (1)، والرقمين 3,3 يحتلان المرتبتين الثانية والثالثة (2 , 3) ، وتكون رتبة كل منهما هي متوسط الرتبتين (2 , 3) أي ((2 + 3) ÷ 2 = 2.5) والرقم 5 يحتل المرتبة الرابعة (4) وهكذا باقي الأرقام.... ونوضح ذلك بالجدول الآتي لقيم س ورتبها.

x	3	5	8	9	2	3	6
رتبة x	2.5	4	6	7	1	2.5	5

ولإيجاد معامل الارتباط للرتب نوضح طريقة حسابه بالمثالين الآتيين:

مثال (4): أوجد معامل ارتباط الرتب لسبيرمان لمدة الخدمة وعدد ساعات الطيران لعينة مكونة من عشرة طيارين حسب البيانات المعطاة في المثال (2) السابق.

الحل:

يمكن تلخيص الحسابات كما في الجدول الآتي:

مدة الخدمة بالسنوات x	عدد ساعات الطيران y	رتبة x	رتبة y	d	d^2
3	2	3.5	1	2.5	6.25
3	5	3.5	5	-1.5	2.25
5	4	5.5	3.5	2	4
2	3	1.5	2	-0.5	0.25
8	9	9.5	9	0.5	0.25
7	8	7.5	7	0.5	0.25
7	9	7.5	9	-1.5	2.25
2	4	1.5	3.5	-2	4
5	7	5.5	6	-0.5	0.25
8	9	9.5	9	0.5	0.25
——	——	—	**المجموع**	0	20

حيث d هي: $d =$ رتبة $x -$ رتبة y .

ومن ذلك يمكن حساب معامل الارتباط الخطي لسبيرمان كما يأتي:

$$r = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6\,(20)}{10(100-1)}$$

$$= 1 - \frac{120}{990} = 1 - 0.136 = 0.864$$

وهو ارتباط طردي وقوي.

ملاحظة: إذا حسبنا معامل الارتباط الخطي اعتماداً على طريقة بيرسون فليس من الضروري أن نحصل على النتيجة نفسها عند حسابه اعتماداً على طريقة سيبرمان.

مثـــــال (5): في دراسة اجتماعية عن الوضع المالي لكل من أسرتي الزوج والزوجة لعينة مكونة من خمسة جنود، وذلك لمعرفة تأثير الحالة المادية في الزواج بين أسر الجنود، حيث كانت المعلومات كما في الجدول الآتي:

الحالة المادية لأسرة الجندي x	جيدة	منخفضة	متوسطة	ممتازة	متوسطة
الحالة المادية لأسرة زوجته y	ممتازة	جيدة	جيدة	ممتازة	متوسطة

أي أن المطلوب هو حساب معامل ارتباط الرتب.

نلخص الحل في الجدول الآتي:

d^2	d	رتبة y	رتبة x	أسرة زوجته y	أسرة الجندي x
0.25	-0.5	4.5	4	ممتازة	جيدة
2.25	-1.5	2.5	1	جيدة	منخفضة
0	0	2.5	2.5	جيدة	متوسطة
0.25	0.5	4.5	5	ممتازة	ممتازة
2.25	1.5	1	2.5	متوسطة	متوسطة
5	0	المجموع	—	—	—

$$r = 1 - \frac{6\,\Sigma\, d^2}{n\,(\,n^2 - 1\,)} = 1 - \frac{6\,(\,5\,)}{5\,(\,25 - 1\,)} = 1 - 0.25 = 0.75$$

أي أنه يوجد ارتباط طردي قوي.

ثانياً: معامل الاقتران (Coefficient of contingency):

لقد سبق أن عرفنا معامل الارتباط لسبيرمان (معامل ارتباط الرتب) للبيانات الوصفية التي يمكن ترتيبها ، ولكن نصادف كثيراً من الدراسات التطبيقية في مختلف أوجه الحياة العملية كعلم النفس علم الاجتماع، العلوم العسكرية، والعلوم الزراعية... إلى آخره، بيانات وصفية ليس في طبيعتها صفة الترتيب أي لا يمكن وضع رتب لها، أو لا معنى للرتب فيها، نذكر مثلاً التدخين له صفتان مدخن ، أو غير مدخن ، التعليم له صفتان متعلم أو غير متعلم، طبيعة العمل لشخص يعمل ، أو لا يعمل ، طبيعة الجسم يكون سليم ، أو غير سليم ، فصائل الدم تكون مثلاً A , A^+ , B , ...

