دولة ليبيا
وزارة التعليم
مركز المناهج التعليمية والبحوث التربوية

مبادئ الإحصاء

للسنة الثانية بمرحلة التعليم الثانوي
«القسم العلمي»

إعداد إدارة المناهج

1440 – 1441هـ
2019 – 2020 م

قائمة المحتويات

الموضوع   الصفحة

الفصل الأول

(1-1) المقدمة 5
(2-1) مصادر البيانات الإحصائية 7
(3-1) أنواع البيانات الإحصائية 8
(4-1) جمع البيانات الإحصائية 8
تمارين (1) 11

الفصل الثاني

العرض الجدولي للبيانات الإحصائية

(1-2) جداول التوزيعات التكرارية 12
أولاً – جدول التوزيع التكراري لبيانات وصفية 13
ثانياً – جدول التوزيع التكراري لبيانات كمية 15
جدول التوزيع التكراري لبيانات منفصلة 15
جدول التوزيع التكراري لبيانات متصلة 20
(2-2) جداول التوزيعات التكرارية المتجمعة 25
(3-2) جداول التوزيعات التكرارية النسبية 28
تمارين (2) 32

الفصل الثالث

العرض البياني للبيانات الإحصائية

(1-3) القطاعات الدائرية 36
(2-3) الأعمدة البيانية 38
(3-3) المدرج التكراري 42
(4-3) المضلع التكراري 46
(5-3) المنحنى التكراري 49

الموضوع الصفحة

(6-3) المنحنى التكراري المتجمع الصاعد والهابط 50
تمارين (3) 56

الفصل الرابع

مقاييس النزعة المركزية (المتوسطات)

(1-4) الوسط الحسابي (المتوسط الحسابي) 60
(2-4) الوسط الحسابي المرجح 70
(3-4) الوسيط 73
(4-4) المنوال 84
تمارين (4) 91

الفصل الخامس

مقاييس التشتت

(1-5) المدى 96
(2-5) التباين...99
(3-5) الانحراف المعيارى...103
(4-5) معامل الاختلاف 107
تمارين (5) 110

الفصل السادس

الارتباط والانحدار

(1-6) الارتباط 113
(1-1-6) الارتباط البسيط 113
(2-1-6) الارتباط الخطي البسيط 115
أ. معامل ارتباط بيرسون 115
ب.معامل ارتباط الرتب لسبيرمان 119
(2-6) الانحدار 124
(1-2-6) الانحدار الخطي البسيط 126
تمارين (6) 131

الفصل الأول

(1-1) المقدمة :

إن علم الإحصاء هو علم قائم بحد ذاته له قوانينه وقواعده الرياضية الخاصة به ، ومجالات تطبيقه تشمل جميع فروع العلوم الطبيعية والاجتماعية والتطبيقية.

وقد كان علم الإحصاء في بدايته يهتم فقط بعملية العد والحصر للأشياء ومن هنا جاءت تسميته العربية (إحصاء) فهي مشتقة من كلمة أحصى وتعني استخدام الحصى أو الحجارة الصغيرة كوسيلة بدائية لعد الأشياء الكثيرة ، فقد كان الإنسان قديماً يستعين بالحصى في عملية العد .

وكان علم الإحصاء مقصوراً على تعدادات السكان وثرواتهم وعدد المواليد والوفيات لمعرفة القوى البشرية المتوفرة لدى الدولة، وذلك للاهتداء بها في تصريف أمور الدولة ورسم سياستها ، وحيث إن الإحصاء كان مقصوراً على الحقائق الخاصة بالدولة فمن هنا جاءت تسميته باللغة الأجنبية (Statistics) فهي مشتقة من كلمة State أي الدولة .

وبمرور الزمن تطور الإحصاء فشمل إلى جانب التعدادات ، تجميع المعلومات عن الظواهر المختلفة في جميع المجالات ، واستخدمت الطريقة الرقمية للتعبير عن هذه الظواهر ، وبالتالي أصبح من السهل الاستعانة بالنظريات الرياضية لشرح الأساليب الإحصائية المستخدمة ، فنظريات ومبادئ الإحصاء تعتمد بدرجة كبيرة جداً على فروع الرياضيات المختلفة كالتفاضل والتكامل والهندسة التحليلية والجبر ... إلخ ، وقد عكف عدد من العلماء على دراسة واستنباط نظريات هذا العلم ، وكيفية تطبيقها في العلوم الأخرى ، وبفضل هذه المجهودات أصبح علم الإحصاء علماً مستقلاً له نظرياته وقواعده ، وفي الآونة الأخيرة تطور علم الإحصاء تطوراً ملموساً في مجال التطبيق وذلك بسبب انتشار الحاسبات الآلية التي أصبح لها دورمهمّ في سهولة وسرعة معالجة حجم كبيرمن البيانات وتطبيق الأساليب الإحصائية المتعددة عليها.

تعريف علم الإحصاء :

علم الإحصاء هو العلم الذي يبحث في المبادئ والقوانين والطرق العلمية المختلفة لجمع البيانات وتبويبها وعرضها وتحليلها، ثم استخلاص النتائج والتعميمات والتوصل إلى أحسن القرارات لحل المشاكل المختلفة على أساس من التحليل العلمي للبيانات المتاحة .

فروع علم الإحصاء:

لقد تم تطويرعلم الاحصاء وأصبحت له فروع عديدة يمكن تقسيمها إلى فرعين :

1 – الإحصاء الوصفي :

وهو يهتم بطرائق جمع ووصف وتلخيص البيانات الإحصائية باستخدام الجداول التكرارية والرسومات البيانية وبعض المقاييس الإحصائية لتوضيح المعالم الأساسية لهذه البيانات .

2 – الإحصاء الاستنتاجي :

يختص هذا الفرع باستنتاج واتخاد القرارات حول بعض الظواهر وتعميم النتائج المتحصل عليها مع حساب درجة الثقة المصاحبة لتلك القرارات والاستنتاجات .

أهمية علم الإحصاء وعلاقته بالعلوم الأخرى :

اكتسب علم الإحصاء أهميته من إمكانية تطبيق نظرياته ومبادئه وأساليبه في كل مجال يمكن التعبير عن ظواهره ببيانات عددية أي مجال تكون فيه المعلومات المجمعة عن الظاهرة محل البحث معلومات عددية أو حتى غير عددية ، ولكن يمكن إعادة صياغتها وتحويلها إلى معلومات عددية .

وهكذا أصبح بالإمكان استخدام الأساليب الإحصائية وتطبيقها في مختلف العلوم ، ففي علم الاقتصاد استخدم علم الإحصاء لتفسير الظواهر الاقتصادية المختلفة كنظريات العرض والطلب والعلاقة بين مستويات الدخل والإنفاق الاستهلاكي ، ونوع هذه العلاقة وكيفية قياسها . وفي مراقبة إنتاج المصانع من حيث كمية ودرجة جودته ومدى ملاءمته لاحتياجات السوق وأذواق المستهلكين ، وغيرها من الدراسات الاقتصادية .

أما في علم النفس فقد استخدمت الطرق الإحصائية في قياس ذكاء الأشخاص ، وفي دراسة العلاقة بين ذكاء الأشخاص ومهاراتهم ...إلخ ، واعتمد علم الأحياء على علم الإحصاء في دراسة الأجناس البشرية والفصائل الحيوانية والنباتية المختلفة ودراسة خواص كل منها ، وعلاقتها ببعضها .

أما في علم الوراثة استخدمت الأساليب الإحصائية لدراسة انتقال الصفات من الآباء إلى الأبناء وإمكانية التمييز بين الصفات الوراثية والصفات المكتسبة .

كما استخدمت الطرق الإحصائية في علم الفلك وفي الدراسات الخاصة بتحديد مدارات الكواكب والنجوم وغيرها من الأجرام السماوية . وللإحصاء أهمية كبيرة في الجغرافيا بشقيها الطبيعي والبشري ، فقد استخدمت الأساليب الإحصائية في دراسة أشكال سطح الأرض والجغرافيا المناخية وجغرافيا البحار والمحيطات ، هذا فضلاً عن تطبيق الطرق الإحصائية في الجغرافيا السكانية وجغرافيا المدن وعلم الخرائط .... إلخ .

أما بالنسبة للعلوم الطبية فيطبق علم الإحصاء في أغلبية الدراسات الطبية لمقارنة الأمراض المختلفة وسبل علاجها وتحديد العلاقة بين بعض الأمراض ومسبباتها .وهكذا نرى أن علم الإحصاء قد أصبح في وقتنا الحاضر علماً مهماً كثير النفع والاستعمال في شتى أنواع العلوم، وبصفة عامة فإن أي تخطيط اقتصادي أو اجتماعي لا يمكن أن يكون تخطيطاً صحيحاً ما لم تستخدم الأساليب والمبادئ والنظريات الإحصائية في تنفيذه وتقييم نتائجه .

أهداف علم الإحصاء :

1. تبسيط البيانات الإحصائية بعرضها في جداول أو رسومات بيانية ، وذلك لتسهيل فهمها وتلخيصها وتحليلها .
2. التعبير عن الحقائق بصورة عددية واضحة ودقيقة بدلاً من عرضها والتعبير عنها بطريقة إنشائية .
3. مقارنة المجموعات المختلفة وإيجاد العلاقات القائمة بينها .
4. التنبؤ ببيانات مستقبلية مما يساعد في عملية التخطيط .
5. استخلاص النتائج واتخاذ القرارات المناسبة بقدر كبير من الصحة ، وذلك بعد قيام الباحث في أي فرع من فروع العلوم المختلفة بتحليل البيانات المتوفرة لديه .

(2-1) مصادر البيانات الإحصائية :

هناك عدة مصادر للحصول على البيانات تختلف باختلاف موضوع الدراسة والغرض منها .

1. النشرات والدوريات والسجلات .
2. التجارب .
3. الاستبيانات .
4. التعدادات العامة .

بعض المصطلحات الإحصائية :

أ. **المجتمع الإحصائي (قيد الدراسة) :** هو عبارة عن مجموعة من العناصر (الوحدات) المشتركة في الصفة التي تهم الباحث .

ب. **العينة :** هي جزء من المجتمع الإحصائي يتم اختيار عناصرها غالباً بطريقة عشوائية (دون تدخل من الباحث)

ج. **الظاهرة :** هي صفة لعناصر تختلف من عنصر لآخر في الشكل أو النوع أو الكمية ويطلق عليها أيضاً تسمية متغير .

د. **المعلمة والإحصاءة:** المقادير الخاصة بالمجتمع تسمى معلمات بينما المقادير الخاصة بالعينة تسمى إحصاءات .

(3-1) أنواع البيانات الإحصائية :

يمكن تصنيف البيانات الإحصائية إلى صنفين :

أ. **بيانات نوعية (كيفية أو وصفية) :** وفيها تكون الظاهرة قيد الدراسة مقسمة إلى صفات أوأنواع أو أزمنة وهي بدورها تنقسم إلى قسمين :

1. بيانات نوعية قابلة للترتيب مثل المستوى التعليمي، تقديرات النجاح ...الخ .
2. بيانات نوعية غير قابلة للترتيب مثل الجنسية ، اللون...الخ .

ب. **بيانات كمية (عددية)** : وهي بيانات عن ظواهر يتم التعبير عنها عددياً وهي بدورها تنقسم إلى قسمين :

1. بيانات كمية منفصلة : وهي بيانات عن ظواهر تأخذ أعداداً صحيحة فقط مثل عدد المواليد، عددالسيارات ،عدد الطلبة ...الخ .
2. بيانات كميّة متصلة : وهي بيانات عن ظواهر تأخذ قيماً متّصلة (عدد صحيح وكسر من وحدة القياس) مثل الطول والوزن والزمن ... الخ .

البيانات الكمية القابلة للعد تكون منفصلة، والقابلة للقياس تكون متصلة .

(4-1) جمع البيانات الإحصائية :

هناك عدة طرائق لجمع البيانات الإحصائية أهمها :

1. **طريقة المسح (الحصر) الشامل :**

لاتّباع هذه الطريقة يتطلب الأمر جمع البيانات من كل الوحدات المكونة للمجتمع محل البحث.

قبل البدء في عملية جمع البيانات الإحصائية يجب على الباحث تحديد الهدف من الدراسة التي سيقوم بها تحديداً دقيقاً ، وعلى ضوء هذا يقوم بتحديد البيانات المطلوبة والضرورية لبحثه ، ثم يقوم بتصميم استمارة إحصائية وهي عبارة عن قائمة تشمل أسئلة تكون الإجابة عنها مؤدية إلى البيانات المطلوبة في البحث. وتوجد بعض الاعتبارات التي يجب مراعاتها عند تصميم الاستمارة الإحصائية والمتمثلة فيما يلي :

1. أن تحتوي الاستمارة على أقل عدد ممكن من الأسئلة بشرط أن تعطي جميع المعلومات المطلوبة في البحث .
2. يجب أن تكون الأسئلة بسيطة واضحة المعنى ، كما يفضل اختيار الألفاظ الشائعة المفهومة .
3. من الأفضل أن تصاغ الأسئلة بحيث تكون إجابتها بنعم أو لا، أو إجابة من ثلاث أو أربع إجابات ، يصيغها مصمم الاستمارة ويضع الشخص الذي يملأ الاستمارة علامة أمام الإجابة الّتي تناسبه .
4. يجب تجنب الأسئلة المحرجة والمتعلّقة بالأمور الشخصية السريّة كالسؤال عن الأرباح مثلاً .

5. تجنّب الأسئلة الإيحائية ، أي التي توحي بإجابات مُعيّنة مثلاً :هل تغيبت عن العمل بسبب المرض ؟ فالإجابة عن هذا النّوع من الأسئلة تكون غالباً بالإيجاب والطريقة الأفضل للسؤال : لماذا تغيبت عن العمل؟
6. تجنب الأسئلة التي تكون إجابتها نسبية مثل السؤال عن الثقافة ، عن التدين ... إلخ .
7. يجب أن يخصّص لكل سؤال نقطة واحدة للإجابة عنها ، فيجب عدم وضع سؤالين أو أكثر في نفس السؤال . ويجب ترتيب الأسئلة ترتيباً منطقياً وإعطاء رقم لكل سؤال حتى يسهل الرجوع إليه .
8. يجب تكرار الأسئلة التي يرى الباحث أن إجابتها مهمّة جداً ، وذلك بأن يعبّر عن السؤال بأكثر من صيغة في مواضع متباعدة من الاستمارة ، وذلك للتأكد من صحة المعلومات التي يعطيها المستجوب، أي: الشخص الذي يملأ الاستمارة .

أنواع الاستمارة الإحصائية :

تنقسم الاستمارات الإحصائية إلى نوعين هما :

1 - كشف البحث :

هو استمارة إحصائية يقوم الباحث أو الشخص المكلف بجمع البيانات بتدوين الإجابة عن الأسئلة التي بها بنفسه وذلك بعد أن يتحصل على الإجابات عن طريق الملاحظة المباشرة للظاهرة محل الدراسة ، كما هو الحال في العلوم المعملية أو عن طريق المقابلة الشخصية للمبحوث أو عن طريق الهاتف حيث إنه عند جمع البيانات بهذه الطرق يقوم الباحث بتوجيه الأسئلة للمبحوث ثم يسجل الباحث إجابات الشخص المبحوث كما هي ، ويستخدم كشف البحث في الدراسات الخاصة بالمجتمعات التي ترتفع فيها نسبة الأمية وكذلك في الدراسات الخاصة بالأطفال .

2 – صحيفة الاستبيان :

هي استمارة إحصائية يقوم المبحوث أي الشخص الذي لديه المعلومات بتدوين الإجابة بنفسه عن الأسئلة التي بها ، في هذه الحالة يقوم الباحث بتسليم الاستمارة للمبحوث ثم يستعيدها منه بعد ملئها أو يرسلها إليه بواسطة البريد العادي أوالإلكتروني ويطلب منه تعبئتها وإعادتها ، ولتشجيع المستجوب على إعادة الاستمارة يجب أن يرفق معها ظرفاً عليه طابع بريد وعنوان الباحث أو الهيئة المشرفة على الدراسة ، أويطلب منه تعبئة الاستمارة إلكترونياً وإعادتها بواسطة البريد الإلكتروني للباحث أو الجهة المشرفة على الدراسة .

2.طريقة العينات:

إذا تعذّر استخدام طريقة المسح الشامل لأسباب اقتصادية أو عملية يتم اختيار جزء من المجتمع محل البحث بطريقة عشوائية دون تدخل من قبل الباحث وبالإمكان تعميم نتائج العينة العشوائية باستخدام الإحصاء الاستنتاجي على المجتمع ككل مع مراعاة أنه كلما زاد حجم العينة وكلما تم اتباع الأسلوب الإحصائي المناسب كانت نتائج العينة قريبة من حقائق المجتمع وأكثر مصداقية .

تمارين (1)

1. عرف علم الإحصاء ، مع ذكر أهدافه .
2. تكلم عن فروع علم الإحصاء .
3. ما هي الطرق المتبعة لجمع البيانات ؟
4. في كل دراسة من الدراسات التالية ، وضح إذا كانت البيانات المتحصل عليها بيانات وصفية أم كمية ؟

أ – دراسة خاصة بمرتبات موظّفي في إحدى الشركات.

ب – دراسة خاصة بأطوال طلبة السنة الأولى في المرحلة الابتدائية .

ج – دراسة خاصة بالعلاقة بين لون شعر الأب ولون شعر الابن .

د – دراسة خاصة بأوزان معلبات تنتجها مصانع المجمعات الغذائية في إحدى المدن.

هـ - دراسة خاصة بجنسية المرضى الأجانب الذين يترددون على أحد المستشفيات .

5. في كل حالة من الحالات التالية ، اذكر ما إذا كان المتغير منفصلاً أو متصلاً :

أ – عدد السيارات التي تدخل أحدى المدن الجامعية يومياً .

ب – درجات الحرارة المسجلة ظهراً في إحدى المناطق .

ج – عدد المواليد في مدينة ما يومياً .

د – أوزان المرضى الذين يترددون على عيادة مرضى السكر بإحدى المدن .

هـ - أعمار الطلبة المشتركين في امتحان الثانوية العامة لسنة معينة.

و – الوقت الذي يستغرقه الطلبة في حل امتحان معين .

6. تكلم عن أنواع الاستمارة الإحصائية ؟

الفصل الثاني

العرض الجدولي للبيانات الإحصائية

(2-1) جداول التوزيعات التكرارية :

إن البيانات التي تم جمعها من أو عن مفردات المجتمع محل الدراسة وقبل أن يجرى عليها أي تنظيم أو تبويب تسمى بيانات خام غير مبوبة ، والبيانات الخام لا توضح لنا اتجاهات الظاهرة المدروسة ، بل لا نستطيع أن نجري عليها أي تحليل رياضي لحساب المقاييس الإحصائية المختلفة ، ولذلك فالخطوة التالية بعد عملية جمع البيانات هي تنظيمها وتبويبها وعرضها في جدول؛ ليسهل فهمها وتحليلها واستخلاص النتائج منها ، ويطلق على هذا الجدول جدول التوزيع التكراري ، ويطلق على البيانات بعد عرضها في جدول توزيع تكراري بيانات مبوبة .

تعريف جدول التوزيع التكراري :

يعرف جدول التوزيع التكراري بأنه جدول ذو عمودين ، العمود الأول نقسم فيه البيانات إلى فئات مصنفة حسب النوع أو القيمة العددية والعمود الثاني نسجل فيه أمام كل فئة عدد القيم التابعة لها ، ويسمى هذا العدد بتكرار الفئة .

كيفية إعداد الجداول التكرارية :

إن أول خطوة لتبويب البيانات الخام ووضعها في جداول تكرارية هي عمل ما يسمى بجدول تفريغ البيانات ،وهو جدول مقسم إلى ثلاثة أعمدة حيث إن :

- العمود الأول تدون فيه الفئات مصنفة حسب النوع أو القيمة العددية، وإذا كان ترتيب البيانات ممكناً فيجب أن تكون الفئات مرتبة ترتيباً تصاعدياً أو تنازلياً .
- العمود الثاني خاص بالعلامات، حيث نقوم بقراءة القيم المشاهدة المذكورة في البيانات والمقصود بالقيم المشاهدة أو المشاهدات هي قيم المتغير التي تحصلنا عليها من أو عن المفردات محل الدراسة عند جمع البيانات – ونضع هذه العلامة (/) أمام الفئة التي تشمل القيمة ، ولتسهيل عملية عد العلامات يستحسن أن نضع كل أربع علامات بجوار بعضها (////) أما العلامة الخامسة فتكتب عكس اتجاه العلامات الأربع على هذا النحو ($\cancel{////}$) . وبذلك يكون لدينا حزم من العلامات كل منها يحتوي على خمس علامات .

• العمود الثالث يُسجل به عدد العلامات التابعة لكل فئة وهذا العدد هو الذي يمثل التكرار.
وبذلك نكون قد تحصلنا على تكرار كل فئة من فئات الجدول ، وبالطبع مجموع التكرارات يجب أن يساوي العدد الكلي للقيم المشاهدة .

وبعد الانتهاء من جدول التفريغ نستطيع معرفة جدول التوزيع التكراري لهذه البيانات فهو عبارة عن العمود الأول والثالث لجدول التفريغ مع حذف العمود الخاص بالعلامات . وفيما يلي سنقوم بدراسة كيفية إعداد الجداول التكرارية لكل نوع من أنواع البيانات .

أولاً – جدول التوزيع التكراري لبيانات وصفية " كيفية " :

إذا كانت البيانات وصفية مثل مستوى التعليم ، الديانة ، الجنسية ، ... إلخ ، فإننا نصنف البيانات في فئات بحيث تمثل كل فئة صفة أو نوعاً ، ثم نقوم بإعداد جدول تفريغ البيانات ومنه نحصل على جدول التوزيع التكراري ، والمثال التالي يوضح ذلك .

مثال (2-1) :

البيانات الخام التالية توضّح المستوى الدراسي لعشرين عاملة في أحد المصانع:

أمي	إعدادي	أمي	ابتدائي	ابتدائي
ثانوي	إعدادي	إعدادي	ابتدائي	أمي
أمي	أمي	ابتدائي	ابتدائي	ابتدائي
إعدادي	ابتدائي	إعدادي	ابتدائي	إعدادي

اعـرض هذه البيانات في جدول تكراري .