ولتسهيل مفهوم الاقتران بين صفات متعددة نفرض أن لكل من المتغيرين y ، x صفتين أولى وثانية ويمكن التعبير عن مثل هذه البيانات كما في الجدول الآتي:

المتغير y / المتغير x	الصفة الأولى y_1	الصفة الثانية y_2
الصفة الأولى x_1	***A***	***B***
الصفة الثانية x_2	***C***	***D***

حيث يدل الرمز A عن التكرارات المشتركة في الصفة الأولى x_1 والصفة الأولى y_1 ، وهكذا بالنسبة لبقية الرموز B ، C ، D .

وقد اقترح بيل بأن يُعرّف معامل الاقتران $(C.C.)$ في هذه الحالة بالعلاقة الآتية:

$$C.C.=\frac{AD-BC}{AD+BC}$$

ونوضح طريقة حساب $C.C.$ بالمثال الآتي.

مثال (6): أوجد معامل الاقتران $C.C.$ بين التدخين والتعليم لمجموعة من الأشخاص حيث كانت البيانات كما يأتي:

التدخين y / التعليم x	مدخن	غير مدخن
متعلم	15	10
غير متعلم	9	16

الحل: ومن ذلك يمكن حساب معامل الاقتران كما يلي:

$$C.C.=\frac{AD-BC}{AD+BC}=\frac{15\times16-10\times9}{15\times16+10\times9}=\frac{150}{330}=0.45$$

أي يوجد ارتباط متوسط بين التدخين والتعليم لمجموعة الأشخاص.

ثالثاً: معامل التوافق:

استخدم كرامر مقياساً آخراً للارتباط في الحالة التي يكون فيها كلا المتغيرين وصفيين أو أحدهما وصفي والآخر كمي ولكل منهما أكثر من حالة، يدعى معامل التوافق.

أي عندما تتكون الظواهر من عدة صفات لكل متغير وليس صفتين فقط كما في الحالة الأولى حالة معامل الاقتران، أي أننا بصدد تعميم الحالة السابقة.

فإذا فرضنا أن للمتغير الأول x الصفات التالية $x_1, x_2, \dots, x_r$ وأن للمتغير الثاني y الصفات التالية $y_1, y_2, \dots, y_s$ ورمزنا بـ f_{ij} لتكرارات العينة التي لها الصفة i للمتغير الأول x ولها الصفة j للمتغير الثاني y ورتبنا هذه الكميات بالجدول التالي:

الصفة y / الصفة x	الصفة الأولى y_1	الصفة الثانية y_2	…	الصفة y_s	المجموع
الصفة الأولى x_1	f_{11}	f_{12}	…	f_{1s}	$f_{1\bullet}$
الصفة الثانية x_2	f_{21}	f_{22}	…	f_{2s}	$f_{2\bullet}$
…	…	…	…	…	…
الصفة x_r	f_{r1}	f_{r2}	…	f_{rs}	$f_{r\bullet}$
المجموع	$f_{\bullet 1}$	$f_{\bullet 2}$	…	$f_{\bullet s}$	$n = f_{\bullet\bullet}$

$f_{i\bullet}$: مجموع التكرارات في العينة التي لها الصفة i للمتغير الأول x .

$f_{\bullet j}$: مجموع التكرارات في العينة التي لها الصفة j للمتغير الثاني y .

فإن معامل التوافق يعطى بالعلاقة التالية:

$$r_a = \sqrt{\frac{B-1}{B}}$$

علماً أن
$$B = \frac{f_{11}^2}{f_{1\bullet} f_{\bullet 1}} + \frac{f_{12}^2}{f_{1\bullet} f_{\bullet 2}} + \cdots + \frac{f_{rs}^2}{f_{r\bullet} f_{\bullet s}}$$

أي أننا نربع تكرار الخلية الأولى، ونقسمه على حاصل ضرب مجموع التكرارات للصف الذي به الخلية الأولى، بمجموع التكرارات للعمود الذي به الخلية الأولى، وهكذا نحسب بقية حدود B .

مثال (7): أوجد معامل التوافق بين لون العيون x ولون الشعر y لعينة مؤلفة من 45 شخصاً باستخدام البيانات الواردة في الجدول التالي:

لون الشعر y / لون العيون x	أشقر	بني	أسود	المجموع
أزرق	6	5	4	15
عسلي	3	6	6	15
أسود	2	7	6	15
المجموع	11	18	16	$n = 45$

الحل: لإيجاد معامل التوافق نوجد أولاً B من العلاقة السابقة بالشكل التالي:

$$B=\frac{f_{11}^2}{f_{1\bullet}f_{\bullet 1}}+\frac{f_{12}^2}{f_{1\bullet}f_{\bullet 2}}+\cdots+\frac{f_{rs}^2}{f_{r\bullet}f_{\bullet s}}$$

$$=\frac{(6)^2}{11\times 15}+\frac{(5)^2}{18\times 15}+\frac{(4)^2}{16\times 15}+\frac{(3)^2}{11\times 15}+\frac{(6)^2}{18\times 15}$$

$$+\frac{(6)^2}{16\times 15}+\frac{(2)^2}{11\times 15}+\frac{(7)^2}{18\times 15}+\frac{(6)^2}{16\times 15}$$

$$=0.22+0.09+0.07+0.05+0.13+0.15+0.02+0.18+0.15=1.07$$

بالتعويض في علاقة التوافق نجد:

$$r_a=\sqrt{\frac{B-1}{B}}=\sqrt{\frac{1.07-1}{1.07}}=\sqrt{0.065}=0.25$$

أي يوجد ارتباط ضعيف بين لون الشعر ولون العينين.