الحلّ :

البيانات السابقة بيانات وصفية تمثّل متغيراً وصفياً وهو المستوى التعليمي والذي ينقسم إلى 4 أنواع هي : أمي ، ابتدائي ، إعدادي ، ثانوي ، ويمكننا أن نضعها في جدول تفريغ البيانات بحيث تمثل كل فئة مستوى تعليمياً مع ترتيب الفئات ترتيباً تصاعدياً أو تنازلياً ، وعندها سيكون لدينا 4 فئات ، ثم نسجل أمام كل فئة العلامات التي تمثل المفردات التي تتبعها ويكون عدد العلامات هو تكرار الفئة وبالطبع يجب أن يكون مجموع التكرارات يساوي العدد الكلي للعاملات ، وجدول (2-1) يوضح جدول تفريغ البيانات المذكورة .

جدول (2-1)

جدول تفريغ للمستويات التعليمية لعشرين عاملة

المستوى التعليمي(الفئة)	العلامات	عدد العاملات(التكرار)
أمي	~~////~~	5
ابتدائي	~~////~~ ///	8
إعدادي	~~////~~ /	6
ثانوي	/	1
المجموع		20

ومن جدول تفريغ البيانات نحصل على جدول التوزيع التكراري المناظر عن طريق الاكتفاء بالعمود الأول والثالث وحذف العمود الخاص بالعلامات والموضح فيما يلي :

جدول (2-2)

جدول التوزيع التكراري للمستوى التعليمي لعشرين عاملة

المستوى التعليمي (الفئة)	عدد العاملات (التكرار)
أمي	5
ابتدائي	8
إعدادي	6
ثانوي	1
المجموع	20

وبعد وضع البيانات الخام في جدول تكراري أصبحت واضحة وسهلة الفهم ، فنفهم من الجدول أنه يوجد في المصنع 5 عاملات أميات ، و8 عاملات مستواهن ابتدائي ، و6 عاملات مستواهن إعدادي وهكذا .

ثانياً – جدول التوزيع التكراري لبيانات كمية :

(1) جدول التوزيع التكراري لبيانات منفصلة (متقطعة) :

إذا كانت البيانات كميّة وتمثل متغيراً منفصلاً فتوجد طريقتان لإعداد الجداول التكرارية :

(أ) **إذا كان عدد القيم المختلفة التي يأخذها المتغير عدداً محدوداً وصغيـراً** ، فنكوّن جدولاً تكرارياً تمثل كل فئة من فئاته قيمة واحدة من القيم التي يأخذها المتغير مع مراعاة ترتيب القيم ترتيباً تصاعدياً، ثم نقوم بإعداد جدول لتفريغ البيانات ونشتق منه جدول التوزيع التكراري .

مثال (2-2) :

إذا ألقينا 4 مكعبات نرد معا 30 مرة وحصلنا على البيانات التالية التي تمثل عدد المكعبات التي يظهر عليها الرقم 6 في كل رمية .

0 ، 1 ، 4 ، 1 ، 2 ، 0 ، 1 ، 0 ، 0 ، 1 ، 2 ، 3 ، 1 ، 0 ، 2 ،

3 ، 0 ، 2 ، 1 ، 0 ، 3 ، 1 ، 0 ، 1 ، 0 ، 1 ، 4 ، 2 ، 1 ، 2.

الحلّ :

المتغير محل الدراسة في هذا المثال هو عدد المكعبات التي يظهر عليها الرقم 6 في كل رمية ، وهو متغير عشوائي منفصل والقيم المختلفة التي يأخذها هي :

0 ، 1 ، 2 ، 3 ، 4

حيث المقصود بكل قيمة من هذه القيم ما يلي :

0 : لا يظهر الرقم 6 على أي مكعب من المكعبات الأربعة .

1 : يظهر الرقم 6 على مكعب واحد فقط من المكعبات الأربعة .

2 : يظهر الرقم 6 على مكعبين من المكعبات الأربعة .

3 : يظهر الرقم 6 على ثلاثة مكعبات من المكعبات الأربعة .

4 : يظهر الرقم 6 على المكعبات الأربعة .

وحيث إن عدد القيم المختلفة التي يأخذها المتغير يساوي 5 وهو عدد محدود وصغيروبالتالي سنجعل كل فئة من فئات جدول التوزيع التكراري تمثل قيمة واحدة فقط من هذه القيم ، وسيكون جدول تفريغ البيانات كما يلي :

جدول (2-3)

جدول تفريغ عدد المكعبات التي ظهر عليها الرقم 6 في ثلاثين رمية.

عدد المكعبات التي ظهر عليها الرقم 6 (الفئة)	العلامات	عددالرميات (التكرار)
0	~~////~~ ////	9
1	~~////~~ ~~////~~	10
2	~~////~~ /	6
3	///	3
4	//	2
المجموع		30

ومن جدول تفريغ البيانات نحصل على جدول التوزيع التكراري المناظر .

جدول (2-4)

جدول التوزيع التكراري لعدد المكعبات التي ظهر عليها الرقم 6 في ثلاثين رمية لأربع مكعبات

عدد المكعبات التي ظهر عليها الرقم 6 (الفئة)	عدد العاملات (التكرار)
0	9
1	10
2	6
3	3
4	2
المجموع	30

ومن جدول التوزيع التكراري نفهم بسهولة أن عدد الرميات التي لم يظهر فيها الرقم 6 على أي مكعب من المكعبات الأربعة هي 9 رميات من 30 رمية وعدد الرميات التي ظهر فيها الرقم 6 على مكعب واحد فقط من المكعبات الأربعة هي 10 رميات وهكذا ... وهذا لا يمكن

فهمه من البيانات الخام بسهولة بالإضافة إلى أننا نستطيع حساب مقاييس إحصائية مهمة وكثيرة عند وضع البيانات الخام في صورة جدول توزيع تكراري .

(ب) إذا كانت البيانات المنفصلة تمثل متغيراً منفصلاً يأخذ عدداً كبيراً من القيم، ففي هذه الحالة نجعل كل فئة في الجدول تمثل عدداً من القيم أي مجموعة من القيم بدلاً من قيمة واحدة ، وذلك حتى لا يكون الجدول مطولاً مما يؤدي إلى تشتت المعلومات فيه ، ولتكوين الجدول التكراري في هذه الحالة نتبع الخطوات التالية :

خطوات إنشاء جداول التوزيعات التكرارية :

1- نحدد المدى الذي تنتشر فيه القيم المشاهدة للبيانات، وذلك بطرح أصغر قيمة في البيانات (الحد الأدنى للبيانات الخام) من أكبر قيمة فيها (الحد الأعلى للبيانات الخام) ويرمز له بالرمز (R) .

المدى = الحد الأعلى للبيانات – الحد الأدنى للبيانات

$$R = \text{max value} - \text{min value}$$

2- نقسم المدى إلى عدد من الفئات وعادة يفضل أن يتراوح عددها من 5 إلى 15 فئة ، ويفضل أن تكون متساوية الطول ، والمقصود بطول الفئة في حالة المتغير المنفصل هو عدد القيم التي تمثلها الفئة ، ونقوم بتحديد طول الفئة كما يلي :

طول الفئة = المدى ÷ عدد الفئات

أو **$L = R/N$** ، حيث (**N**) عدد الفئات

وإذا كان خارج القسمة كسرٌ فنستعمل العدد الصحيح الأكبر منه مباشرة حتى لا تكون حدود فئات الجدول محتوية على كسور مما يجعل الجدول في صورة معقدة وغير واضحة ويرمز لطول الفئة بالرمز (**L**) .

3- بعد تحديد طول الفئة نحدد الحد الأدنى والحد الأعلى لكل فئة ، حيث يكون الحد الأدنى للفئة الأولى أقل أو يساوي أصغر قيمة مشاهدة في البيانات الخام ، والحد الأعلى للفئة الأخيرة أكبر من أو يساوي أكبر قيمة مشاهدة في البيانات الخام ، وذلك حتى نضمن بأن فئات الجدول التكراري ستشمل كل القيم المشاهدة .

4- نكوّن جدول تفريغ البيانات الذي تحدثنا عنه سابقاً ، ومنه نشتق جدول التوزيع التكراري عن طريق الاكتفاء بالعمود الأول والثالث وحذف العمود الخاص بالعلامات . والمثال

التالي يوضح هذه الخطوات .

مثال (2-3) :

البيانات التالية تمثل عدد زبائن محل تجاري في إحدى المدن يومياً خلال شهرين متتاليين :

40	57	60	55	52	50	45	38	41	39
44	53	56	36	35	60	50	52	44	40
54	46	45	49	41	48	50	49	47	55
47	48	47	62	58	54	48	46	45	51
41	43	36	39	44	46	48	49	51	58
49	46	43	47	48	61	39	38	43	44

اعرض هذه البيانات في جدول توزيع تكراري يتكون من 6 فئات متساوية الطول .

الحل :

لتكوين جدول التوزيع التكراري لهذه البيانات نتبع الخطوات المذكورة سابقاً ، وذلك كما يلي :

- نحدد الحد الأعلى والحد الأدنى للبيانات ثم نحسب قيمة المدى ، فنجد أن :

الحد الأعلى = 62 ، الحد الأدنى = 35

إذن : المدى = 62 - 35 = 27 أو $R = 62 - 35 = 27$

- بما أننا نريد تقسيم هذا المدى إلى 6 فئات فسيكون طول كل فئة كما يلي :

طول الفئة = 27 ÷ 6 = $4\frac{3}{6}$ أو $L = 27 \div 6 = 4\frac{3}{6}$

وعندها يكون طول الفئة مساوياً 5 ، أي كل فئة تمثل 5 قيم من القيم التي يمكن أن يأخذها المتغير .

حيث إن أقل قيمة في البيانات 35 فسيكون الحد الأدنى للفئة الأولى 35 ، وبما أن طول الفئة يساوي 5 ، إذن الفئة الأولى ستمثل القيم التالية :

35 ، 36 ، 37 ، 38 ، 39

وبالتـالي يكون الحد الأعلى للفئة الأولى مساوياً 39، وتكتب هذه الفئة كما يلي :

35 – 39 ، أي تمثل قيم المتغير من 35 إلى 39 (بما فيها 35 و 39)، وبنفس الطريقة نكتب الفئات الأخرى ، وعندئذ ستكون الفئات السّت هي :

35 - 39 وتمثل القيم 35 ، 36 ، 37 ، 38 ، 39 .
40 - 44 وتمثل القيم 40 ، 41 ، 42 ، 43 ، 44 .
45 - 49 وتمثل القيم 45 ، 46 ، 47 ، 48 ، 49 .
50 - 54 وتمثل القيم 50، 51 ، 52 ، 53 ، 54 .
55 - 59 وتمثل القيم 55، 56 ، 57 ، 58 ، 59 .
60 - 64 وتمثل القيم 60، 61 ، 62 ، 63 ، 64 .

كما ذكرنا سابقاً أن الحد الأعلى للفئة الأخيرة يجب أن يساوي أو يزيد عن أكبر قيمة في البيانات الخام ، نحدد عدد القيم المشاهدة التابعة لكل فئة من الفئات الستة ؛ أي نحدد تكرار كل فئة ، وذلك باستخدام جدول تفريغ البيانات المشار إليه سابقاً ، فنمثل القيمة المشاهدة الأولى في البيانات وهي 39 بعلامة أمام الفئة التي تحتويها وهي الفئة الأولى ، والقيمة الثانية وهي 41 بعلامة أمام الفئة التي تحتويها وهي الفئة الثانية ، والقيمة الثالثة وهي 38 نمثلها بعلامة أمام الفئة الأولى ، ... وهكذا ، فنحصل على جدول تفريغ البيانات التالي :

جدول (2-5)

جدول تفريغ لعدد الزبائن لمحل تجاري في 60 يوماً

الفئة	العلامات	التكرار
35 – 39	/// ~~////~~	8
40 – 44	// ~~////~~ ~~////~~	12
45 – 49	~~////~~ ~~////~~ ~~////~~ ~~////~~	20
50 – 54	~~////~~ ~~////~~	10
55 – 59	/ ~~////~~	6
60 - 64	////	4
المجموع		60

ومن جدول تفريغ البيانات نحصل على جدول التوزيع التكراري المناظر عن طريق الاكتفاء بالعمود الأول والثالث وحذف العمود الخاص بالعلامات، وذلك كما هو واضح في جدول (2 – 6).

جدول (2-6)

جدول التوزيع التكراري لزبائن محل تجاري في 60 يوماً

عدد الزبائن(الفئة)	عدد الأيام (التكرار)
35 – 39	8
40 – 44	12
45 – 49	20
50 – 54	10
55 – 59	6
60 - 64	4
المجموع	60

وبسهولة نفهم من هذا الجدول أنه في 8 أيام من هذين الشهرين كان عدد زبائن هذا المحل يتراوح بين 35 إلى 39 زبوناً، وفي 12 يوماً كان عدد الزبائن يتراوح بين 40 إلى 44 وهكذا ... وهذه المعلومات بالطبع لم تكن واضحة عندما كانت البيانات خاماً أي غير مبوبة في جدول تكراري .

(2) جدول التوزيع التكراري لبيانات متصلة (مستمرة) :

إذا كانت البيانات كمية وتمثل متغيراً متصلاً ، فلقد علمنا سابقاً أن المتغير المتصل يمكنه أن يأخذ كل القيم الموجودة بمدى معين ، لذلك عند إعداد جدول تكراري لبيانات تمثل متغيراً متصلاً سنتعامل مع فترات وليس مع قيم محددة كما في حالة المتغير المنفصل ، أي كل فئة ستمثل فترة بها عدد لا نهائي من القيم المتصلة ، ولتكوين هذا الجدول نتبع نفس الخطوات المذكورة سابقاً في حالة المتغير المنفصل مع مراعاة الفرق بين النوعين عند وضع حدود الفئات .

في المثال السابق كان المتغير عبارة عن عدد الزبائن وهو متغير منفصل ، ولذلك قسمنا الفئات بحيث إن الفئة الأولى تبدأ من 35 وتنتهي عند 39 ، والفئة الثانية تبدأ من 40 وتنتهي عند 44، لأن المتغير في هذا المثال لا يأخذ القيم المحصورة بين 39 و 40 كالقيمة 39.72 مثلاً ، ولكن هذا التقسيم الذي يحتوي على فجوات بين الفئات لا يناسب المتغير المتصل؛ لأن المتغيرالمتصل يمكنه أن يأخذ أي قيمة بين الحد الأدنى والحد الأعلى للبيانات ، لذلك عند تكوين جدول التوزيع التكراري لمتغير متصل يجب أن نجعل كل الفئات متصلة ببعضها وليس

بينها فجوات ، فكل فئة تنتهي عند بداية الفئة اللاحقة لها ، وبذلك تكون الفئات شاملة لجميع القيم التي يمكن أن يأخذها المتغير المتصل .
وطول الفئة في حالة المتغير المتصل هو طول الفترة أو المسافة التي تمثلها هذه الفئة ، ونحصل على طول الفئة بطرح حدها الأدنى من حدها الأعلى ، أي أن :

طول الفئة = الحد الأعلى للفئة - الحد الأدنى للفئة

ويفضل أن يكون طول الفئة خالياً من الكسور لتسهيل فهم الجدول وتسهيل العمليات الحسابية اللازمة عند حساب المقاييس الإحصائية ، وعند ظهور كسر نقربه إلى أعلى ، فمثلاً إذا كان طول الفئة يساوي 7.23 فنعتبره 8 وهكذا ....
ثم نقوم بتحديد حدود الفئات ، فيجب أن يكون الحد الأدنى مساوياً أقل قيمة في البيانات وإذا كانت أقل قيمة تحتوي على كسر فنعتبر القيمة الصحيحة التي أقل منها مباشرة كحد أدنى للفئة الأولى، وذلك لتسهيل الجدول وتسهيل العمليات الحسابية ، ثم نكوّن جدول تفريغ للبيانات ونحدد تكرار كل فئة بطريقة العلامات المذكورة سابقاً . ثم نشتق منه جدول التوزيع التكراري عن طريق الاكتفاء بالعمود الأول والثالث وحذف العمود الخاص بالعلامات .

مثال (2-4) :

البيانات التالية تمثل الراتب الأسبوعي بالدينار لثلاثين موظفاً من الموظفين العاملين بإحدى الشركات :

50.239	51.020	71.312	65.981	56.801	55.231
75.021	62.839	50.231	35.900	59.012	70.900
55.923	55.210	48.311	55.212	54.999	74.923
60.100	57.390	77.610	49.201	64.981	71.231
60.998	66.000	62.321	40.230	59.030	30.510

المطلوب عرض هذه البيانات في جدول توزيع تكراري يتكون من 5 فئات .

الحل :

في هذا المثال المتغير محل الدراسة هو راتب الموظف وهو متغير متصل ، ولتكوين جدول التوزيع التكراري لهذا المتغير سنتبع نفس الخطوات التي اتبعناها في المثال (2 – 3) مع مراعاة تحديد حدود الفئات بالطريقة التي تلائم المتغير المتصل ، وذلك كما يلي :

• نحدد الحد الأعلى والحد الأدنى للبيانات ثم نحسب قيمة المدى ، فنجد أن :

الحد الأدنى = 30.510 (أقل راتب في البيانات)

الحد الأعلى = 77.610 (أعلى راتب في البيانات)

المـــــدى = 77.610 – 30.510 = 47.100

• إذا أردنا تقسيم هذا المدى إلى 5 فئات فسيكون طول كل فئة كما يلي :

طول الفئة = 47.100 ÷ 5 = 9.420

وعندئذ سنعتبر طول الفئة (الفترة) مساوياً 10 ، وحيث إن الحد الأدنى في البيانات 30.510 ، فنستطيع أن نجعل الفئة الأولى تبدأ من العدد الصحيح الأقل من الحد الأدنى مباشرة وذلك لتسهيل صورة الجدول . وبالتالي سيكون الحد الأدنى للفئة الأولى مساوياً 30 ، وحيث إن طول الفئة يساوي 10 ،فستكون الفئات كما يلي :

30 إلى أقل من 40

40 إلى أقل من 50

50 إلى أقل من 60

60 إلى أقل من 70

70 إلى أقل من 80

وبكتابة الفئات بهذه الطريقة لن يكون لدينا فجوات تفصل الفئات وبالتالي نكون قد عبرنا عن المتغير المستمر بطريقة صحيحة .

بعد تحديد الفئات ، نقوم بتكوين جدول تفريغ البيانات الموضح فيما يلي :

جدول (2-7)

جدول تفريغ لمرتبات ثلاثين موظفاً بالشركة

الفئة	العلامات	التكرار
30 إلى أقل من 40	//	2
40 إلى أقل من 50	///	3
50 إلى أقل من 60	//// //// //	12
60 إلى أقل من 70	//// //	7
70 إلى أقل من 80	//// /	6
المجموع		30

ويكون جدول التوزيع التكراري للراتب الأسبوعي لثلاثين موظفاً ، كما يلي :

جدول (2-8)

جدول التوزيع التكراري للراتب الأسبوعي بالدينار لثلاثين موظفاً

الراتب الأسبوعي بالدينار (الفئة)	عدد الموظفين (التكرار)
30 إلى أقل من 40	2
40 إلى أقل من 50	3
50 إلى أقل من 60	12
60 إلى أقل من 70	7
70 إلى أقل من 80	6
المجموع	30

مركز الفئة :

مركز الفئة هو القيمة الواقعة في منتصفها ، ويحسب كما يلي :

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

ونطبق هذه الصيغة سواء كانت البيانات منفصلة أم متصلة ، فمثلاً في المثال (2 – 3) الخاص ببيانات منفصلة ، فإن مركز الفئة الأولى يحسب كما يلي :

$\frac{39+35}{2}$ =37 ، وهكذا بالنسبة لمراكز بقية الفئات .

وفي المثال (2 – 4) الخاص ببيانات متصلة ، فإن مركز الفئة الأولى يحسب كما يلي :

$\frac{40+30}{2}$ = 35 وهكذا بالنسبة لمراكز بقية الفئات .

ومركز الفئة هو عبارة عن قيمة نظرية أي: ليس بالضروري أن يأخذها المتغير محل الدراسة ، وفي العمليات الحسابية اللازمة لبعض المقاييس الإحصائية تستعمل قيمة مركز الفئة كممثل عن كل القيم التي تحتويها الفئة ، فمثلاً في الجدول التكراري (2 – 8) الفئة الأولى تبدأ من 30 إلى أقل من 40 وتكرارها مساوٍ 2، فهذا يعني أن موظفين راتبهما الأسبوعي في هذه الفئة ولكن لا يوضح الجدول التكراري راتب كل موظف منهما ، وبافتراض أن المفردات موزعة توزيعاً عادلاً داخل كل فئة فنعتبر أن كل راتب مساوٍ لمركز الفئة والذي يساوي 35 وبالتالي سيكون مجموع الرواتب الأسبوعية للموظفين التابعين للفئة الأولى مساوياً 35 × 2 = 70 ديناراً ، وبنفس الطريقة نستطيع أن نحسب المجموع الكلي للرواتب الأسبوعية للموظفين التابعين لأي فئة في الجدول التكراري، وذلك بضرب مركز الفئة في تكرارها ، وبجمع مجاميع الرواتب الخاصة بجميع الفئات نحصل على المجموع الكلي للرواتب الأسبوعية لثلاثين موظفاً.

ونستعمل هذه المجاميع عند حساب كثير من المقاييس الإحصائية . كما سنلاحظ في الفصول القادمة من هذا الكتاب .

الجداول غير المنتظمة :

يسمى جدول التوزيع التكراري جدولاً غير منتظم إذا كانت فئاته غير متساوية الطول ، أما إذا كانت فئاته ذات أطوال متساوية فيسمى الجدول جدولاً منتظماً . وعادة تفضل الجداول المنتظمة لسهولة التعامل معها حسابياً عند حساب المقاييس الإحصائية المختلفة ، ولكن أحياناً تكون البيانات مركزة في جزء من التوزيع بينما مبعثرة في الأجزاء الأخرى ، فإذا جعلنا أطوال الفئات متساوية فسنجد أن كثيراً من الفئات خالية من التكرارات ولهذا يفضل أن تكون الفئات غير متساوية الطول فنجعل الفئات في الأجزاء التي تتركز فيها القيم ذات طول قصير بينما تكون الفئات في الأجزاء التي تتبعثر فيها القيم أوسع .

كذلك نضطر لتكوين جدول غير منتظم إذا أردنا توضيح خصائص معينة للمتغير محل الدراسة ، فمثلاً عند تكوين جدول تكراري لدرجات امتحان في مادة من المواد لتوضيح تقديرات الطلبة ، فسيكون هذا الجدول غير منتظم، وذلك لأن الفئات الخاصة بالتقديرات غير متساوية الطول حيث :

0 – 29	تقدير ضعيف جداً	65 – 74	تقدير جيد
30 – 49	تقدير ضعيف	75 – 84	تقدير جيد جداً
50 – 64	تقدير مقبول	85 – 100	تقدير ممتاز

الجداول المفتوحة :

يسمى جدول التوزيع التكراري جدولاً مفتوحاً إذا لم يُحدد الحد الأدنى للفئة الأولى أو الحد الأعلى للفئة الأخيرة أو كلاهما ، وذلك لعدم توفر البيانات اللازمة لتحديد البداية أو النهاية ، أو نظراً لوجود قيم قليلة متباعدة في أعلى التوزيع أو أسفله .
والجدول المفتوح من طرفه الأدنى هو الجدول الذي يكون فيه الحد الأدنى للفئة الأولى غير محدد ، والجدول المفتوح من طرفه الأعلى هو الجدول الذي يكون فيه الحد الأعلى للفئة الأخيرة غير محدد ، أما إذا كان الحد الأدنى للفئة الأولى غير محدد والحد الأعلى للفئة الأخيرة غير محدد فيكون الجدول مفتوحاً من الطرفين، ويمكن توضيحها بالأمثلة التالية:

مفتوح من طرفه الأدنى	مفتوح من طرفه الأعلى	مفتوح من الطرفين
أقل من 5	5 – 1	أقل من 5
10 – 6	10 – 6	10 – 6
15 – 11	15 – 11	15 – 11
20 – 16	20 – 16	20 – 16
25 – 21	21 أو أكثر	21 أو أكثر

(2-2) جداول التوزيعات التكرارية المتجمعة :

التوزيعات التكرارية العادية التي درسناها في البند السابق توضح لنا عدد المفردات التابعة لكل فئة من فئات الجدول على حدة، ولكننا أحياناً نرغب في معرفة عدد المفردات التي قيمتها أقل من قيمة معينة أو عدد المفردات التي قيمتها أكثر من أو تساوي قيمة معينة ، فمثلاً في المثال (2 – 4) قد يهمنا معرفة عدد الموظفين الذين مرتباتهم أقل من 50 ديناراً أسبوعياً أو عدد الموظفين الذين مرتباتهم 50 أو أكثر ، ولتوضيح معلومات كهذه يجب وضع التوزيع التكراري في شكل آخر يطلق عليه التوزيع التكراري المتجمع ويوجد نوعان من التوزيعات التكرارية المتجمعة وهي :

1 – التوزيع التكراري المتجمع الصاعد :

يوضح التوزيع التكراري المتجمع الصاعد عدد المفردات التي قيمتها أقل من الحد الأدنى لأي فئة من فئات جدول التوزيع التكراري العادي ، فمثلاً في مثال (2 – 4) نجد أن عدد الموظفين الذين راتبهم الأسبوعي أقل من 30 ديناراً يساوي صفراً ، أي لا يوجد

موظفون يتقاضون أقل من هذه القيمة ، أما عدد الموظفين الذين راتبهم الأسبوعي أقل من 40 ديناراً يساوي 2، وهو عدد الموظفين التابعين للفئة الأولى وبالنسبة لعدد الموظفين الذين راتبهم أقل من 50 يساوي (2 + 3 = 5)، وهم الموظفون التابعون للفئة الأولى والفئة الثانية ، وهكذا نستمر في عملية التجميع إلى أن نحصل على عدد الموظفين الذين يتقاضون أقل من الحد الأدنى لكل فئة وبالتالي سنحصل على ما يسمى بجدول التكرار المتجمع الصاعد ، وقد سمّي بالصاعد لأن قيمة التكرار في صعود مستمر ، وعند تكوين هذا النوع من الجداول نفترض وجود فئة أكبر من الفئة الأخيرة في جدول التوزيع التكراري العادي ، وبالتالي تكون كل المفردات قيمتها أقل من الحد الأدنى لهذه الفئة ، وسنجد أن أي جدول تكراري متجمع صاعدٍ يبدأ من الصفر وينتهي بالعدد الكلي للمفردات ، والمثال التالي يوضح ذلك :

مثال (2-5) :

كون جدول التوزيع التكراري المتجمع الصاعد للمثال (2 – 4) الخاص بالراتب الأسبوعي لثلاثين موظفاً .

الحل :

الجدول التالي يبين التوزيع التكراري المتجمع الصاعد للموظفين الذين مرتباتهم الأسبوعية أقل من قيمة معينة :

جدول (2-9)

الراتب الأسبوعي بالدينار (الفئة)	التكرار المتجمع الصاعد	
أقل من 30	0	
أقل من 40	2	
أقل من 50	5	(2 + 3)
أقل من 60	17	(2 + 3 + 12)
أقل من 70	24	(2 + 3 + 12 + 7)
أقل من 80	30	(2+3+12+7+6)

2 – التوزيع التكراري المتجمع الهابط :

التوزيع التكراري المتجمع الهابط يوضح عدد المفردات التي قيمتها تساوي أو أكثر من الحد الأدنى لأي فئة من فئات جدول التوزيع التكراري العادي ، فمثلاً في مثال (2 – 4) نجد أن عدد الموظفين الذين راتبهم الأسبوعي 30 أو أكثر هو 30، وهو عبارة عن تكرار الفئة

الأولى مضافاً إليه تكرارات كل الفئات اللاحقة بها ، وفي نفس الوقت هو عبارة عن العدد الكلي للموظفين، لأن كل الموظفين الذين تمثلهم القيم المشاهدة للبيانات المتوفرة لدينا يتقاضون أسبوعياً 30 ديناراً أو أكثر ، أما عدد الموظفين الذين راتبهم الأسبوعي 40 أو أكثر يساوي 28 موظفاً ، وهم عدد الموظفين التابعين للفئة الثانية والفئات اللاحقة بها ، وهكذا نستطيع أن نحسب عدد المفردات التي قيمتها تساوي أو أكثر من الحد الأدنى لكل فئة وذلك بجمع تكرار الفئة التي نتعامل مع حدها الأدنى مع تكرارات كل الفئات اللاحقة بها ، ولذلك سمي بالتكرار المتجمع ، وسمي بالهابط ؛لأن قيمة التكرار في هبوط مستمر ، وهنا كذلك نفترض وجود فئة أكبر من الفئة الأخيرة في جدول التوزيع التكراري العادي ، وبالتالي يكون عدد المفردات التي قيمتها تساوي الحد الأدنى لهذه الفئة أو أكثر منه هو صفر ، وبالتالي سنجد أن أي جدول تكراري متجمع هابط يبدأ من العدد الكلي للتكرارات وينتهي بالصفر ، وذلك كما هو موضح بالمثال التالي :

مثال (2-6) :

كوّن جدول التوزيع التكراري المتجمع الهابط في مثال (2 – 4) .

الحل :

الجدول التالي يبين التوزيع التكراري المتجمع الهابط للموظفين الذين راتبهم الأسبوعي يساوي أو أكثر من قيمة معينة :

جدول (2-10)

الراتب الأسبوعي بالدينار (الفئة)	التكرار المتجمع الهابط	
30 أو أكثر	30	
40 أو أكثر	28	(2-30)
50 أو أكثر	25	(3-2-30)
60 أو أكثر	13	(12-3-2-30)
70 أو أكثر	6	(7-12-3-2-30)
80 أو أكثر	0	(6-7-12-3-2-30)

ونستطيع وضع جدول التكرارالمتجمع الصاعد وجدول التكرارالمتجمع الهابط في جدول واحد، فإذا وضعنا جدول (2 – 9) وجدول (2 – 10) معاً فسنحصل على ما يلي :

جدول (2-11)

الفئة	التكرار المتجمع الصاعد	الفئة	التكرار المتجمع الهابط
أقل من 30	0	30 أو أكثر	30
أقل من 40	2	40 أو أكثر	28
أقل من 50	5	50 أو أكثر	25
أقل من 60	17	60 أو أكثر	13
أقل من 70	24	70 أو أكثر	6
أقل من 80	30	80 أو أكثر	0

ونلاحظ من هذا الجدول أن مجموع التكرارين الصاعد والهابط المتقابلين يساوي دائماً المجموع الكلي للتكرارات ، وبالتالي نستطيع أن نستنتج أحد الجدولين بمعرفة الجدول الآخر .

(2-3) جداول التوزيعات التكرارية النسبية :

نستطيع حساب التكرار النسبي لجميع جداول التوزيعات التكرارية سواء خاصة ببيانات وصفية أو بيانات كمية ،وسواء كانت عادية أو متجمعة ، فبصفة عامة التكرار النّسبي لأي فئة نحصل عليه بقسمة تكرار هذه الفئة على العدد الكلي للمشاهدات (التكرارات) أي أن :

$$\textbf{التكرار النسبي} = \frac{\textbf{تكرار الفئة}}{\textbf{العدد الكلي للمشاهدات (التكرارات)}}$$

ويمكن التعبير عن التكرار النّسبي بكسر عشري أو نسبة مئوية .

مثال (2-7) :

أوجد جدول التكرار النّسبي لجدول التوزيع التكراري في المثال (2 – 1) الذي يمثل المستوى الدراسي لعشرين عاملة في أحد المصانع .

الحل :

في المثال (2 - 1) الخاص بالمستوى التعليمي لعشرين عاملة ، نجد أن :

التكرار النّسبي للفئة الأولى $\frac{5}{20} = 0.25$

أي أن 25 % من العاملات أميات ، وبنفس الطريقة نستطيع أن نحسب التكرار النسبي لجميع الفئات بالجدول، وبالتالي نحصل على الجدول التالي :

جدول (2-12)

المستوى التعليمي (الفئة)	التكرار النسبي
أمي	0.25
ابتدائي	0.40
إعدادي	0.30
ثانوي	0.05
المجموع	1.00

ومن جدول (2 – 12) نفهم أن 25 % من العاملات أميات، 40 % يحملن الشهادة الابتدائية، 30% يحملن الشهادة الإعدادية ، و5% يحملن الشهادة الثانوية ، ويمكن كتابة التكرار النسبي على شكل نسب مئوية بدلاً من كسور عشرية ، وفي هذه الحالة سيكون المجموع 100% .

مثال (2-8) :

أوجد التكرار النسبي لجدول التوزيع التكراري في مثال (2 – 3) الذي يمثل عدد زبائن محل تجاري خلال شهرين .

الحل :

في المثال (2 – 3) الخاص بعدد الزبائن لمحل تجاري ، نجد أن :

التكرار النسبي للفئة الأولى = $\frac{8}{60} = 0.13$

أي أن حوالى 13% من الأيام كان عدد الزبائن فيها يتراوح من 35 إلى 39 .
وبنفس الطريقة نستطيع أن نحسب التكرار النسبي لجميع الفئات بالجدول ، وبالتالي نحصل على الجدول التالي :

جدول (2-13)

عدد الزبائن(الفئة)	التكرار النسبي
39 – 35	0.13
44 – 40	0.20
49 – 45	0.33
54 – 50	0.17
59 – 55	0.10
64 - 60	0.07
المجموع	1.00

مثال (2-9) :

كوّن جدول التكرار المتجمّع الصاعد النّسبي للمثال (2 – 5) ، وجدول التكرار المتجمع الهابط النسبي للمثال (2 – 6) .

الحل :

من جدول (2 – 9) بالمثال (2 – 5) نجد أن :

التكرار المتجمع الصاعد لعدد الموظفين الذين راتبهم أقل من 60 هو 17 ، وبالتالي فإن :

التكرار المتجمع الصاعد النسبي لعدد الموظفين الذين راتبهم أقل من 60

هو $\frac{17}{30} = 0.57$

وبنفس الطريقة نستطيع الحصول على التكرار المتجمع النسبي لجميع فئات جدول التوزيع التكراري المتجمع الصاعد ولجميع فئات جدول التوزيع التكراري المتجمع الهابط الموضح في جدول (2 – 10) بالمثال (2 – 6) ، وعندئذ نحصل على الجدولين التاليين:

جدول (2-14)

الفئة	التكرار المتجمع الصاعد النسبي
أقل من 30	0.00
أقل من 40	0.07
أقل من 50	0.17
أقل من 60	0.57
أقل من 70	0.80
أقل من 80	1.00

جدول (2-15)

الفئة	التكرار المتجمع الهابط النسبي
30 أو أكثر	1.00
40 أو أكثر	0.93
50 أو أكثر	0.83
60 أو أكثر	0.43
70 أو أكثر	0.20
80 أو أكثر	0.00

تمارين (2)

1- البيانات التالية تمثل هوايات 40 طالباً :

قراءة رياضة رياضة قراءة موسيقى رسم رياضة قراءة
رياضة موسيقى رياضة قراءة رسم قراءة رياضة رسم
رسم رياضة موسيقى رياضة رياضة رسم رياضة رياضة
قراءة رسم رياضة قراءة رسم قراءة رسم رياضة
قراءة قراءة رياضة رسم موسيقى رياضة رياضة رياضة

أعرض هذه البيانات في جدول توزيع تكراري ، ثم أحسب التكرار النسبي .

2- البيانات التالية تبين تقديرات 30 طالباً في مادة الرياضيات :

مقبول ، جيد ، ضعيف ، جيد جداً ، ضعيف ، جيد ، مقبول ، ممتاز ، جيد ،
مقبول ، جيد ، ضعيف ، ضعيف ، جيد ، جيد ، جيد جداً ، مقبول ، مقبول ،
ضعيف، مقبول ، جيد ، ضعيف ، مقبول ، جيد جداً ، جيد ،مقبول ، ضعيف،
مقبول ، ضعيف ، ممتاز .

أ – أعرض هذه البيانات في جدول توزيع تكراري .

ب – أوجد جدول التكرار النسبي .

3- من سجلات إحدى الشركات ، وُجد أن عدد الغائبين من العاملين خلال شهر كان كما يلي :

3 ، 2 ، 0 ، 0 ، 3 ، 1 ، 2 ، 0 ، 4 ، 5 ، 3 ، 1 ، 1 ، 0 ، 1
1 ، 2 ، 2 ، 0 ، 0 ، 4 ، 1، 1 ، 1 ، 0 ، 2 ، 0 ، 1 ، 2 ، 1

أعرض هذه البيانات في جدول توزيع تكراري بحيث كل فئة تمثل قيمة واحدة فقط ، ثم أوجد جدول التكرار النسبي لعدد الغائبين .

4- البيانات التالية تبين المبيعات الأسبوعية من السيارات خلال 50 أسبوعاً لأحد معارض السيارات :

2 ، 1 ، 4 ، 0 ، 4 ، 1 ، 2 ، 0 ، 1 ، 3 ، 3 ، 0 ، 1 ، 3 ، 2 ، 2 ، 2 ،1،

1 ' 2 ، 2 ، 0 ، 3 ، 4 ، 0 ، 3 ، 2 ، 1 ، 1 ، 1 ، 1 ، 3 ، 1 ، 3 ، 4 ،2،

0 ، 1 ، 2 ، 2 ، 0 ، 2 ، 3 ، 1 ، 1 ، 0 ، 1 ، 2 ، 1 ، 2 .

أ – أعرض هذه البيانات في جدول توزيع تكراري بحيث كل فئة تمثل قيمة واحدة فقط .

ب – أوجد جدول التكرار النسبي لعدد السيارات المبيعة أسبوعياً .

5- البيانات التالية تبين عدد المدرسين في 25 مدرسة ابتدائية في إحدى المدن :

30 ، 24 ، 23 ، 27 ، 28 ، 31 ، 21 ، 25 ، 18 ، 29 ، 29 ،

20 ، 27، 28 ، 26 ، 26 ، 24 ، 23 ، 27 ، 28 ، 22 ، 26 ،

26 ، 28 ، 30

أ – كوّن جدول توزيع تكراري لهذه البيانات وذلك باستخدام 5 فئات متساوية في الطول .

ب – أوجد جدول التوزيع التكراري النسبي .

ج – أوجد جدول التكرار المتجمع الصاعد لهذه البيانات .

د – من الجدول الذي تحصلت عليه في (ج) أوجد جدول التكرار المتجمع الهابط لهذه البيانات .

6- البيانات التالية تمثل درجات 50 طالباً في مادة الإحصاء :

72	69	18	75	71	89	12	69	65	68
24	81	67	59	69	74	39	66	87	76
63	52	51	60	62	50	59	64	72	69
77	10	60	80	49	64	80	84	30	75
70	70	15	69	72	66	63	20	73	82

أ – كوّن جدول التوزيع التكراري لهذه البيانات ، وذلك باستخدام 8 فئات متساوية الطول بافتراض أن الدرجات تمثل متغيراً منفصلاً .

ب – أوجد جدول التوزيع التكراري النسبي لهذه الدرجات .

ج – أوجد جدول التوزيع التكراري المتجمع الصاعد النسبي لهذه البيانات .

7- البيانات التالية عبارة عن قيمة مبيعات أحد المحلات التجارية خلال خمسة وأربعين يوماً :

128.44	240.81	149.82	153.72	153.80	202.20
143.35	151.20	153.92	166.40	174.52	188.92
151.43	154.03	168.64	176.68	191.63	102.39
156.39	170.15	177.84	111.70	124.27	249.10
170.98	180.91	113.64	129.07	223.18	186.48
184.70	115.60	137.34	143.94	189.28	198.79
115.89	141.81	149.51	152.47	104.21	230.00
184.04	191.62	103.20			

أ – كوّن جدول التوزيع التكراري لهذه البيانات ، وذلك باستخدام 6 فئات متساوية الطول .

ب – كوّن جدول التوزيع التكراري النسبي لهذه البيانات .

ج – أوجد جدول التوزيع التكراري المتجمع الهابط .

د – أوجد جدول التوزيع التكراري المتجمع الهابط النسبي .

8- فيما يلي جدول التكرار المتجمع الهابط لرواتب 68 موظفاً (باعتبار أن الراتب متغير متصل) :

الفئة	التكرار المتجمع الهابط
240 أو أكثر	68
250 أو أكثر	60
260 أو أكثر	50
270 أو أكثر	34
280 أو أكثر	20
290 أو أكثر	8
300 أو أكثر	0

أ – كوّن جدول التوزيع التكراري المتجمع الصاعد النسبي .
ب – كوّن جدول التوزيع التكراري العادي لرواتب الموظفين .
ج – كوّن جدول التوزيع التكراري النسبي لرواتب الموظفين .

9- فيما يلي جدول التوزيع التكراري لعدد الأطفال لدى 55 عائلة :

عدد الأطفال (الفئة)	0 - 3	4 - 7	8 - 11	12 - 15	16 - 19
عدد العائلات (التكرار)	10	24	11	7	3

أ – كـوّن جدول التوزيع التكراري النسبي .
ب – كـوّن جدول التوزيع التكراري المتجمع الصاعد النسبي .
ج – كـوّن جدول التوزيع التكراري المتجمع الهابط النسبي .

الفصل الثالث
العرض البياني للبيانات الإحصائية

شرحنا في الفصل السابق كيفية عرض عدد كبير من البيانات في صورة جـداول تكرارية، وذلك لتبسيط هذه البيانات وتسهيل فهمها ، ولكن الجداول التكرارية قد لا تكفي أحياناً للتوضيح وخاصة أن بعض الناس يجدون صعوبة كبيرة في إدراك مدلولات الأرقام التي تعرض عليهم في جداول ، ولزيادة الإيضاح يحاول الإحصائي عرض البيانات بطريقة أخرى وهي الرسومات البيانية ، وتوجد عدة طرق لتمثيل البيانات الإحصائية بيانياً ، وسندرس أهمها والمتمثلة فيما يلي :

- **القطاعات الدائرية .**
- **الأعمدة البيانية .**
- **المدرج التكراري .**
- **المضلع التكراري .**
- **المنحنى التكراري .**
- **المنحنى التكراري المتجمع الصاعد والهابط.**

(1-3) القطاعات الدائرية :

في هذه الطريقة نمثل العدد الكلي للمفردات (المجموع الكلي للتكرارات) بدائرة كاملة ونمثل عدد المفردات التابعة لكل فئة (تكرار الفئة) بقطاع من الدائرة بحيث تتناسب مساحات القطاعات مع تكرارات الفئات التي تمثلها ، ولما كانت مساحة القطاع تتناسب مع زاويته المركزية في الدائرة الواحدة ، فما علينا إلا تقسيم الزاوية المركزية للدائرة والتي تساوي 360° إلى زوايا تتناسب مع تكرارات فئات التوزيع ، ولتحقيق ذلك تحسب زاوية القطاع الذي يمثل كل فئة ، كما يلي :

$$\text{زاوية القطاع الذي يمثل الفئة} = \frac{\text{تكرار الفئة}}{\text{المجموع الكلي للتكرارات}} \times 360^\circ$$

وعلى أساس قياسات الزوايا نقسم مساحة الدائرة إلى قطاعات، كل قطاع يمثل فئة ويمكن

تظليل هذه القطاعات أو تلوينها للتمييز بينها ويكتب داخل أو بجوار كل قطاع اسم الفئة التي يمثلها، وعادة تستخدم طريقة القطاعات الدائرية في حالة البيانات الوصفية.

مثال (3-1) :

يبين الجدول التالي قيمة الواردات حسب الموانئ لسنة 1982 ، وفقاً لنشرات اتجاهات التجارة الخارجية التي تعدها مصلحة الإحصاء والتعداد :

الموانئ	القيمة بالدنانير الليبية (لأقرب مليون)
ميناء طرابلس	1530
ميناء بنغازي	453
موانئ أخرى	142
المجموع	2125

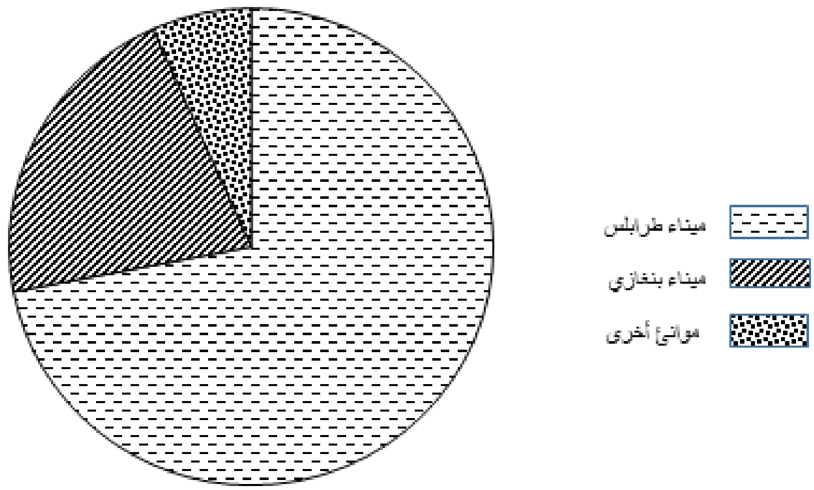
في هذا المثال ، الفئات هي عبارة عن الموانئ ، والتكرار هو قيمة الواردات ، ولتمثيل هذه البيانات بقطاعات دائرية سنحصل على زوايا القطاعات كما يلي :

زاوية القطاع الذي يمثل ميناء طرابلس = $\frac{1530}{2125} \times 360 = 259.2^\circ$

زاوية القطاع الذي يمثل ميناء بنغازي = $\frac{453}{2125} \times 360 = 76.7^\circ$

زاوية القطاع الذي يمثل موانئ أخرى = $\frac{142}{2125} \times 360 = 24.1^\circ$

نرسم دائرة بنصف قطر مناسب ، ونقسم زاويتها المركزية التي تساوي 360° إلى الأجزاء السابقة ، فنحصل على الشكل (3 – 1):



شكل (3-1)

(3-2) الأعمدة البيانية :

الأعمدة البيانية هي مستطيلات ذات قواعد متساوية وعمودية على المحور الأفقي ، بحيث إن كل مستطيل يمثل فئة من فئات جدول التوزيع التكراري ، ويكون ارتفاع كل مستطيل مساوياً لتكرار الفئة التي يمثلها ، مع مراعاة أن تكون المسافة بين هذه المستطيلات متساوية حتى يكون الرسم أكثر وضوحاً .