مثال (8): يمثل الجدول الآتي أوضاع 280 مدخناً ومصنفة حسب درجة إدمانهم من جهة وإصابتهم بالضغط الشرياني من جهة أخرى:

غير مدخن	مدخن وسط	مدخن بكثرة	التدخين x / الضغط الشرياني y
21	36	30	مصاب بالضغط الشرياني
148	26	19	غير مصاب بالضغط الشرياني

والمطلوب:

1 – احسب معامل التوافق بين التدخين وارتفاع الضغط الشرياني.

2 – هل تعتقد أن هناك علاقة بين التدخين وارتفاع الضغط الشرياني.

الحل:

المجموع	غير مدخن	مدخن وسط	مدخن بكثرة	التدخين x / الضغط الشرياني y
87	21	36	30	مصاب بالضغط الشرياني
193	148	26	19	غير مصاب بالضغط الشرياني
280	169	62	49	المجموع

لإيجاد معامل التوافق نوجد أولاً B من العلاقة السابقة بالشكل التالي:

$$B=\frac{f_{11}^2}{f_{1\bullet}f_{\bullet 1}}+\frac{f_{12}^2}{f_{1\bullet}f_{\bullet 2}}+\cdots+\frac{f_{rs}^2}{f_{r\bullet}f_{\bullet s}}$$
$$=\frac{(30)^2}{87\times 49}+\frac{(36)^2}{87\times 62}+\frac{(21)^2}{87\times 169}+\frac{(19)^2}{193\times 49}+\frac{(26)^2}{193\times 62}+\frac{(148)^2}{193\times 169}$$
$$=0.211+0.240+0.03+0.038+0.056+0.671=1.208$$

بالتعويض في علاقة التوافق نجد:

$$r_a=\sqrt{\frac{B-1}{B}}=\sqrt{\frac{1.208-1}{1.208}}=\sqrt{0.172}=0.415$$

2 – نستنتج أنه لا يوجد أي ارتباط بين الدخل ومراجعة الطبيب.

3 – خط الانحدار (Regression line):

سبق أن درسنا في هذا الفصل طرائق حساب قيمة معامل الارتباط الخطي بين متغيرين x،y كما تعرفنا على كيفية إيجاد قيمة معامل الاقتران لكارل بيرسون وغيره.

ومن الملاحظ أن جميع المقاييس الســـابقة تبين أو تعطي قوة الارتباط بين أي متغيرين فقط، ولكن إذا كان يود الدارس أو الباحث يرغب اســتقصـــاءً أو بحثاً لأحد المتغيرين عند معرفة قيمة محددة للمتغير الآخر فإنه لا يمكن اســـتخدام معامل الارتباط أو معاملي الاقتران والتوافق، ولكن لا بد للوصول إلى إيجاد علاقة جبرية محددة بين المتغيرين (x,y).

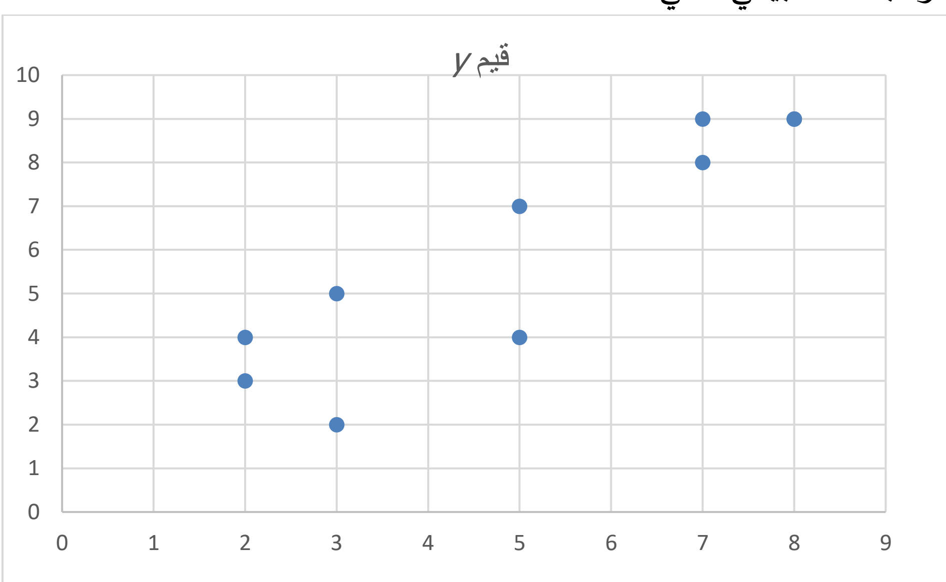
تسمى عادة العلاقة الرياضية التي تفرض التوقع أو التنبؤ بسلوك أحد المتغيرين عند معرفة الآخر بمعادلة خط الانحدار.