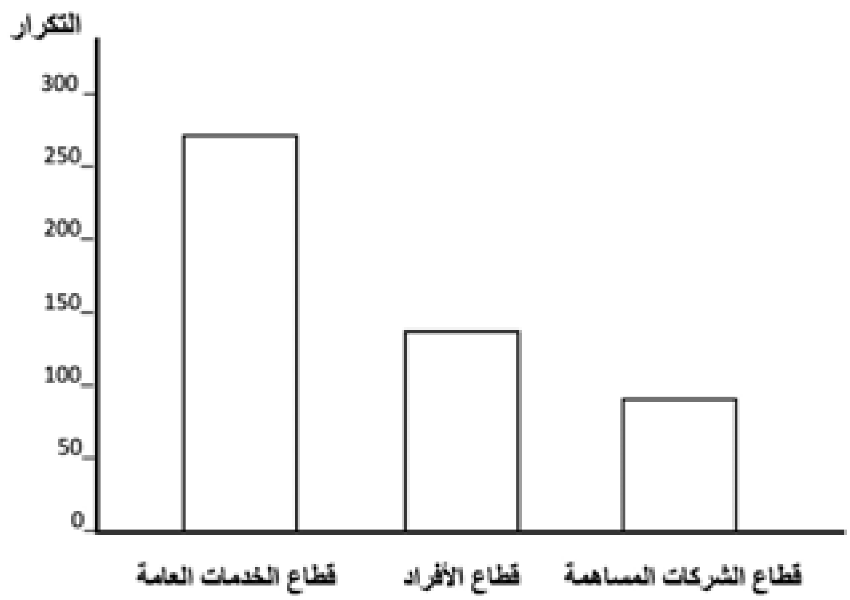
وعادة تستخدم طريقة الأعمدة البيانية لتمثيل الجداول التكرارية الخاصة ببيانات وصفية أو بيانات كمية خاصة بمتغيرات منفصلة فئاتها تمثل قيمة واحدة فقط من القيم التي يأخذها المتغير المنفصل محل الدراسة .

مثال (3-2) :

يبين الجدول التالي توزيع القوى العاملة الليبية وغير الليبية ، حسب القطاعات ، وذلك طبقاً لحصر القوى العاملة لعام 1980م .

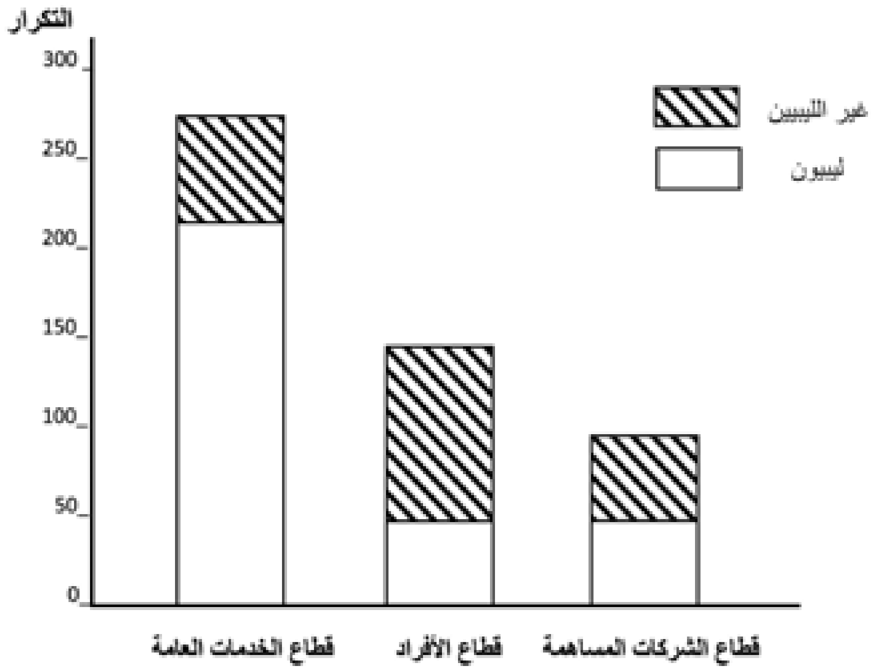
القطاع	عدد الليبيين (لأقرب ألف)	عدد غير الليبيين (لأقرب ألف)	العدد الكلي
قطاع الخدمات العامة	216	55	271
قطاع الأفراد	39	97	136
قطاع الشركات المساهمة	47	47	94

اذا أردنا تمثيل الفئات المختلفة (القطاعات) دون التمييز بين الليبيين وغير الليبيين فسنمثل كل فئة (قطاع) بمستطيل بحيث تكون قواعد المستطيلات متساوية (وفي هذا المثال جعلنا طول كل قاعدة مستطيل مساوية 1 سم) ، وارتفاع كل مستطيل يساوي المجموع الكلي للقوى العاملة التابعة للقطاع الذي يمثله هذا المستطيل ، وذلك كما هو موضح في شكل (3 – 2) .



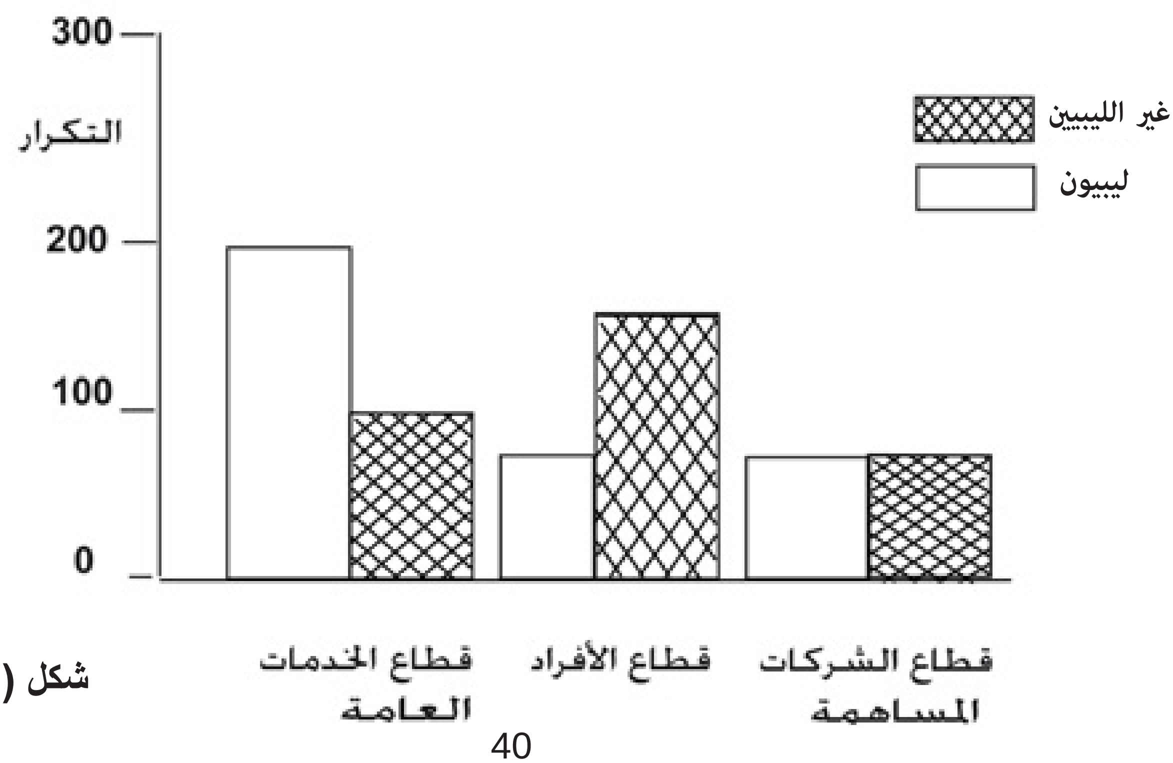
شكل (3-2)

ونستطيع باستخدام الأعمدة البيانية مقارنة مكونات ظاهرة معينة بأجمالها ، فباستخدام بيانات المثال نستطيع مقارنة القوى العاملة غير الليبية في كل قطاع بإجمالي القوى العاملة في هذا القطاع وذلك بتقسيم كل مستطيل يمثل قطاعاً إلى جزأين عن طريق خط مستقيم يوازى القاعدة ، بحيث الجزء الأول يمثل عدد القوى العاملة الليبية والجزء الثاني يمثل عدد القوى العاملة غير الليبية وذلك كما هو موضح في شكل (3-3) ، ويطلق على طريقة العرض بهذه الصورة طريقة الأعمدة المجزأة .



شكل (3-3)

كذلك نستطيع باستخدام الأعمدة البيانية مقارنة ظاهرتين أو أكثر وذلك برسم أعمدة متلاصقة للظواهر المراد مقارنتها ، فمثلاً باستخدام بيانات المثال السابق الخاص بالقوى العاملة ، نستطيع مقارنة القوى العاملة الليبية بالقوى العاملة غير الليبية في كل قطاع من القطاعات وذلك برسم مستطيلين متلاصقين لكل قطاع أحدهما يمثل عدد الليبيين والآخر يمثل غير الليبيين ، بحيث تكون قواعد المستطيلات متساوية والمسافة بين ازواج المستطيلات متساوية ، وذلك كما هو موضح في شكل (3 – 4) ، ويطلق على هذه الطريقة طريقة الأعمدة المتلاصقة .



شكل (3-4)

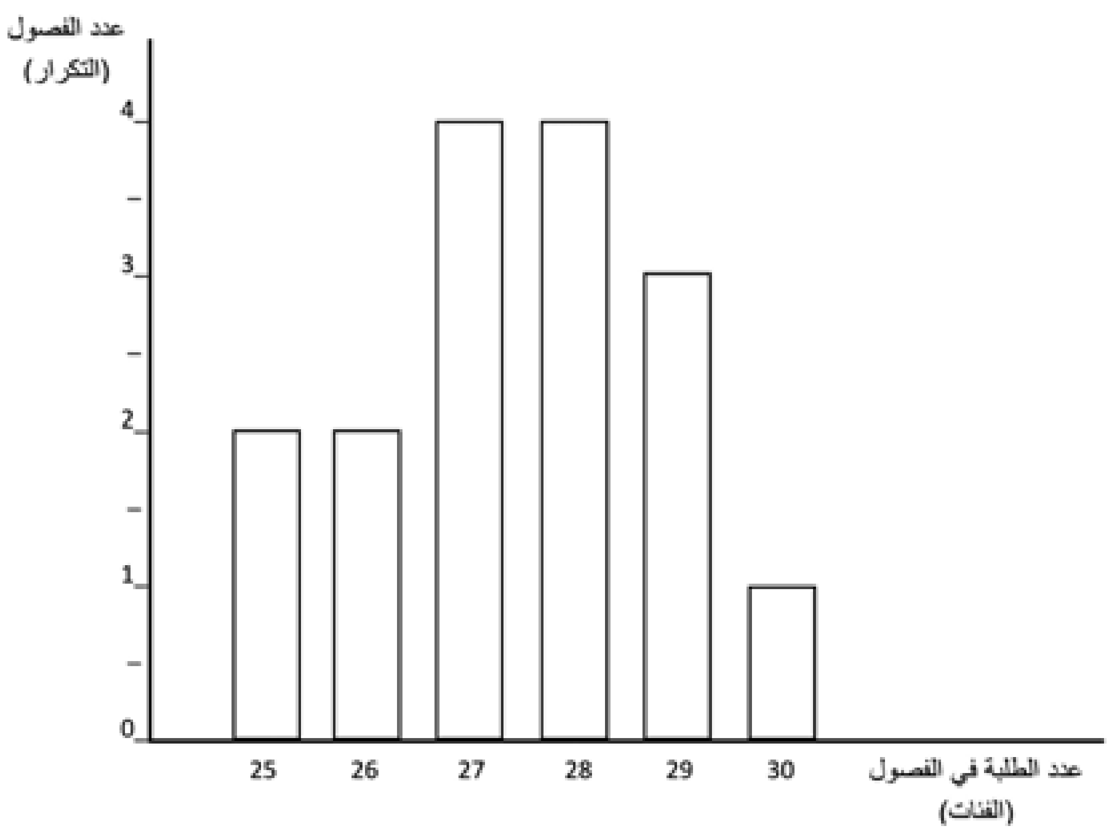
وكما ذكرنا سابقاً أن الأعمدة البيانية تُستعمل كذلك لتمثيل الجداول التكرارية والخاصة ببيانات منفصلة عندما تمثل كل فئة من فئات الجدول قيمة واحدة للمتغير المنفصل محل الدراسة ، والمثال التالي يوضح ذلك :

مثال (3-3) :

مدرسة إعدادية بها 16 فصلاً دراسياً، وجدول التوزيع التكراري التالي يوضح توزيع الفصول الدراسية حسب أعداد الطلبة بمدرسة بها (16) فصلاً :

عدد الطلبة (الفئات)	25	26	27	28	29	30	المجموع
عدد الفصول (التكرار)	2	2	4	4	3	1	16

ونفهم من هذا الجدول أنه يوجد فصلان بالمدرسة كل واحد منهما يشمل 25 طالباً ، وفصلان يشمل كل واحد منهما 26 طالباً ، وأربعة فصول يشمل كل واحد منها 27 طالباً ، وهكذا ... ونلاحظ أن هذه البيانات تمثل متغيراً منفصلاً، وكل فئة في الجدول تحتوي على قيمة واحدة ، وبالتالي نستطيع أن نمثل هذه البيانات بأعمدة بيانية كما في الشكل (3 – 5) .



شكل (3-5)

(3-3) المدرج التكراري :

المدرج التكراري هو مجموعة من المستطيلات المتلاصقة ، كل مستطيل منها يمثل فئة من فئات التوزيع التكراري ، بحيث تكون مساحات المستطيلات متناسبة مع تكرارات الفئات التي تمثلها.

وهذه الطريقة هي الأكثر شيوعاً،وتستخدم لتمثيل الجداول التكرارية الخاصة بمتغيرات متصلة وكذلك الخاصة بمتغيرات منفصلة عندما تمثل الفئة أكثر من قيمة واحدة . ولرسم المدرج التكراري نتبع الخطوات التالية :

- نرسم المحورين ، المحور الأفقي نمثل عليه الفئات، ويجب أن يقسم تقسيماً ملائماً بحيث يسمح بظهور جميع الفئات الموجودة في جدول التوزيع التكراري ، والمحور الرأسي يمثل التكرارات، ويجب تقسيمه تقسيماً ملائماً بحيث يسمح بظهور أكبر التكرارات .
- نرسم فوق كل فئة مستطيلاً ارتفاعه مساوٍ لتكرار الفئة التي يمثلها مع عدم ترك مسافات بين المستطيلات ، وبعد رسم جميع المستطيلات الممثلة لجميع الفئات نحصل على ما يسمى بالمدرج التكراري .

وعندما يكون جدول التوزيع التكراري منتظماً أي تكون فئاته متساوية الطول، تكون مساحات المستطيلات متناسبة مع تكرارات الفئات التي تمثلها ، أما إذا كان الجدول غير منتظم أي فئاته غير متساوية الطول فلكي تكون مساحات المستطيلات متناسبة مع التكرارات يجب أن نعدل التكرار قبل القيام بالرسم ، حيث التكرار المعدل للفئة التي طولها يختلف عن الطول السائد في الجدول أي الطول القياسي للفئات يحسب كما يلي :

$$\text{التكرار المعدل للفئة} = \text{التكرار الأصلي} \div \frac{\text{طول الفئة}}{\text{الطول القياسي للفئات}}$$

مثال (3-4) :

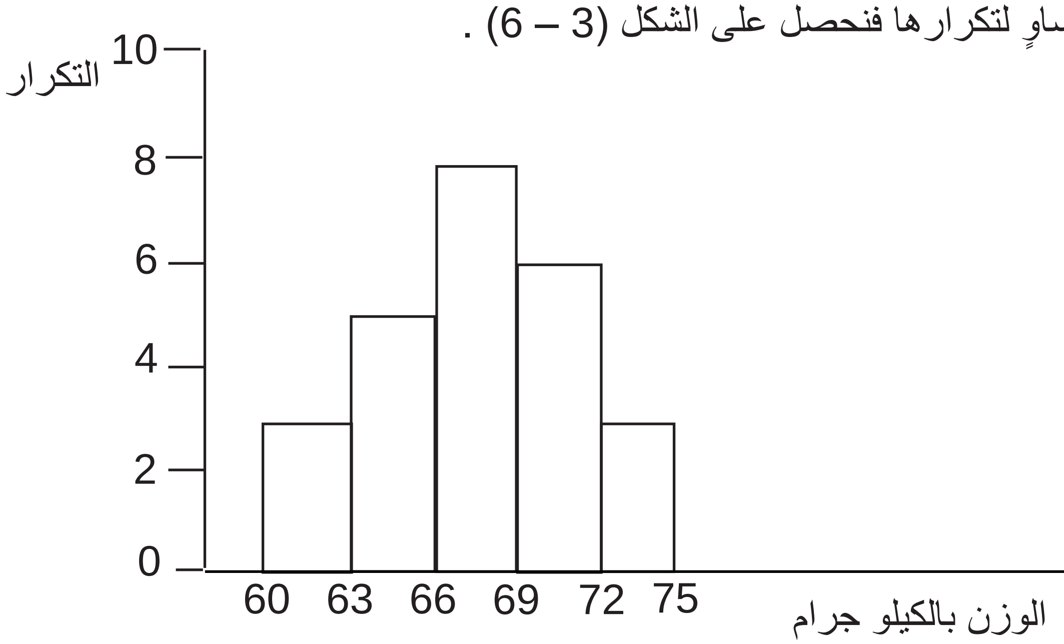
الجدول التالي يمثل التوزيع التكراري لأوزان 25 طالباً بالكيلوجرامات ، مثّل هذا التوزيع باستخدام المدرج التكراري .

الوزن	60 إلى أقل من 63	63 إلى أقل من 66	66 إلى أقل من 69	69 إلى أقل من 72	72 إلى أقل من 75
عدد الطلبة	3	5	8	6	3

الحلّ :

نمثل حدود الفئات (الوزن) على المحور الأفقي ، ونمثل التكرارات (عدد الطلبة) على المحور الرأسي ، وذلك باستخدام مقياس رسم مناسب .

بما أن فئات الجدول متساوية الطول (الجدول منتظم) فنعبر عن كل فئة بمستطيل ارتفاعه مساوٍ لتكرارها فنحصل على الشكل (3 – 6) .



شكل (3-6)

مثال (3-5) :

مثّـل جدول التوزيع التكراري التالي بيانياً مستخدماً المدرج التكراري :

الفئة	**التكرار**
10 إلى أقل من 30	12
30 إلى أقل من 40	20
40 إلى أقل من 50	33
50 إلى أقل من 60	17
60 إلى أقل من 75	6
المجموع	88

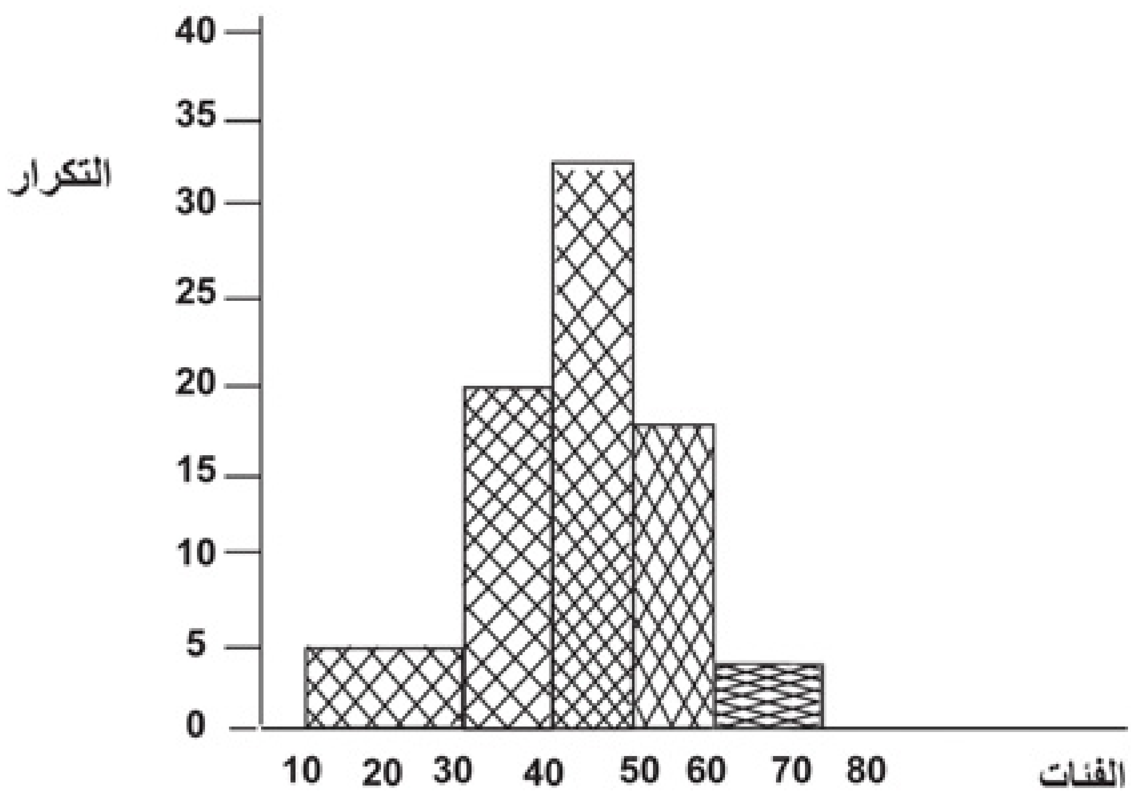
الحلّ :

نلاحظ أن هذا الجدول غير منتظم، أي أنّ فئاته غير متساوية الطول ، فيه الفئة الأولى والفئة الأخيرة يختلف طولهما عن بقية الفئات ، فالفئات الأخرى كلها متساوية الطول وطول كل منها يساوي 10، وهذا الطول هو الذي نعتبره الطول القياسي ، ويجب هنا قبل القيام برسم المدرج أن نحسب التكرار المعدل للفئتين الأولى والأخيرة ، وذلك كما يلي :

التكرار المعدل للفئة الأولى $= 12 \div \left(\frac{20}{10} \right) = 6$

التكرار المعدل للفئة الأخيرة $= 6 \div \left(\frac{15}{10} \right) = 4$

فنمثل الفئة الأولى بمستطيل ارتفاعه 6 والفئة الأخيرة بمستطيل ارتفاعه 4 ، وبالتالي ستكون مساحات المستطيلات متناسبة مع تكرارات الفئات التي تمثلها ، وشكل (3 – 7) يوضح المدرج التكراري المطلوب .



شكل (3-7)

وكما أشرنا سابقاً أن المدرج التكراري يستخدم كذلك في حالة التوزيعات التكرارية الخاصة بمتغيرات منفصلة ، وحيث إنّ المدرج عبارة عن مستطيلات متلاصقة ، لذلك يجب أن نتخلص من الفجوات الموجودة بين فئات جدول التوزيع التكراري الخاص ببيانات منفصلة قبل تمثيله بمدرج تكراري ، ويتم ذلك بتقسيم المسافة بين أي فئتين إلى قسمين متساويين ، قسم يضم إلى الفئة السابقة وقسم يضم إلى الفئة اللاحقة ، ويطلق على الحدود الجديدة بين الفئات الحدود الحقيقية ، وهي التي تستخدم عند رسم المدرج التكراري .

مثال (3-6) :

الجدول التكراري التالي يوضح توزيع درجات 60 طالباً في امتحان لمادة الإحصاء ، مثّل هذا الجدول باستخدام المدرج التكراري .

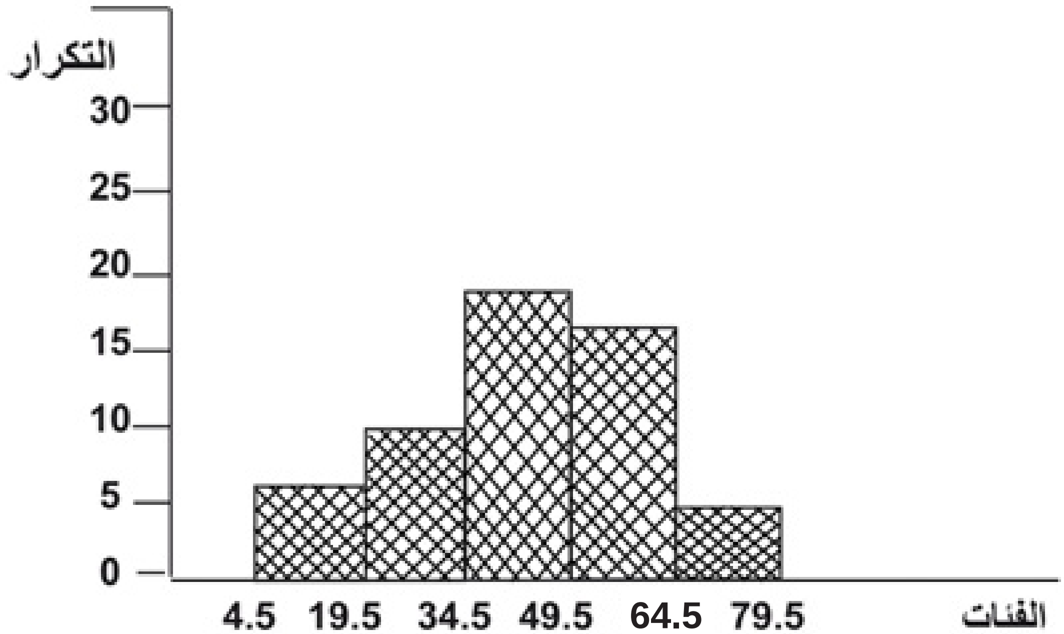
الدرجات	5-19	20-34	35-49	50-64	65-79	المجموع
عدد الطلبة	7	10	20	18	5	60

الحـلّ :

حيث إن الجدول الذي نرغب في تمثيله هو جدول خاص ببيانات منفصلة ، لذلك يجب التخلص من الفجوات التي تفصل فئات الجدول وذلك بإيجاد الحدود الحقيقة لكل فئة .
وللتخلص من المسافة بين الفئة الأولى والفئة الثانية ، نضم نصف المسافة إلى الفئة الأولى والنصف الآخر إلى الفئة الثانية ، وعندئذ ستنتهي الفئة الأولى عند القيمة 19.5 بدلاً من القيمة 19 ، وتبدأ الفئة الثانية من القيمة 19.5 بدلاً من القيمة 20 وبالتالي نكون قد تخلصنا من الفجوة الموجودة بين الفئة الأولى والفئة الثانية.
وبنفس الطريقة نستطيع التخلص من كل فجوات الجدول ونحصل على الحدود الحقيقة التي سنستخدمها عند رسم المدرج التكراري والموضحة في جدول (3-1) ، فمن هذا الجدول نستطيع تمثيل التوزيع التكراري للدرجات بالمدرج التكراري الموضح في شكل (3-8) .

جدول (3-1)

الدرجات	4.5-19.5	19.5-34.5	34.5-49.5	49.5-64.5	64.5-79.5
عدد الطلبة	7	10	20	18	5



شكل (3-8)

(3-4) المضلّع التكراري :

المضلع التكراري هو خط منكسر يصل بين نقط تمثل كل نقطة منها فئة من فئات التوزيع ، حيث تكون إحداثي أي نقطة (x ، y) كما يلي :

x : مركز الفئة التي تمثلها النقطة .

y : تكرار هذه الفئة .

ويجب أن يغلق المضلع التكراري من طرفيه على المحور الأفقي ، ويتم ذلك بإضافة فئة قبل الفئة الأولى وفئة بعد الفئة الأخيرة وبالطبع سيكون تكرار كل منهما يساوي صفراً وبالتالي ستقع النقطتان الخاصتان بهاتين الفئتين على المحور الأفقي فتؤدي إلى غلق المضلع التكراري.

وعند رسم المضلع التكراري نقوم بالخطوات التالية :

- نرسم محورين متعامدين ، ونقوم بتدريج المحور الأفقي بطريقة تناسب فئات التوزيع وتدريج المحور الرأسي بطريقة تناسب التكرارات
- نضيف للجدول فئة سابقة للفئة الأولى ،وفئة لاحقة للفئة الأخيرة حيث تكرار كل منهما يساوي صفراً ، ثم نقوم بحساب مراكز كل الفئات .
- نمثل كل فئة بنقطة على الرسم البياني ، حيث الإحداثي السيني للنقطة هو مركز الفئة والإحداثي الصادي هو تكرار الفئة .
- نصل كل نقطتين متتاليتين بخط مستقيم فنحصل على خط منكسر، هذا الخط المنكسر يطلق عليه المضلع التكراري .

مثال (3-7) :

مثّـل التوزيع التكراري التالي بمضلع تكراري :

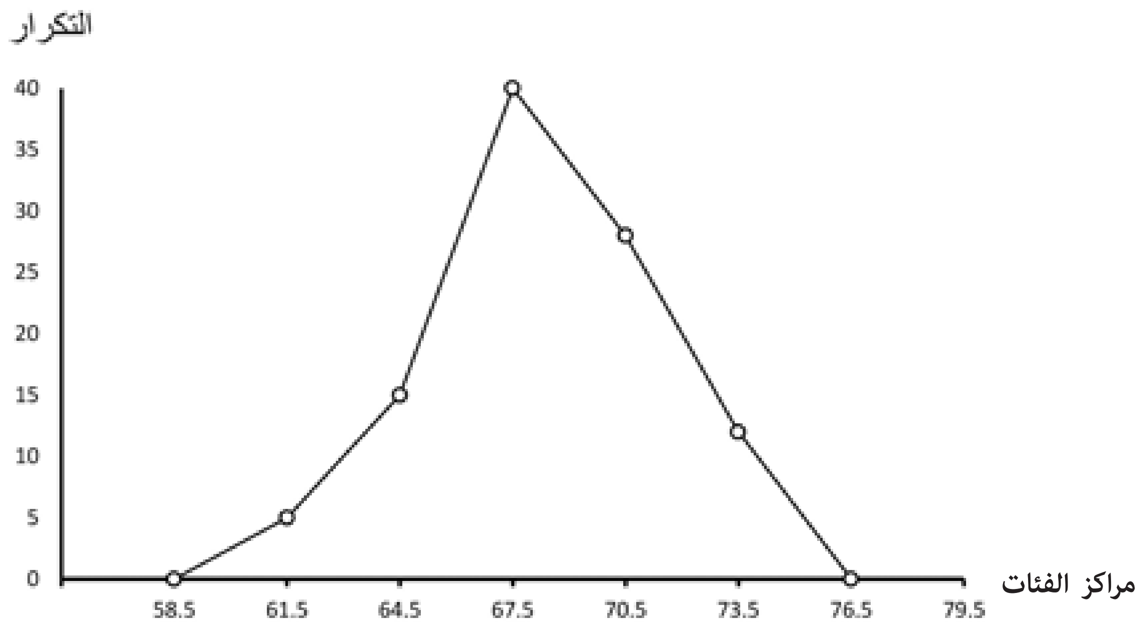
الفئة	التكرار
60 إلى أقل من 63	5
63 إلى أقل من 66	15
66 إلى أقل من 69	40
69 إلى أقل من 72	28
72 إلى أقل من 75	12

الحل :

لرسم المضلع التكراري نفترض وجود فئة قبل الفئة الأولى وفئة بعد الفئة الأخيرة ، وتكرار كلٍّ منها يساوي صفراً ونحدد مراكز الفئات ثم نحدد نقط المضلع ، وذلك كما هو موضح في جدول (3 – 2)، وبرسم هذه النقط على الرسم البياني والتوصيل بينها بخطوط مستقيمة نحصل على المضلع التكراري الموضح في الشكل (3 – 9) .

جدول (3-2)

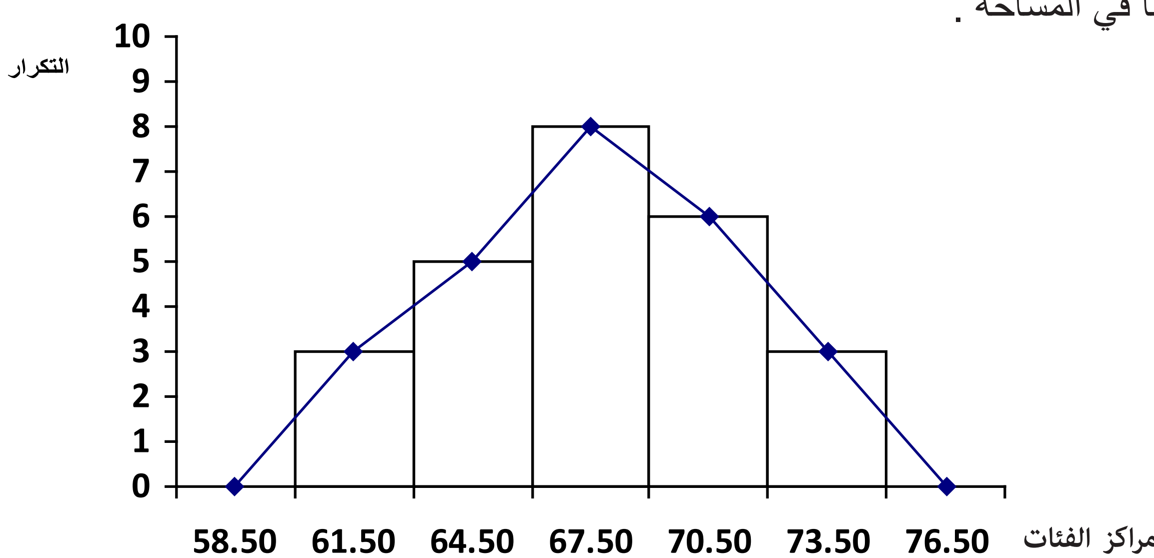
الفئة	مركز الفئة	التكرار	نقط المضلع
57 إلى أقل من 60	58.5	0	(0 ، 58.5)
60 إلى أقل من 63	61.5	5	(5 ، 61.5)
63 إلى أقل من 66	64.5	15	(15 ، 64.5)
66 إلى أقل من 69	67.5	40	(40 ، 67.5
69 إلى أقل من 72	70.5	28	(28 ، 70.5)
72 إلى أقل من 75	73.5	12	(12 ، 73.5)
75 إلى أقل من 78	76.5	0	(0 ، 76.5)



شكل (3-9)

عندما يكون جدول التوزيع التكراري غير منتظم ، أي يحتوي على فئات طولها يختلف عن الفئات الأخرى ، فيجب قبل رسم المضلع التكراري إجراء تعديل للتكرارات مثل الذي أجريناه في حالة المدرج التكراري .

ونستطيع رسم المضلع التكراري اعتماداً على المدرج التكراري بتوصيل منتصفات القواعد العليا للمستطيلات المكونة للمدرج ، وذلك كما هو موضح في شكل (3 – 10) ، ومن هذا الشكل نستنتج أن المساحة الكلية تحت المضلع التكراري مساوية تماماً للمساحة الكلية تحت المدرج التكراري ؛ لأن المضلع يحذف من كل مستطيل مثلثاً ويضيف له مثلثاً آخر مساوياً له تماماً في المساحة .



شكل (3-10)

(5-3) المنحنى التكراري :

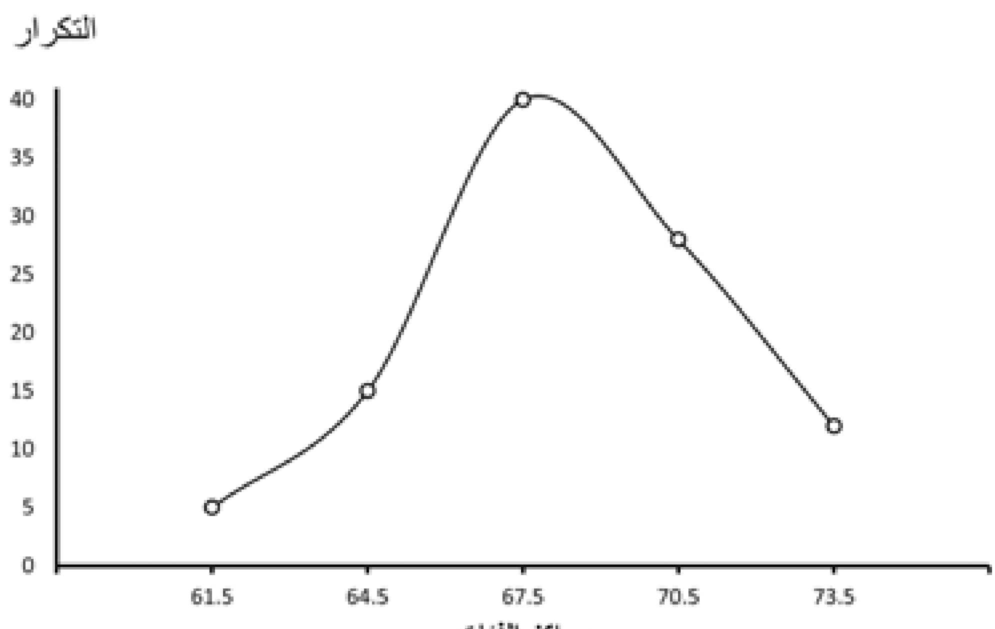
يرسم المنحنى التكراري باتّباع الخطوات نفسها التي أجريناها في حالة المضلع التكراري والفرق بين الحالتين هو في طريقة توصيل النقط التي تمثل الفئات ، ففي حالة المنحنى التكراري نصل بين النقط بمنحنى ممهد مستمر بحيث يمر بأكبر عدد ممكن من النقط ويمر من خلال الباقي بتوازن ، كذلك في حالة المنحنى التكراري لا نستعين بالفئتين: السابقة للأولى واللاحقة للأخيرة . وبسبب التقريب في الرسم وعدم التقيد بالنقط تماماً ، سنجد أن المساحة الواقعة تحت المنحنى والمحصورة بينه وبين المحور الأفقي غير مساوية للمساحة تحت المضلع وبالتالي لا تساوي المساحة تحت المدرج التكراري ، ويكون المنحنى التكراري ممهداً بطريقة جيدة عندما تكون النقط التي تمثل الفئات قريبة من بعضها أي كلما زاد عدد الفئات وقل طولها .

مثال (3-8) :

مثّل الجدول التكراري المذكور في المثال (3 – 7) بيانياً باستخدام المنحنى التكراري .

الحلّ :

- نقوم بتعيين النقط التي تمثل الفئات كما في حالة المضلع التكراري .
- نصل بين النقط بمنحنى ، وذلك كما هو موضح بالشكل (3 – 11) .



شكل (3-11)

(6-3) المنحنى التكراري المتجمع (الصاعد أوالهابط) :

لتمثيل جدول التوزيع التكراري المتجمع الصاعد أو الهابط نجري الخطوات التالية :

- نرسم محورين متعامدين ، نخصص المحور الأفقي كالمعتاد للفئات والمحور الرأسي للتكرارات مع مراعاة اختيار مقياس للرسم يتسع للتكرار الكلي وذلك لأن أكبر عدد في عمود التكرار المتجمع هو التكرار الكلي .
- نمثل كل فئة في جدول التكرار المتجمع بنقطة في الرسم البياني ، بحيث يكون الإحداثي الأفقي لهذه النقطة يساوي الحد الأدنى للفئة، والإحداثي الرأسي لها هو التكرار المتجمع الخاص بهذه الفئة .
- بعد تحديد النقط على الرسم نصل بينها بخط منحني منتظم فنحصل على المنحنى المتجمع الصاعد في حالة التوزيع التكراري المتجمع الصاعد ،ونحصل على المنحنى الهابط في حالة التوزيع التكراري المتجمع الهابط ، والمثال التالي يوضح ذلك :

مثال (3-9) :

ارسم المنحنى التكراري المتجمع الصاعد والمنحنى التكراري المتجمع الهابط للتوزيع التكراري التالي الذي يمثل أوزان 100 طالب بالكيلو جرامات .

الوزن (بالكيلو جرامات)	عدد الطلبة
60 إلى أقل من 63	5
63 إلى أقل من 66	15
66 إلى أقل من 69	40
69 إلى أقل من 72	28
72 إلى أقل من 75	12
المجموع	100

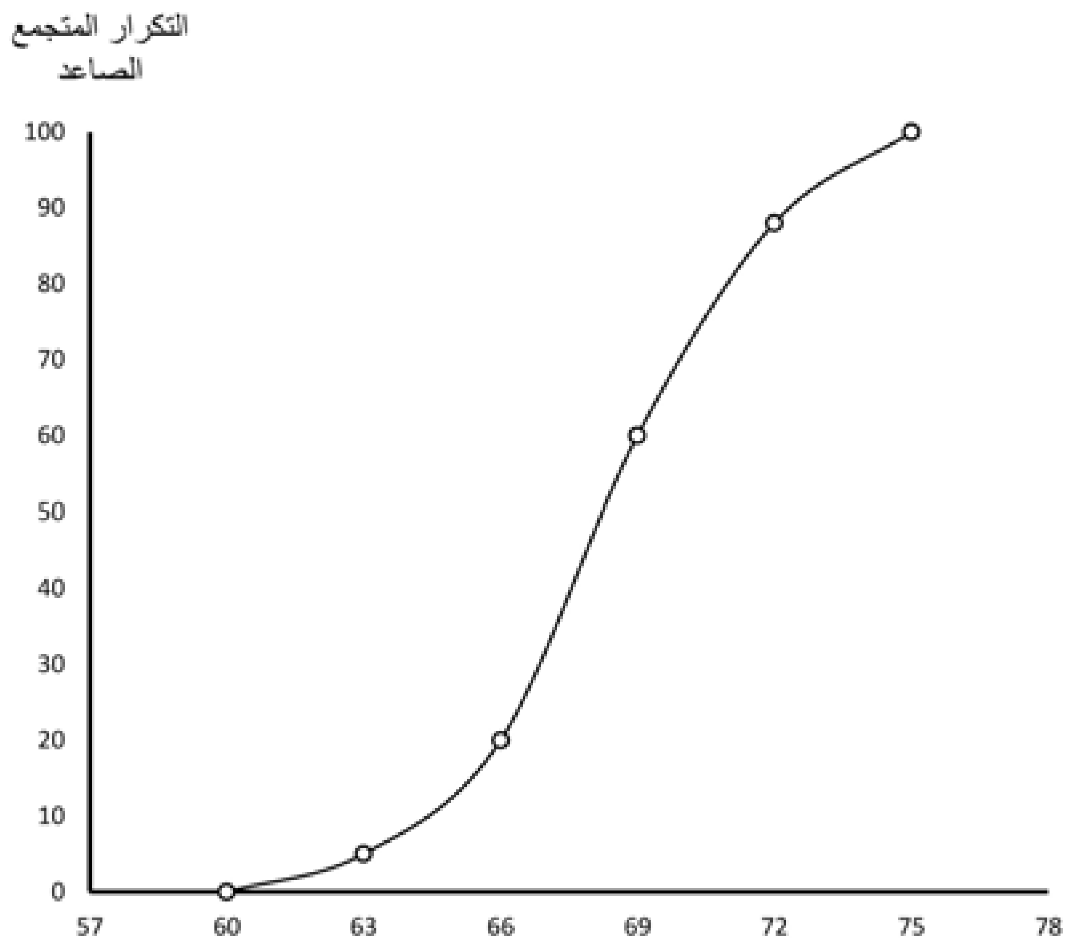
الحلّ :

نكون جدول التكرار المتجمع الصاعد والموضح في جدول (3 – 3) ، ثم نقوم برصد النقط على الرسم ، حيث إحداثي النقطة الأولى (60 ، 0)، وإحداثي النقطة الثانية (63 ، 5) وهكذا .... ونصل بين النقط بمنحنى فنحصل على المنحنى التكراري المتجمع الصاعد الموضح في شكل (3 – 12) .