وتشــير معادلة خط الانحدار إلى انحدار أحد المتغيرين على المتغير الآخر، وسـوف ندرس فيما يلي معادلة خط انحدار المتغير y على المتغير x ويمكن تلخيص الصورتين الجبرية والبيانية لخطي الانحدار فيما يأتي:

انحدار y على x :

إن معادلة خط انحدار y على x يعطى بالعلاقة: $y=a+bx$

ويمكن توضيح المقصود بالشكل البياني الآتي:



حيث إن أحد المتغيرين يعتمد على الأخر سواء كان الاعتماد طردياً أو عكسياً.

ويمكن تحديد ورسم خط الانحدار بعدة طرائق: منها تمهيد خط مناسب بعد رسم شكل الانتشار للبيانات الخاصة بالمتغيرين (x,y) وهذه الطريقة تقريبية جداً، ولا تستخدم كثيراً، لأنها تختلف من شخص لآخر، ولهذا كان لابد من إيجاد طريقة لتوفيق خط الانحدار بحيث لا تعتمد على انطباع الأشخاص، ولكن تعتمد على البيانات الخاصة بالمتغيرين (x,y) فقط من المعادلة السابقة.

ومن ذلك يمكن تحديد الصيغة الرياضية لخط انحدار y على x بالضبط إذا عُلمت قيمتا الثابتين a ، b الذين يمكن حسابهما باتباع طريقة المربعات الصغرى فنجد مقدر لكل من الثابتين بالشكل التالي:

$$\hat{b}=\frac{n\sum x\,y-\sum x\sum y}{n\sum x^2-(\sum x)^2}$$

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

ونحصل على مقدر المربعات الصغرى لمعادلة الانحدار الخطي بالشكل التالي:

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

ويسمى الثابت b عادة بمعامل انحدار y على x .

حيث إن a تمثل الجزء الذي يقطعه خط انحدار y على x من محور y .

مثال (8): يبين الجدول التالي نسبة تركيز الكالسيوم في الدم x ونسبة تركيز الكالسيوم في البول y لستة أشخاص من المرضى مقدراً بالملغ لكل 100 سنتمتر مكعب:

المطلوب:

14	13	11	12	10	15	نسبة تركيز الكالسيوم في الدم x
13	12	10	8	9	14	نسبة تركيز الكالسيوم في البول y

1 – أوجد معامل الارتباط الخطي r بينهما.

2 – أوجد معادلة خط انحدار y على x .

3 – أوجد تقديراً لنسبة تركيز الكالسيوم في البول إذا كان نسبة تركيز الكالسيوم في الدم 15.

الحل:

1 – لتسهيل الحسابات ننشئ الجدول التالي:

نسبة تركيز الكالسيوم في الدم x	نسبة تركيز الكالسيوم في البول y	xy	x^2	y^2
15	14	210	225	196
10	9	90	100	81
12	8	96	144	64
11	10	110	121	100
13	12	156	169	144
14	13	182	196	169
75	66	844	955	754

بالتعويض في قانون الارتباط الخطي نجد:

$$r = \frac{n\sum xy-\sum x\sum y}{\sqrt{[n\sum x^2-(\sum x)^2][n\sum y^2-(\sum y)^2]}} = \frac{6(844)-(75)(66)}{\sqrt{[6(955)-(75)^2][6(754)-(66)^2]}} = 0.858$$

2 – إن معادلة انحدار تركيز الكالسيوم في البول y على نسبته في الدم x هي من الشكل:

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

حيث $\hat{a}$،$\hat{b}$ يعطيان بالشكل:

$$\hat{b} = \frac{n\sum xy-\sum x\sum y}{n\sum x^2-(\sum x)^2} = \frac{6(844)-(75)(66)}{6(955)-(75)^2} = 1.0857$$

$$\hat{a} = \overline{y} - \hat{b}\,\overline{x} = 11 - 1.0857(12.5) = -2.57$$

وبالتالي معادلة الانحدار تصبح بالشكل:

$$\hat{y} = -2.57 + 1.0857x$$

3 – إن تقدير نسبة تركيز الكالسيوم في البول إذا كان نسبة تركيز الكالسيوم في الدم 15 هو13.7.