ونكوّن جدول التكرار المتجمع الهابط الموضح في جدول (3 – 4) ، ثم نقوم برصد النقط على الرسم ، حيث إحداثي النقطة الأولى (60 ، 100)، وإحداثي النقطة الثانية (63 ، 95)، وهكذا ..... ونصل بين النقط بمنحنى فنحصل على المنحنى التكراري المتجمع الهابط الموضح في شكل (3 – 13)

جدول (3-3)

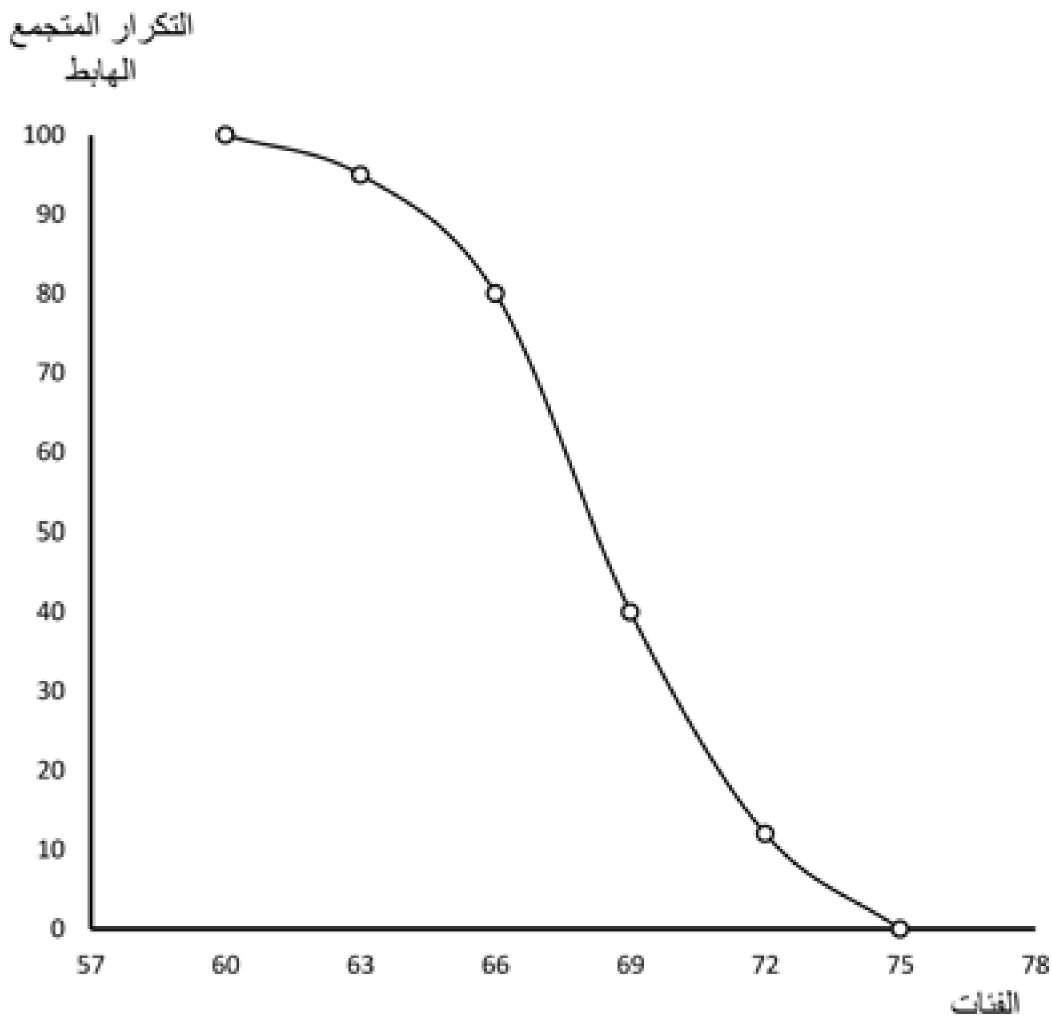
الفئة	التكرار
أقل من 60	0
أقل من 63	5
أقل من 66	20
أقل من 69	60
أقل من 72	88
أقل من 75	100



شكل (3-12)

جدول (3-4)

الفئة	التكرار المتجمع الهابط
60 أو أكثر	100
63 أو أكثر	95
66 أو أكثر	80
69 أو أكثر	40
72 أو أكثر	12
75 أو أكثر	0



شكل (3-13)

مع أن المنحنيين المتجمعين يختلفان عن بعضهما من ناحية الشكل إلا أنهما يخدمان نفس الغرض ، فباستخدام المنحنى التكراري المتجمع الصاعد نستطيع معرفة عدد المفردات التي أقل من قيمة معينة ، وبطرح هذا العدد من العدد الكلي للمفردات نحصل على عدد المفردات

التي أكبر من أو تساوي هذه القيمة وهو العدد الذي نحصل عليه عند استخدام المنحنى التكراري المتجمع الهابط .

وإذا رسمنا هذين المنحنيين في نفس الشكل سنجد أنهما سيلتقيان عند نقطة واحدة فقط وهي القيمة التي يكون عدد المفردات التي أقل منها يساوي عدد المفردات التي أكبر منها أي عندما يتساوى التكرار المتجمع الصاعد مع التكرار المتجمع الهابط ويساوي كل منهما 50% من العدد الكلي للمفردات ، وبالتالي فنقطة التقاطع تعطي لنا القيمة الوسطى في البيانات أو ما يسمى بالوسيط .

مثال (3-9) :

الجدول التالي يوضح أوزان 100 قطعة منتجة بالجرامات :

الوزن	عدد القطع
20 إلى أقل من 25	6
25 إلى أقل من 30	14
30 إلى أقل من 35	43
35 إلى أقل من 40	32
40 إلى أقل من 45	5

ارسم المنحنى التكراري المتجمع الصاعد للأوزان ، ومن الرسم أوجد :

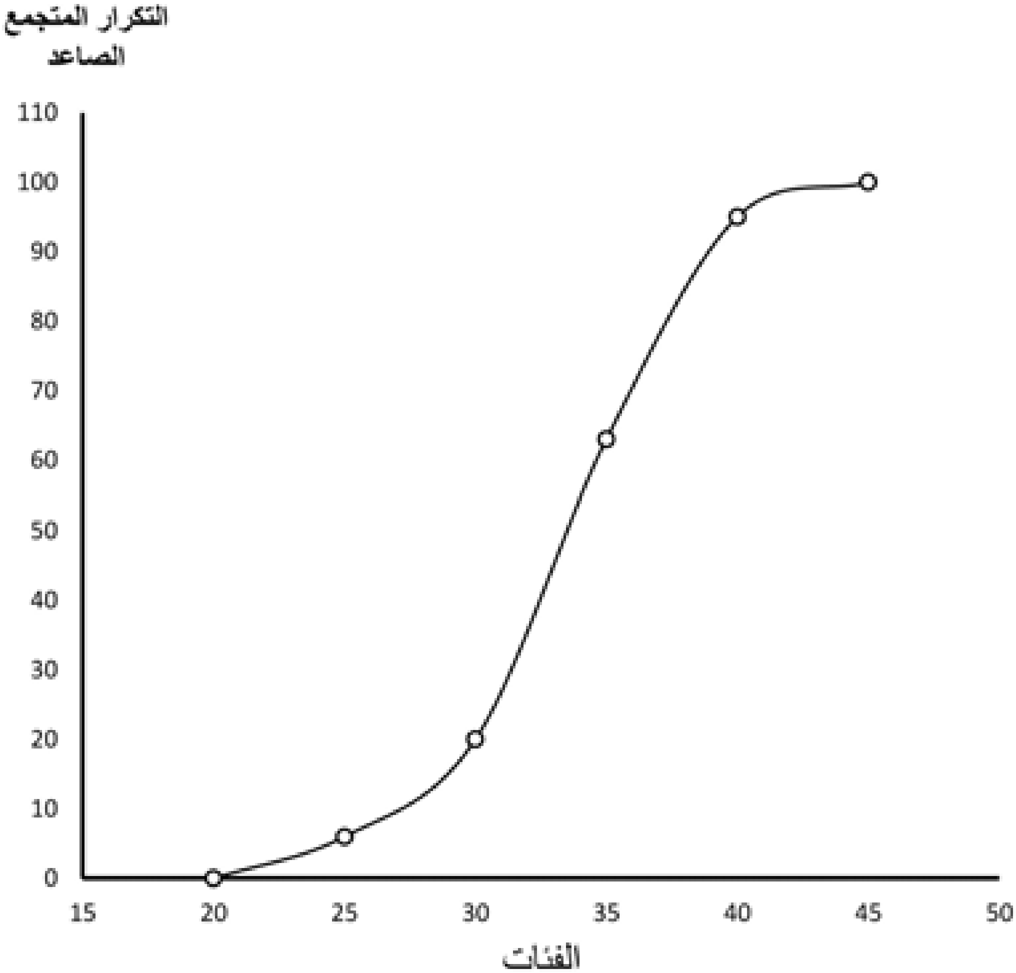
- عدد القطع التي أوزانها من 24 إلى أقل من 32 .

الحلّ :

نكوّن جدول التكرارالمتجمع الصاعد للأوزان والموضح في جدول (3 – 5) ، ثم نقوم برصد النقط التي تمثل الفئات ونصل بينها بمنحنى فنحصل على المنحنى التكراري المتجمع الصاعد الموضح في شكل (3 – 14) .

جدول (3-5)

الفئة	التكرار
أقل من 20	0
أقل من 25	6
أقل من 30	20
أقل من 35	63
أقل من 40	95
أقل من 45	100



شكل (3-14)

- نستطيع الحصول على عدد القطع التي أوزانها تتراوح من 24 جراماً إلى أقل من 32 جراماً من المنحنى التكراري المتجمع الصاعد ، وذلك برسم خط عمودي على المحور الأفقي عند الوزن 24 ونمده إلى أن يلتقي بالمنحنى التكراري المتجمع الصاعد ثم نرسم من نقطة الالتقاء خطاً أفقياً عمودياً على المحور الرأسي ، والنقطة التي يلتقي فيها هذا الخط مع المحور الرأسي تمثل عدد القطع التي أوزانها أقل من 24 ، ومن شكل (3 – 14) نجد أن هذه النقطة هي القيمة 5 تقربياً ، ويعنى ذلك أنه توجد 5 قطع أوزانها أقل من 24 جراماً . وبنفس الطريقة نجد أن هناك 35 قطعة تقربياً أوزانها أقل من 32 جراماً ، وبالتالي فإن عدد القطع التي أوزانها تتراوح من 24 إلى أقل من 32 يساوي عدد القطع التي أوزانها أقل من 32 مطروحاً منه عدد القطع التي أوزانها أقل من 24 . أي أن :

عدد القطع التي تتراوح أوزانها بين 24 جراماً إلى أقل من32 جراماً هو :

35 – 5 = 30 قطعة

- نستطيع إيجاد القيمة الوسطى للبيانات من شكل (3-14)، فهي القيمة التي أقل منها 50 % من القيم ، وحيث إن التكرار الكلي يساوي 100 ، إذن 50 % من القطع يعنى 50 قطعة $\frac{50}{100} \times 100$ ، وبالتالي فإن القيمة الوسطى للبيانات هي عبارة عن الوزن الذي تزن 50 قطعة أقل منه .

أي أن: أقل من القيمة الوسطى 50 قطعة.

أي أن: التكرار المتجمع الصاعد للقيمة الوسطى في هذه الحالة يساوى 50 ، فبرسم خط يوازى المحورالأفقي من القيمة 50 ونمده إلى أن يلتقي مع المنحنى التكراري المتجمع الصاعد ثم من نقطة الالتقاء نسقط خطّاً عمودياً على المحور الأفقي فنحصل على القيمة الوسطى للبيانات والتي في شكل (3-14) نجدها تساوى تقريبا 34 جراماً . وهذا يعنى أن 50 قطعة أوزانها أقل من 34 جراماً .

تمارين (3)

1- الجدول التالي يبين قيمة الصادرات موزعة حسب المناطق الجغرافية لسنة 1981 وفقاً لنشرات اتجاهات التجارة الخارجية التي تعدها مصلحة الإحصاء والتعداد .

البلدان	القيمة بالدنانير الليبية (لأقرب مليون)
بلدان أوربية	2796
بلدان أفريقية	1430
بلدان أسيوية	341
أقطار الجامعة العربية وبلدان أخرى	44

مثّل البيانات السابقة :

- باستخدام القطاعات الدائرية .
- باستخدام الأعمدة البيانية .

2- البيانات التالية توضح إنتاج البطاطس والبصل الجاف والطماطم في ليبيا بآلاف الأطنان لسنة 1991 م، وذلك وفقاً للمجموعة الإحصائية لدول الوطن العربي الصادرة عن جامعة الدول العربية عام 1994م .

المحصول	الإنتاج
البطاطس	180
البصل الجاف	147
الطماطم	167

المطلوب تمثيل هذه البيانات باستخدام قطاعات دائرية .

3- فيما يلي قيمة الصـــادرات والواردات بملايين الدنانير للسنوات من 1978م إلى 1981م:

السنة	الواردات	الصادرات
1978	1363	2933
1979	1572	4762
1980	2006	6489
1981	2481	4611

مثّل البيانات السابقة باستخدام الأعمدة البيانية ، مع توضيح الواردات والصادرات باستخدام :

أ – الأعمدة البيانية المجزأة .

ب – الأعمدة البيانية المتلاصقة .

4- البيانات التالية توضح إجمالي قيمة الصادرات والواردات الليبية إلى ومن الدول العربية بملايين الدولارات الأمريكية، للسنوات من 88 إلى 91 وذلك وفقاً للمجموعة الإحصائية الصادرة عن جامعة الدول العربية عام 1994 :

السنة	الصادرات	الواردات
1988	99	193
1989	111	184
1990	460	334
1991	421	465

قارن بين الصادرات والواردات في هذه السنوات باستخدام الأعمدة البيانية المتلاصقة .

5- الجدول التالي يمثل أطوال 100 طالبٍ بالسنتيمتر :

الطول	140 إلى أقل من 150	150 إلى أقل من 160	160 إلى أقل من 170	170 إلى أقل من 180
عدد الطلبة	8	27	45	20

أ – مثّل التوزيع بمدرج تكراري .

ب – مثّل التوزيع بمضلع تكراري .

ج – مثّل التوزيع بمنحنى تكراري .

د – كوّن جدول التكرار المتجمع الصاعد ومثّله بيانياً .

6- البيانات التالية توضح توزيع الموظفين على حسب دخلهم الشهري ، وذلك بالنسبة لعدد 70 موظفاً يعملون في إحدى المؤسسات :

عدد الموظفين	فئة الدخل
7	130 إلى أقل من 150
8	150 إلى أقل من 170
13	170 إلى أقل من 190
18	190 إلى أقل من 210
15	210 إلى أقل من 230
9	230 إلى أقل من 250

أ – ارسم المدرج التكراري ، ثم ارسم على نفس الرسم المضلع التكراري .

ب – ارسم المنحنى التكراري .

ج – ارسم المنحنى التكراري المتجمع الصاعد واحسب من الرسم القيمة الوسطى للرواتب ، وعدد الموظفين الذين رواتبهم من 175 إلى أقل من 200 .

7- فيما يلي الكمية المستهلكة من الدقيق بالكيلوجرامات بالنسبة لمائة وعشرين مصنع حلويات في مدينة ما :

عدد المصانع (التكرار)	الكمية المستهلكة (الفئات)
5	20 إلى أقل من 30
15	30 إلى أقل من 40
27	40 إلى أقل من 50
33	50 إلى أقل من 60
22	60 إلى أقل من 70
18	70 إلى أقل من 80
120	المجموع

أ – ارسم المدرج التكراري .

ب – ارسم المضلع التكراري باستخدام المدرج التكراري .

ج – ارسم المنحنى التكراري المتجمع الصاعد ومن الرسم أوجد عدد المصانع التي تستهلك أقل من 55 كيلو جراماً .

د – أوجد القيمة الوسطى للكمية المستهلكة من الدقيق يومياً .

8- الجدول التالي يوضح درجات 110 طلّاب في مادة الإحصاء :

الدرجة	19-0	39-20	59-40	79-60	99-80
عدد الطلبة	5	25	35	40	5

أ – كوّن جدول التكرار المتجمع الصاعد ، ومثّله بيانياً .

ب – كوّن جدول التكرار المتجمع الهابط ، ومثّله بيانياً على نفس الرسم المتحصل عليه في (أ) ثم أوجد القيمة الوسطى من الرسم .

9- فيما يلي جدول التكرار المتجمع الصاعد لأعمار 65 مدرساً من مدرسي المرحلة الابتدائية في مدينة ما :

الأعمار(الفئات)	عدد المدرسين (التكرار المتجمع)
أقل من 25	0
أقل من 35	21
أقل من 45	38
أقل من 55	53
أقل من 65	63
أقل من 75	65

كوّن جدول التوزيع التكراري الأصلي لهذه البيانات ، ثم ارسم المدرج التكراري .

10- فيما يلي جدول التكرار المتجمع الهابط لأوزان 150 طالباً في المرحلة الثانوية :

الأوزان بالكيلو جرام (الفئات)	عدد الطلبة (التكرار المتجمع)
45 أو أكثر	150
50 أو أكثر	130
55 أو أكثر	94
60 أو أكثر	40
65 أو أكثر	15
70 أو أكثر	0

كوّن جدول التوزيع التكراري الأصلي لهذه البيانات ، ثم ارسم المضلع التكراري .

الفصل الرابع

مقاييس النزعة المركزية " المتوسطات "

الفكرة الأساسية التي تعتمد عليها مقاييس النزعة المركزية هي تمثيل مجموعة كبيرة من البيانات بقيمة واحدة ، وهذه القيمة عادة تكون في وسط البيانات أي في مركزها وهي القيمة التي تميل إليها بقية القيم وتتجمع حولها ، وبالتالي سميت هذه الخاصية التي تتمتع بها معظم البيانات بخاصية النزعة المركزية ، وأطلق على المقاييس التي تقيس هذه القيمة المتوسطة التي يحصل حولها التجمع، مقاييس النزعة المركزية أو المتوسطات .

وتؤخذ قيمة المتوسط كممثل للمجموعة كلها ، على أساس أنها قيمة غير متطرفة بل هي قيمة تتجمع حولها أغلبية القيم ، وبالتالي هي أولى من غيرها في تمثيل البيانات . ومعرفة القيمة الوسطى للبيانات تفيدنا في دراسة خصائص التوزيع التكراري والمقارنة بين التوزيعات التكرارية المختلفة لنفس الظاهرة ، ومن أهم المتوسطات التي سنتعرض لدراستها ما يلي :

- **الوسط الحسابي (المتوسط الحسابي) .**
- **الوسيط .**
- **المنوال .**

(4 – 1) الوسط الحسابي :

يعرف الوسط الحسابي بأنه القيمة التي تمثل مركز ثقل البيانات أي نقطة الارتكاز التي يحصل عندها التوازن ، فإذا كان لدينا مجموعة من البيانات ومثلناها بأثقال متساوية الوزن على لوح مدرج ، فسنجد أن هذا اللوح سيتزن إذا علق أو ثبت من مركز ثقله ومركز الثقل هذا هو الذي يمثل قيمة الوسط الحسابي لهذه القيم ، ويحسب الوسط الحسابي لأي مجموعة من القيم بجمع هذه القيم ثم قسمة المجموع على عددها .

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

ويطلق كذلك على الوسط الحسابي مصطلح المتوسط الحسابي ، وسنرمز لمتوسط العينة بالرمز $\overline{X}$ ، ونرمز لمتوسط المجتمع بالرمز μ .

وسنتعرض فيما يلي لكيفية تطبيق صيغة الوسط الحسابي في حالةالبيانات غير المبوبة أي غير المعروضة في جداول توزيعات تكرارية ، وفي حالة البيانات المبوبة أي المعروضة في جداول توزيعات تكرارية .

1 – ةةةة ةةةةة ةةةةةةة ةة ةةةة ةةةةةةةة ةةة ةةةةةةة :

إذا رمزنا للمتغير محل الدراسة بالرمز (X)، وللقيم المشاهدة لهذا المتغير بالرموز التالية :
$X_1 , X_2 , X_3 , \dots , x_n$

حيث n يرمز إلى عدد القيم المشاهدة ، فنحسب الوسط الحسابي ($\overline{X}$) كما يلي :

$$\boxed{\overline{X} = \frac{x_1 + x_2 + \cdots + x_n}{n}}$$

وإذا عبرنا عن عملية الجمع باستخدام الرمز ($\sum$) للدلالة على مجموع القيم المشاهدة ، فإننا نستطيع كتابة صيغة قانون الوسط الحسابي في حالة البيانات غير المبوبة كما يلي :

$$\boxed{\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

ةةةة (1-4) :

البيانات التالية تمثل درجات 6 في مادة الإحصاء ، (4 ، 6 ، 3 ، 7 ، 9 ، 1)، احسب الوسط الحسابي لهذه الدرجات .

ةةةة :

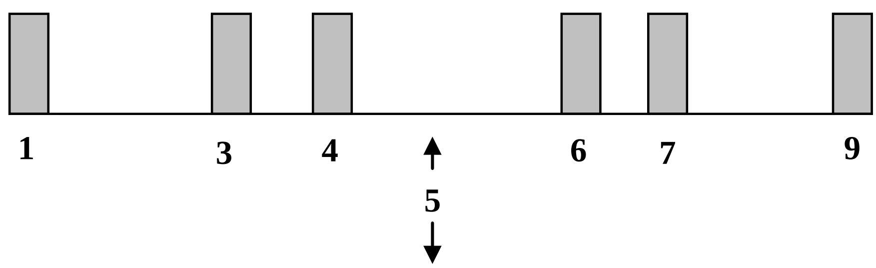
المتغير محل الدراسة في هذا المثال هو درجة الإحصاء ، وسنرمز له بالرمز X، والقيم المشاهدة التي تمثل درجات 6 طلبة سنرمز لها كما يلي :

$X_1 = 4$ ، $X_2 = 6$ ، $X_3 = 3$ وهكذا إلى ... $X_6 = 1$

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$= \frac{4+6+3+7+9+1}{6} = 5$$

وهذا يعني أن مركز الثقل أي: نقطة الارتكاز التي يحصل عندها التوازن هي الدرجة 5 ، وذلك كما هو مبيّن في شكل (4 – 1) .



الوسط الحسابي (مركز الثقل)

شكل (4 – 1)

مثال (2-4) :

البيانات التالية توضح قيمة الواردات (بملايين الدنانير الليبية) لأحد الموانى في السنوات من 1976 إلى 1980 وذلك وفقاً للنشرات الصادرة عن مصلحة الإحصاء والتعداد .

السنوات	1976	1977	1978	1979	1980
قيمة الواردات	381	272	293	333	391

احسب الوسط الحسابي لقيمة واردات هذا الميناء في هذه السنوات الخمس .

الحلّ :

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$= \frac{381+272+293+333+391}{5} = \frac{1670}{5} = 334$$

2 – حساب الوسط الحسابي للبيانات المبوبة :

أ – إذا كانت البيانات مبوبة في جدول توزيع تكراري بحيث تمثل كل فئة من فئات الجدول قيمة واحدة فقط ، ففي هذه الحالة لكي نحسب الوسط الحسابي نجرى الخطوات التالية :

• نحسب المجموع الكلي للقيم التابعة لكل فئة بضرب القيمة x_i التي تمثلها الفئة في تكرار الفئة f_i ، أي نحسب ($x_i f_i$) لكل الفئات .

• نحسب المجموع الكلي لجميع القيم وذلك بجمع المجاميع الخاصة بكل الفئات والتي تحصلنا عليها في الخطوة السابقة، أي نحسب :

($\sum_{i=1}^{k} x_i f_i$) ، حيث k = عدد الفترات (أو عدد القيم المختلفة).

• نحسب العدد الكلي للقيم المشاهدة وهو بالطبع يساوي المجموع الكلي للتكرارات ($\sum_{i=1}^{k} f_i$) .

• نحسب الوسط الحسابي بقسمة المجموع الكلي للقيم ($\sum_{i=1}^{k} x_i f_i$) على العدد الكلي للقيم ($\sum_{i=1}^{k} f_i$) .

ب– إذا كانت البيانات مبوبة في جدول تكراري بحيث كل فئة من فئات الجدول تمثل أكثر من قيمة واحدة ، ففي هذه الحالة لا نستطيع معرفة القيم المشاهدة التابعة لكل فئة ، والذي نعرفه هو عددها فقط والمتمثل في تكرار الفئة ، ولحساب المجموع الكلي للقيم التابعة لكل فئة سنفترض أن القيم المشاهدة موزعة حول مركز الفئة داخل كل فئة من فئات الجدول توزيعاً عادلاً ، وبالتالي تكون قيمة مركز الفئة هي القيمة الافتراضية لجميع القيم داخل الفئة ، ولكي نحسب الوسط الحسابي في هذه الحالة نجري الخطوات التالية :

• نحسب مركز كل فئة ونرمز للمركز بالرمز x_i .

• نحسب المجموع الكلي للقيم التابعة لكل فئة بضرب القيمة x_i (حيث x_i مركزالفئة) في تكرار الفئة f_i اي نحسب ($x_i f_i$) لكل الفئات.

• نحسب المجموع الكلي لجميع القيم وذلك بجمع المجاميع الخاصة بكل الفئات والتي تحصلنا عليها في الخطوة السابقة، أي نحسب ($\sum_{i=1}^{k} x_i f_i$) .

• نحسب العدد الكلي للقيم المشاهدة وهو بالطبع يساوي المجموع الكلي للتكرارات ($\sum_{i=1}^{k} f_i$) .

• نحسب المتوسط الحسابي بقسمة المجموع الكلي للقيم ($\sum_{i=1}^{k} x_i f_i$) على العدد الكلي للقيم ($\sum_{i=1}^{k} f_i$) .

وهكذا تكون صيغة المتوسط الحسابي للبيانات المبوبة كما يلي :

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}$$

حيث :

x : القيمة التي تمثلها الفئة (عندما الفئة تمثل قيمة واحدة فقط) ،وهي عبارة عن مركز الفئة (عندما الفئة تمثل أكثر من قيمة واحدة)

f : تكرار الفئة .

مثال (3-4) :

تمثل البيانات المنفصلة الموضحة في الجدول التكراري التالي توزيع الفصول على حسب عدد الطلبة في مدرسة ابتدائية بها 20 فصلاً ، فما هو الوسط الحسابي لعدد طلبة الفصل في هذه المدرسة؟

عدد طلبة الفصل	30	31	32	33	34	35
عدد الفصول	1	2	4	5	6	2

الحلّ :

بما أن كل فئة في هذا الجدول التكراري تمثل قيمة واحدة فقط فستكون x هي عبارة عن القيمة المذكورة في الفئة ، وسنجري الخطوات اللازمة للحصول على الوسط الحسابي والموضحة في جدول (4 – 1) :

جدول (4-1)

عدد طلبة الفصل (x_i)	عدد الفصول (f_i)	$x_i f_i$
30	1	30
31	2	62
32	4	128
33	5	165
34	6	204
35	2	70
المجموع	20	659

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} = \frac{659}{20} = 32.95$$

مثال (4-4) :

تمثل البيانات الموضحة في الجدول التكراري التالي درجات 50 طالباً :

الدرجة	عدد الطلبة
39 – 30	4
49 – 40	6
59 – 50	8
69 – 60	12
79 – 70	9
89 – 80	7
99 - 90	4

احسب الوسط الحسابي لهذه الدرجات .

الحلّ :

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

بما أن كل فئة في هذا الجدول التكراري تمثل أكثر من قيمة فستكون x_i عبارة عن مركز الفئة ، وسنجري الخطوات اللازمة للحصول على الوسط الحسابي والموضحة في جدول (4 – 2)

جدول (4-2)

الدرجة	عدد الطلبة (f_i)	مراكز الفئات (x_i)	$x_i f_i$
30 – 39	4	34.5	138.0
40 – 49	6	44.5	267.0
50 – 59	8	54.5	436.0
60 – 69	12	64.5	774.0
70 – 79	9	74.5	670.5
80 – 89	7	84.5	591.5
90 - 99	4	94.5	378.0
المجموع	50		3255.0

$$\bar{x} = \frac{3255}{50} = 65.1$$ درجـــة

مثال (4-5) :

احسب الوسط الحسابي للبيانات التالية التي تمثل أوزان 210 قطع منتجة .

الوزن (بالجرام)	عدد القطع
50 إلى أقل من 54	10
54 إلى أقل من 58	30
58 إلى أقل من 62	90
62 إلى أقل من 66	60
66 إلى أقل من 70	20

الحلّ :

الوسط الحسابي هو =
$$\overline{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}$$

وجدول (4 – 3) يوضح الحسابات اللازمة للحصول على الوسط الحسابي .

جدول (3-4)

الوزن (بالجرام)	**عدد القطع (f_i)**	**مراكز الفئات (x_i)**	$x_i f_i$
50 إلى أقل من 54	10	52	520
54 إلى أقل من 58	30	56	1680
58 إلى أقل من 62	90	60	5400
62 إلى أقل من 66	60	64	3840
66 إلى أقل من 70	20	68	1360
المجموع	210		12800

الوسط الحسابي هو :

$$\overline{x} = \frac{12800}{210} = 60.95$$ جرام

خواص الوسط الحسابي :

1- أهم مقاييس النزعة المركزية وأكثرها استعمالاً نظراً لسهولة حسابه وإمكانية التعامل معه رياضياً ، ولذلك له أهمية قصوى في التحليل الإحصائي إذ إنه يدخل في حساب كثير من المقاييس الأخرى.

2- يأخذ بعين الاعتبار جميع القيم المشاهدة ، ونستطيع أن نحصل على مجموع القيم المشاهدة إذا عرفنا قيمة الوسط الحسابي كما يلي :

مجموع القيم = الوسط الحسابي × عدد القيم

3- الوسط الحسابي هو قيمة نظرية وليس بالضروري أن تكون واحدة من القيم التي يمكن أن يأخذها المتغير محل الدراسة ، ففي المثال (4 – 3) وجدنا أن الوسط الحسابي لعدد الطلبة في الفصل يساوي 32.95 ، وهذه القيمة لا يمكن أن يأخذها المتغير محل الدراسة وهو عدد الطلبة في الفصل .

4- يتأثر بوجود قيم متطرفة في البيانات وينحاز لها ، فمثلاً إذا تبرع خمسة عشر فرداً لعمل خيري بالمبالغ التالية :

10 ، 20 ، 18 ، 10 ، 10 ، 20 ، 12 ، 15 ، 15 ، 20 ، 15 ، 15 ، 10 ، 10 ، 1000

الوسط الحسابي = $\frac{1200}{15}$ = 80 ديناراً

نــلاحظ أن الوسط الحسابي للتبرعات كبير ولا يمثل ما دفعه معظم الأفراد ، فهنا انحاز الوسط الحسابي للقيمة المتطرفة 1000، وبالتالي يعدُّ الوسط الحسابي في هذه الحالة مقياساً مضللاً . ولكن لو استبعدنا القيمة المتطرفة فسنلاحظ أن الوسط الحسابي سيكون واقعياً .

5- لا يمكن حسابه في حالة البيانات النوعية (الوصفية) .

6- لا يمكن إيجاده من الرسم (أي بيانياً) .

7- لا يمكن حسابه في حالة جداول التوزيعات التكرارية المفتوحة ؛ لأننا لا نستطيع حساب مراكز الفئات المفتوحة .

8- مجموع انحرافات القيم عن وسطها الحسابي يساوي صفراً ، حيث انحراف القيمة عن الوسط الحسابي المقصود به القيمة مطروحاً منها الوسط الحسابي ، وبالتالي فهذه الخاصية تعني أن:

$$\sum_{i=1}^{n}(x_i-\bar{x})=0$$

فمثلا : إذا كانت لدينا البيانات التالية 6 ، 8 ، 5 ، 11 ، 10

الوسط الحسابي لهذه البيانات هو: $\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}$

$$\frac{6+8+5+11+10}{5}=8$$

فسنجد أن مجموع انحرافات هذه القيم عن وسطها الحسابي يساوي صفراً ، كما هو مبين فيما يلي :

الانحرافات $(x_i - \bar{x})$	القيمة (x_i)
6 – 8 = -2	6
8 – 8 = 0	8
5 – 8 = -3	5
11– 8 = 3	11
10 – 8 =2	10
صفر	المجموع

9- مجموع مربعات انحرافات القيم عن وسطها الحسابي أقل ما يمكن ، أي أن :

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \text{نهاية صغرى}$$

ويعني ذلك أنه لو اخترنا أي قيمة أخرى سواء أكانت أقل أو أكبر من الوسط الحسابي ، فسنجد أن مجموع مربعات انحرافات القيم عن هذه القيمة دائماً أكبر من مجموع مربعات انحرافات القيم عن الوسط الحسابي .

فمثلاً : نستطيع حساب مجموع مربعات انحرافات القيم عن وسطها الحسابي بالنسبة للبيانات المذكورة في المثال السابق في الخاصية (9) ، وذلك كما يلي:

مربعات الانحرافات عن الوسط الحسابي $(x_i - 8)^2$	**انحرافات القيم عن الوسط الحسابي** $(x_i - 8)$	القيمة (x_i)
4	2-	6
0	0	8
9	3-	5
9	3	11
4	2	10
26		المجموع

وهذا يعني أن أقل قيمة يمكن أن يأخذها مجموع مربعات الانحرافات هو القيمة 26 ، ولو حسبنا مجموع مربعات انحرافات هذه القيم عن أي قيمة أخرى أكبر أو أقل من الوسط الحسابي (8) فسنجد أن مجموع مربعات الانحرافات دائماً أكبر من 26 ، فمثلاً لو حسبنا مجموع مربعات انحرافات القيم عن القيمة 9 وعن القيمة 7.5

فسنحصل على ما يلي :

القيمة (x_i)	$(x_i - 9)^2$	$(x_i - 7.5)^2$
6	9	2.25
8	1	0.25
5	16	6.25
11	4	12.25
10	1	6.25
المجموع	31	27.25

ونلاحظ أن 31 أكبر من 26 ، وكذلك 27.25 أكبر من 26 .

(2-4) الوسط الحسابي المرجح :

إذا كانت القيم المشاهدة ليس لها نفس الأهمية أو الوزن ، فعندئذ لحساب الوسط الحسابي يجب ألا نعامل جميع القيم نفس المعاملة بل نكرر كل قيمة عدداً من المرات حسب أهميتها أو وزنها أي: نرجح كل قيمة بوزنها، وعليه يطلق على الوسط الحسابي في هذه الحالة الوسط الحسابي المرجح ويرمز له بالرمز (w) .

فإذا فرضنا أن لدينا القيم التالية :

$$X_1 , X_2 , \ldots\ldots X_n$$

وكانت أهمية كل قيمة متناسبة مع الأوزان التالية :

$$w_1 , w_2 , \ldots\ldots , w_n$$

فإن الوسط الحسابي المرجح يحسب بالصيغة التالية :

$$\text{الوسط الحسابي المرجح} = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}$$

ونلاحظ أن صيغة الوسط الحسابي المرجح مماثلة لصيغة الوسط الحسابي في حالة البيانات المبوبة في جدول تكراري ؛لأنّ التكرار يعبر عن وزن أو أهمية القيمة .
كذلك إذا كان لدينا مجموعات من البيانات ، وكانت هذه المجموعات تحتوى على أعداد مختلفة من القيم المشاهدة ، وعلمنا الوسط الحسابي لكل مجموعة ، فلكي نحصل على الوسط الحسابي العام لكل المجموعات لو دمجت معا ، يجب أن نجعل عدد القيم المشاهدة في كل مجموعة يمثل وزنها، أو أهميتها، وبالتالي يحسب الوسط الحسابي العام لكل المجموعات لو دُمجت معا كما يلي :

الوسط الحسابي للمجموعات : $\overline{X}_1 , \overline{X}_2 , \quad .... \overline{X}_n$

عدد مفردات المجموعات (الوزن): $n_1 , n_2 , \; .... n_k$

فإن الوسط الحسابي العام للمجموعات (الوسط المرجح) هو :

$$W = \frac{n_1\overline{X}_1 + n_2\overline{X}_2 + \; .... \; n_k\overline{X}_k}{n_1 + n_2 \; + \; .... \; n_k} = \frac{\sum_{i=1}^{k} x_i \, n_i}{\sum_{i=1}^{k} n_i}$$

مثال (6-4) :

إذا كان لدينا 3 سلع ، وكان سعر الوحدة من السلعة الأولى 26 دينارا وسعر الوحدة من السلعة الثانية 20 دينارا وسعر الوحدة من السلعة الثالثة 6 دنانير ، فإذا كانت أهمية السلعة الثانية ضعف أهمية السلعة الأولى وكانت أهمية السلعة الثالثة خمسة أمثال أهمية السلعة الأولى ، فاحسب الوسط الحسابي لسعر الوحدة من السلع الثلاثة .

الحل :

بما أن الأهمية تختلف من سلعة إلى أخرى ، فيجب حساب الوسط الحسابي المرجح وذلك كما يلي :

	السلعة الأولى	السلعة الثانية	السلعة الثالثة
سعر الوحدة (X)	26	20	6
الأهمية (W)	1	2	5

وبتطبيق الصيغة التالية للوسط الحسابي المرجح نحصل على :

$$w = \frac{(26\times1)+(20\times2)+(6\times5)}{1+2+5} = 12 \text{ دينار}$$

مثال (4-7) :

شركتان الأولى بها 40 موظفاً وكان الوسط الحسابي لأعمارهم 48 سنة ، والثانية بها 60 موظفاً وكان الوسط الحسابي لأعمارهم 43 سنة ، فإذا انضمت الشركتان معا ، فما الوسط الحسابي العام لأعمار الموظفين في الشركة الجديدة الناتجة عن انضمام الشركتين؟

الحل :

	الشركة الأولى	الشركة الثانية
الوسط الحسابي للأعمار($\overline{x}_i$) :	48	43
الوزن (عدد الموظفين) (w_i) :	40	60

وبتطبيق الصيغة التالية نحصل على الوسط الحسابي العام للأعمار وهو :

$$w = \frac{(48\times40)+(43\times60)}{40+60} = 45 \text{ سنة}$$

(3-4) الوسيط :

الوسيط هو القيمة الموجودة في منتصف البيانات بعد ترتيبها تصاعدياً أو تنازلياً ، أي هو القيمة الوسطى التي يكون نصف البيانات أقل منها والنصف الآخر أكبر منها ويرمزله m ، ويحسب الوسيط كما يلي :

1 – في حالة البيانات غير المبوبة :

يحسب الوسيط في حالة البيانات غير المبوبة (الخام) باتّباع الخطوات التالية :

أ ــ نرتب القيم المشاهدة تصاعدياً ، أي نبدأ بأصغر قيمة ثم نرتب ما بعدها الأكبر فالأكبر .

ب ــ نحدد ترتيب الوسيط بين القيم وذلك كما يلي :

ترتيب الوسيط هو $\frac{n+1}{2}$

حيث n عدد القيم المشاهدة . ويجب الانتباه إلى أن $[(n+1) \div 2]$ هو ترتيب الوسيط وليس قيمته .

ج ــ نحدد قيمة الوسيط وتكون هي القيمة التي ترتيبها $[(n+1) \div 2]$ عندما يكون عدد القيم المشاهدة (n) عدداً فردياً ، أما إذا كانت (n) عدداً زوجياً سيقع الترتيب بين قيمتين ويكون الوسيط هو الوسط الحسابي لهاتين القيمتين .

مثال (8-4) :

أوجد قيمة الوسيط للبيانات التالية :

16 ، 20 ، 19 ، 18 ، 15 ، 17 ، 13 .

الحل :

القيم مرتبة ترتيباً تصاعدياً :

13 ، 15 ، 16 ، $\boxed{17}$ ، 18 ، 19 ، 20

↑ الوسيط

ترتيب الوسيط = $(n+1) \div 2 = (7+1) \div 2 = 4$

أي أن قيمة الوسيط هي القيمة الرابعة في البيانات بعد ترتيب البيانات تصاعدياً وبالتالي فإن :

الوسيط = 17

مثال (4-9) :

أوجد قيمة الوسيط للبيانات التالية : 3 ، 8 ، 9 ، 6 ، 5 ، 11

الحل :

نرتب القيم ترتيباً تصاعدياً :

3 ، 5 ، $\boxed{6 ، 8}$ ، 9 ، 11

ترتيب الوسيط = $(n+1) \div 2 = (6+1) \div 2 = 3.5$

• قيمة الوسيط تقع بين القيمتين الثالثة والرابعة ، إذاً الوسيط يساوي الوسط الحسابي لهاتين القيمتين أي :

$$m = \frac{6+8}{2} = 7$$

2- حساب الوسيط للبيانات المبوبة :

لحساب الوسيط في حالة البيانات المعروضة في جداول تكرارية بحيث تكون فئات الجدول مرتبة ترتيباً تصاعدياً أو تنازلياً، نتبع الخطوات التالية :

1– نحدد ترتيب الوسيط ويساوي في البيانات المبوبة = $\dfrac{\sum_{i=1}^{k} f_i}{2}$

2 – نحدد الفئة الوسيطية وهي الفئة التي تحتوي على الوسيط ويتم تحديدها بالاستعانة بالتكرار المتجمع الصاعد ، فتكون الفئة الوسيطة هي أول فئة يكون تكرارها المتجمع الصاعد أكبر من أو يساوي ترتيب الوسيط .

3 – نقوم بالتعويض في القانون التالي للحصول على قيمة الوسيط :

$$m = L + \left(\frac{\dfrac{\sum_{i=1}^{k} f_i}{2} - F}{f_m} \right) \times c$$

حيث :

L : الحد الأدنى للفئة الوسيطية ، وفي حالة التوزيعات الخاصة بمتغير منفصل نستخدم الحد الأدنى الحقيقي للفئة الوسيطية وهو عبارة عن الحد الأدنى للفئة مطروحاً منه نصف وحدة من وحدات القياس المستخدمة فمثلاً :إذا كانت الفئة الوسيطية تبدأ بالقيمة 20 فيكون الحد الأدنى الحقيقي يساوي 19.5، وهكذا...

F : التكرار المتجمع الصاعد للفئة السابقة للفئة الوسيطية .

f_m : تكرار الفئة الوسيطية .

c : طول الفئة الوسيطية .

والقاعدة الرياضية التي يعتمد عليها قانون الوسيط هو افتراض أن القيم التابعة لكل فئة موزعة حول مركزها توزيعاً عادلاً ، وبالتالي سيكون هناك تناسب بين بعد الوسيط عن الحد الأدنى للفئة الوسيطية والذي نرمز له بالرمز (L) وطول الفئة الوسيطية (C) وبين عدد القيم السابقة للوسيط داخل الفئة الوسيطية (ترتيب الوسيط - F) والعدد الكلى للقيم المشاهدة التابعة للفئة الوسيطية (f_m) ، أي أن هناك تناسباً بين المسافات والتكرارات .

مثال (4-10) :

احسب الوسيط للبيانات التالية التي تمثل أوزان 210 قطع منتجة بالجرام .

الوزن (بالجرام)	عدد القطع
50 إلى أقل من 54	10
54 إلى أقل من 58	30
58 إلى أقل من 62	90
62 إلى أقل من 66	60
66 إلى أقل من 70	20

الحل :

- نحدد ترتيب الوسيط : $\frac{\sum_{i=1}^{k} f_i}{2} = \frac{210}{2} = 105$

- نحدد الفئة الوسيطية أي نحدد الفئة التي تحتوي على الوسيط وهو القيمة التي ترتيبها 105 ويتم ذلك بالاستعانة بالتكرار المتجمع الصاعد الموضح في جدول (4 – 4) .

جدول (4-4)

الوزن (الفئة)	عدد القطع (التكرار)	التكرار المتجمع الصاعد
50 إلى أقل من 54	10	10
54 إلى أقل من 58	30	40
58 إلى أقل من 62	90	130 ← الفئة الوسيطية
62 إلى أقل من 66	60	190
66 إلى أقل من 70	20	210

حيث إن الفئة الثانية تكرارها المتجمع يساوي 40، والفئة الثالثة تكرارها المتجمع يساوي 130 ، فيعني ذلك أن التسعين قيمة الموجودة في الفئة الثالثة ترتيباتها تبدأ من الترتيب 41 إلى الترتيب 130 ، وحيث إن ترتيب الوسيط يساوي 105 ، إذن يجب أن يكون ضمن قيم الفئة الثالثة ، وبالتالي فإن الفئة الثالثة هي الفئة الوسيطية.
وبعد تحديد الفئة الوسيطية نطبق قانون الوسيط ، حيث :

$$f_m = 90 \qquad C=4 \qquad L=58 \quad F=40$$

$$m = L + \left(\frac{\frac{\sum_{i=1}^{k} f_i}{2} - F}{f_m} \right) \times c$$

$$= 58 + \frac{(105 - 40)}{90} \times 4$$

$= 58 + 2.89 = 60.89$ جرام

ويعني ذلك أن 50% من الوحدات المنتجة أي نصف الوحدات أوزانها أقل من 60.89 جراماً، و50% منها أي النصف الآخر أوزانها أكثر من 60.89 جراماً .

مثال (4-11) :
احسب الوسيط للبيانات المنفصلة التالية التي تمثل درجات 50 طالباً .

الدرجة	39-30	49-40	59-50	69-60	79-70	89-80
عدد الطلبة	4	6	8	12	11	9

الحل :

- نحدد ترتيب الوسيط حيث :

ترتيب الوسيط هو: $\sum_{i=1}^{k} f_i \div 2 = 50 \div 2 = 25$

أي أن القيمة المشاهدة التي ترتيبها 25 هي عبارة عن قيمة الوسيط ولتحديد هذه القيمة يجب إيجاد التكرار المتجمع الصاعد والموضح في جدول (4 – 5)

جدول (4-5)

الدرجة (الفئة)	عدد الطلبة (التكرار)	التكرار المتجمع الصاعد	
39 – 30	4	4	
49 – 40	6	10	
59 – 50	8	18	
69 – 60	12	30	الفئة الوسيطية
79 – 70	11	41	
89 – 80	9	50	

إذن الفئة الوسيطية هي الفئة 60 – 69، وبتطبيق قانون الوسيط ، حيث : L= 59.5
وذلك؛ لأننا نتعامل مع بيانات منفصلة ، وبالتالي نستعمل الحد الأدنى الحقيقي للفئة .

$f_m = 12 \qquad C=10 \qquad F=18$

$$m = L + \left(\frac{\frac{\sum_{i=1}^{k} f_i}{2} - F}{f_m} \right) \times c$$

$$= 59.5 + \frac{25-18}{12} \times 10$$

$$= 59.5 + 5.83 = 65.33$$

مثال (4-12) :

الجدول التالي يبين التوزيع التكراري لدخول ستين موظفاً :

الـدخـل	عدد الموظفين
150 إلى أقل من 175	8
175 إلى أقل من 200	15
200 إلى أقل من 225	20
225 إلى أقل من 250	12
250 فأكثر	5

فاحسب وسيط هذه الدخول .

الحل :

لاحظ أن هذا الجدول التكراري يحتوي على فئة مفتوحة وهي الفئة الأخيرة ومع ذلك سنحسب الوسيط كالمعتاد ، كما يلي :

جدول (4-6)

الدخل (الفئة)	عدد الموظفين (التكرار)	التكرار المتجمع الصاعد
150 إلى أقل من 175	8	8
175 إلى أقل من 200	15	23
200 إلى أقل من 225	20	43
225 إلى أقل من 250	12	55
250 فأكثر	5	60

الفئة الوسيطية

ترتيب الوسيط هو $\sum_{i=1}^{k} f_i \div 2 = 60 \div 2 = 30$

من جدول (4 – 6) نجد أن الفئة الوسيطية هي الفئة (200 إلى أقل من 225) إذن :

$f_m = 20 \qquad C=25 \qquad L=200 \qquad F=23$

$$m = L + \left(\frac{\frac{\sum_{i=1}^{k} f_i}{2} - F}{f_m} \right) \times c$$

$$= 200 + \left(\frac{30-23}{20} \right) \times 25$$

$$= 200 + 8.75$$

$= 208.75$ دينار

تحديد الوسيط بيانياً :

يحدد الوسيط بيانياً باستخدام المنحنى التكراري المتجمع الصاعد أو المنحنى التكراري المتجمع الهابط ، أو كلاهما وذلك كما يلي :

- نحدد أولاً ترتيب الوسيط $\sum_{i=1}^{k} f_i \div 2$ على المحور الرأسي والذي يمثل التكرارات المتجمعة .
- نرسم من هذه النقطة خطاً أفقياً يوازي محور السينات حتى يلاقي المنحنى المتجمع الصاعد أو الهابط في نقطة .
- من نقطة التلاقي نسقط عموداً يلاقي محور السينات في نقطة تكون هي قيمة الوسيط ، وذلك كما هو موضح في شكل(4 – 2) ، وشكل (4 – 3) .
- كذلك نستطيع إيجاد الوسيط برسم المنحنيين (الصاعد والهابط) معاً في رسم واحد ، فيكون الإحداثي السيني لنقطة تقاطع المنحنيين هو قيمة الوسيط ، وهي القيمة التي كنا نطلق عليها القيمة الوسطى، وذلك عند دراسة المنحنى المتجمع الصاعد والمنحنى المتجمع الهابط .

مثال (4-13) :

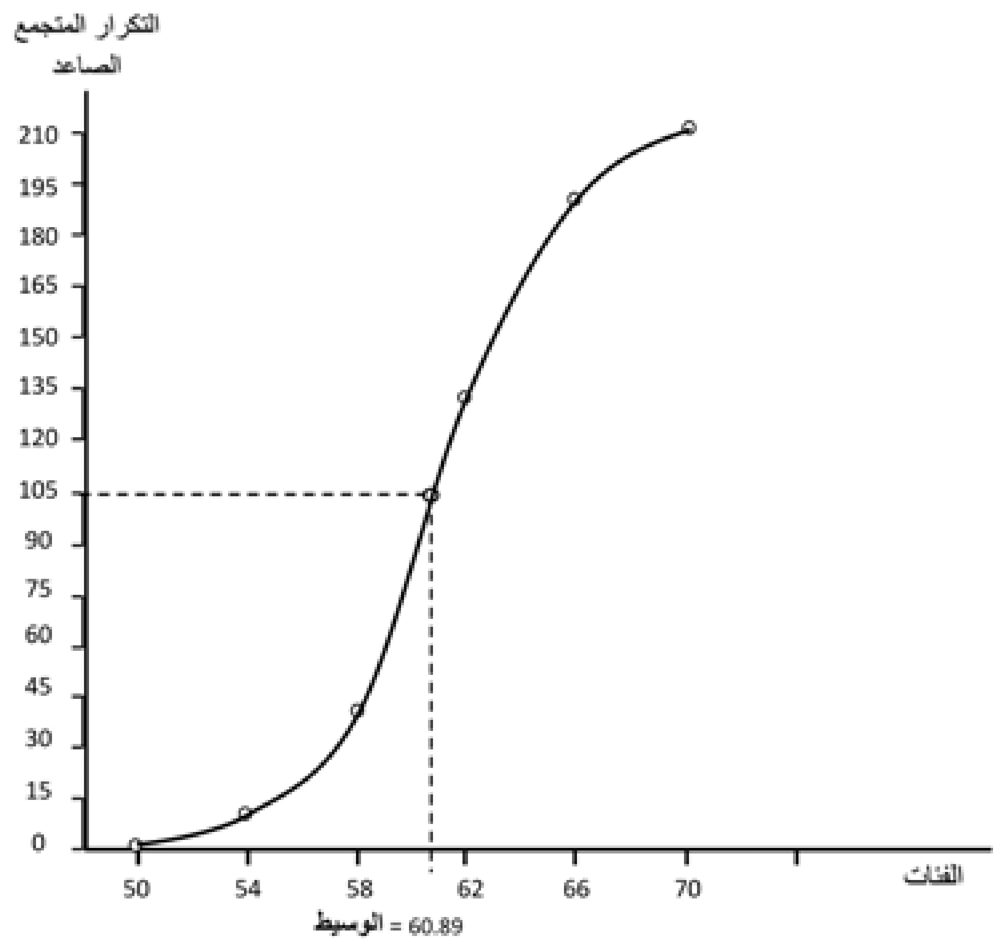
حدد قيمة الوسيط بيانياً للتوزيع التكراري الخاص بأوزان 210 وحدة منتجة المذكورة في مثال (4 – 10) .

الحلّ :

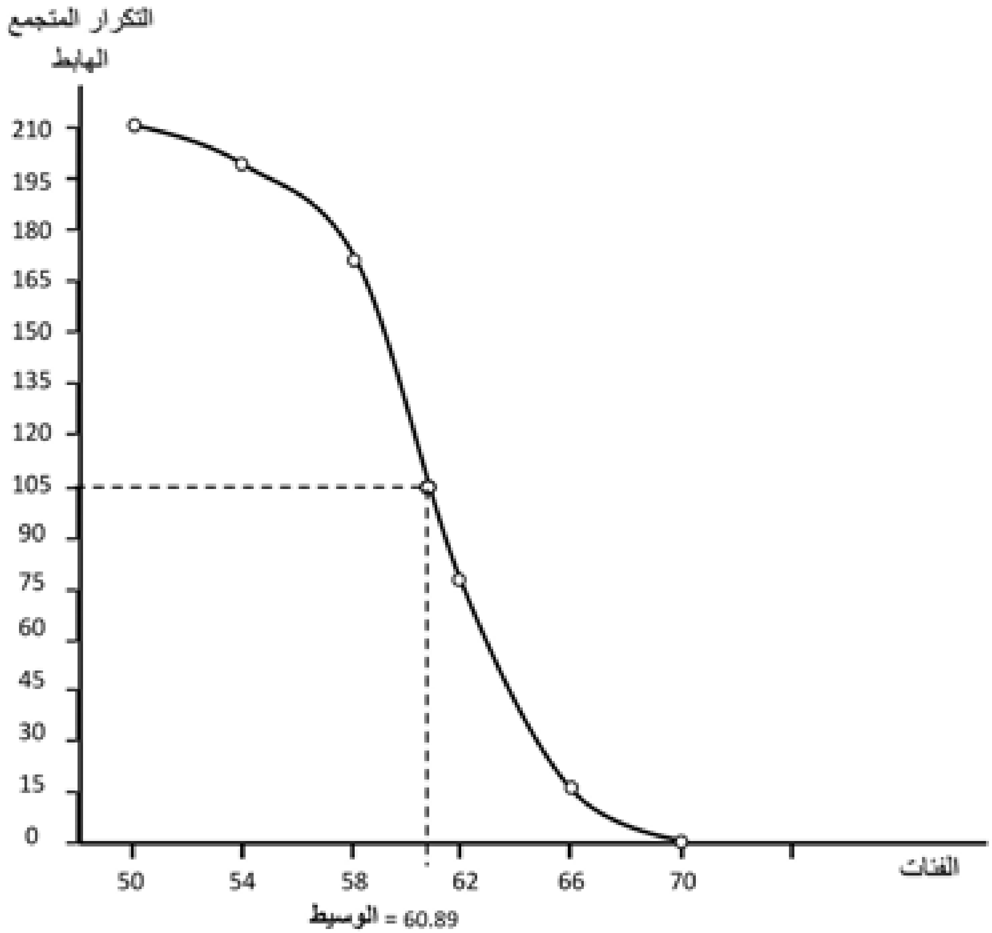
نكوّن أولاً جدول التكرار المتجمع الصاعد أو جدول التكرار المتجمع الهابط لهذه البيانات والموضحة معاً في جدول (4 – 7) ، وبعد رسم المنحنى المتجمع وتحديد ترتيب الوسيط الذي يساوي 105 على المحور الرأسي ، نرسم من هذه النقطة خطاً أفقياً يوازي محور السينات حتى يلاقى المنحنى المتجمع ومن نقطة التلاقي نسقط عموداً ليلاقي محور السينات فنحصل على قيمة الوسيط ، وذلك كما هو موضح في شكل (4 – 2)، وشكل (4 – 3) .

جدول (4 – 7)

الفئة	التكرار المتجمع الصاعد	الفئة	التكرار المتجمع الهابط
أقل من 50	0	50 أو أكثر	210
أقل من 54	10	54 أو أكثر	200
أقل من 58	40	58 أو أكثر	170
أقل من 62	130	62 أو أكثر	80
أقل من 66	190	66 أو أكثر	20
أقل من 70	210	70 أو أكثر	0



شكل (4-2)



شكل (4-3)

ومن شكل (4 – 2)، وشكل (4 – 3)، نجد أن قيمة الوسيط تقريباً 61 جراماً .

خواص الوسيط :

1- سهل التعريف وسهل الحساب .

2- يعتمد على القيمة الوسطى فقط إذا كانت (n) فردية وعلى القيمتين الوسيطيتين إذا كانت (n) زوجية ، ويهمل بقية القيم .

3- لا يتأثر بوجود قيم متطرفة .

فمثلاً: إذا كان لدينا القيم التالية : 2 ، 7 ، 6 ، 4 ، 10 ، 1 ، 9

الترتيب التصاعدي لهذه البيانات : 1 ، 2 ، 4 ، 6 ، 7 ، 9 ، 10

ترتيب الوسيط = $(7 + 1) \div 2 = 4$

إذن الوسيط هو القيمة الرابعة بعد الترتيب التصاعدي أي الوسيط = 6 وإذا وضعنا بدلاً من القيمة 10 ، القيمة 84 مثلاً وهي قيمة متطرفة بالنسبة لقيم المجموعة ، فسنجد أن الوسيط لن يتغير وسيظل مساوياً 6 ، وبالتالي فالوسيط لا يتأثر بالقيم المتطرفة .

4- يمكن إيجاد الوسيط للبيانات النوعية (الوصفية) بشرط إمكانية ترتيبها تصاعدياً أو تنازلياً ويكون عددها فردياً .

فمثلاً إذا كان لدينا البيانات النوعية التالية والتي تمثل تقديرات 9 طلبة في مادة الإحصاء : مقبول، جيد، ضعيف، ممتاز، جيد جداً، مقبول، مقبول، جيد، مقبول الترتيب التصاعدي للبيانات :

ضعيف،مقبول،مقبول، مقبول،مقبول،جيد،جيد، جيد جداً، ممتاز

ترتيب الوسيط = $(9 + 1) \div 2 = 5$

إذن الوسيط هو التقدير الخامس بعد ترتيب البيانات تصاعدياً ،أي أن الوسيط هو تقدير مقبول.

5- يمكن حسابه من جداول التوزيعات التكرارية المفتوحة وذلك كما هو واضح في مثال (4 – 12) .

6- يمكن إيجاد الوسيط بيانياً .

7- يستعمل الوسيط في الحالات التي تكون فيها بعض البيانات ناقصة بشرط أن نعرف ترتيبها؛ فمثلاً إذا أردنا إيجاد الوسيط للمدة التي يقضيها العامل في إنتاج سلعة معينة نكتفي هنا بتسجيل المدة التي يستغرقها 50% من العاملين؛ لأن المدة التي سيستغرقها النصف الآخر من العاملين الذين لم ينتهوا بعد ستكون أكبر من الوسيط .

(4-4) المنوال :

المنوال هو القيمة أو الصفة الأكثر شيوعاً في البيانات ، أي القيمة أو الصفة التي لها أكبر تكرار ، أي التي تتكرر أكثر من غيرها من القيم أو الصفات .

ويحسب المنوال كمايلي:

1 – في حـالة البيانات غير المبوبة :

في هذه الحالة لا توجد أي عمليات حسابية لإيجاد المنوال ، كل ما يتطلبه إيجاد المنوال هو معرفة القيمة التي تتكرر أكثر من غيرها ، وذلك كما هو واضح في المثال التالي :

مثال (4-14) :

أوجد قيمة المنوال للبيانات التالية : 11 ، 12 ، 9 ، 10 ، 9 ، 11 ، 13 ، 9

الحل :

إن منوال هذه القيم يساوي 9 ، وذلك لأنها القيمة التي تتكرر أكثر من غيرها .
إذا كان في البيانات منوال واحد فتسمى بيانات وحيدة المنوال ، وإذا وجد في البيانات منوالان فتسمى بيانات ثنائية المنوال ، وإذا وجد في البيانات أكثر من منوالين فتسمى بيانات عديدة المنوالات ، وأحياناً لا توجد في البيانات قيمة أو صفة تتكرر أكثر من غيرها من القيم فتسمى بيانات عديمة المنوال .

مثال (4-15) :

أوجد المنوال للبيانات المذكورة في كل مجموعة من المجموعات التالية :

المجموعة الأولـى : 18 ، 14 ، 15 ، 19 ، 20 ، 24
المجموعة الثانيــة : 12 ، 10 ، 12 ، 16 ، 10 ، 15 ، 14
المجموعة الثالثـة : 12 ، 12 ، 15 ، 13 ، 15 ، 13
المجموعة الرابعة : 12 ، 15 ، 19 ، 13 ، 16 ، 14، 15، 15
المجموعة الخامسة : 10 ، 9 ، 15، 12، 5 ، 11، 5 ، 10 ، 12

الحل :

في المجموعة الأولى لا يوجد منوال (بيانات عديمة المنوال) .
في المجموعة الثانية يوجد منوالان 12،10 (بيانات ثنائية المنوال).
في المجموعة الثالثة يوجد 3 منوالات 12، 13، 15 (بيانات عديدة المنوال) .
في المجموعة الرابعة يوجد منوال واحد 15 (بيانات أحادية المنوال) .
في المجموعة الخامسة 3 منوالات 5،12،10 (بيانات عديدة المنوال) .

مثال (4 – 16) :

أوجد المنوال للبيانات النوعية التالية التي تبين جنسيات 6 مدرسين يُدرسون في مدرسة لتعليم اللغة الإنجليزية :

عربي ، إنجليزي ، أمريكي ، إنجليزي ، إنجليزي ، استرالي

فهنا المنوال هو الجنسية الإنجليزية ، أي الجنسية المتكررة أكثر من غيرها من الجنسيات

2 – في حالة البيانات المبوبة :

لحساب المنوال لبيانات مبوبة في جداول تكرارية منتظمة أي فئاتها متساوية الطول نتبع الخطوات التالية :

- نحدد أولاً الفئة التي تحتوي المنوال ويطلق عليها الفئة المنوالية وهي الفئة المقابلة لأكبر تكرار .
- نحسب قيمة المنوال باستخدام القانون التالي :

$$\boxed{M = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C}$$

حيث :

L: الحد الأدنى للفئة المنوالية (أو الحد الأدنى الحقيقي في حالة البيانات التي تمثل متغيراً منفصلاً) .

Δ_1 = تكرار الفئة المنوالية – تكرار الفئة السابقة لها .

Δ_2 = تكرار الفئة المنوالية – تكرار الفئة اللاحقة لها .

C: طول الفئة المنوالية .

ملاحظة : إذا كان جدول التوزيع التكراري غير منتظم يجب تعديل تكراراته قبل تطبيق خطوات إيجاد المنوال التي أشرنا إليها .

مثال (17-4) :

أوجد قيمة المنوال للبيانات المذكورة في الجدول التالي والتي تمثل أوزان 210 قطع منتجة بالجرام :

الوزن	عدد القطع
50 إلى أقل من 54	10
54 إلى أقل من 58	30
58 إلى أقل من 62	90
62 إلى أقل من 66	60
66 إلى أقل من 70	20
المجموع	210

الحل :

- الفئة المنوالية هي الفئة (58 إلى أقل من 62) لأنها هي الفئة التي لها أكبر تكرار ونجد أن :

$L = 58$ ، $C = 4$

$\Delta_1 = 90-30 = 60$ ، $\Delta_2 = 90-60 = 30$

المنوال هو :

$$M = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C$$

$$= 58 + \frac{60}{60 + 30} \times 4 = 58 + \frac{24}{9} = 60.67$$

مثال (4-18) :

احسب قيمة المنوال للبيانات التالية ، التي تمثل درجات 50 طالباً .

الدرجة	39-30	49-40	59-50	69-60	79-70	89-80
عدد الطلبة	4	6	8	12	11	9

الفئة المنوالية

الحل :

الفئة المنوالية هي الفئة (60 – 69) ونجد أن :

$L = 59.5$ (لأن البيانات هنا تمثل متغيراً منفصلاً) ، $C = 10$

$\Delta_1 = 12 - 8 = 4$ $\Delta_2 = 12 - 11 = 1$

$$M = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C$$

$$= 59.5 + \frac{4}{4+1} \times 10 = 67.5$$

حساب المنوال بيانياً :

يمكن تحديد قيمة المنوال بيانياً باستخدام المنحنى التكراري أو المضلع فتكون قيمة المنوال هي القيمة المقابلة لقمة المنحنى ، لأن القمة تمثل أكبر تكرار ، وحيث إن المنحنى يكون ممهداً باليد فغالباً ستكون قيمة المنوال التي نتحصل عليها بهذه الطريقة غير دقيقة ، كما يمكن تحديد قيمة المنوال باستخدام المدرج التكراري.

مثال (4-19) :

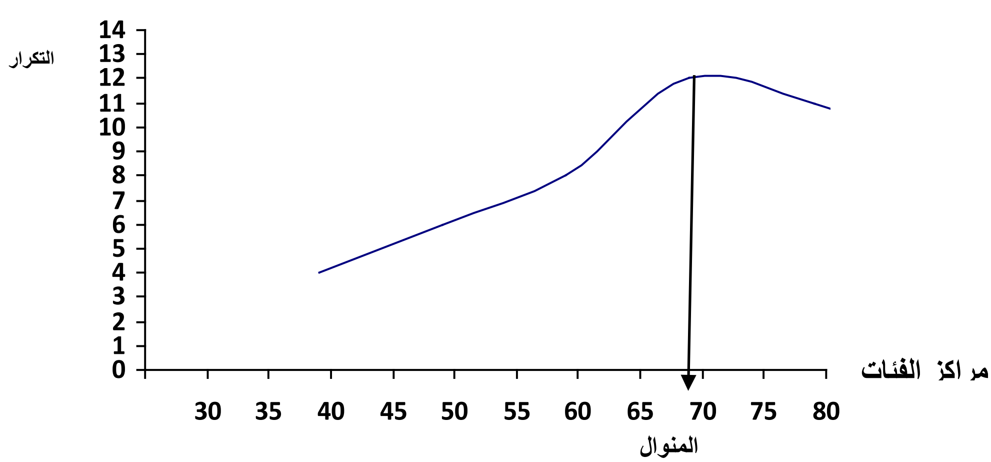
احسب المنوال بيانياً ، للبيانات المذكورة في مثال (4 – 18) باستخدام المنحنى التكراري.

الحلّ :

نحدد مراكز الفئات ، وذلك لتحديد النقاط التي تمثل الفئات على الرسم البياني ، والجدول التالي يوضح ذلك ، ثم نرسم المنحنى التكراري ونحدد منه قيمة المنوال والتي تساوي 67.5 كما هو واضح في شكل (4 – 4) .

جدول (4-8)

الفئة	التكرار	مركز الفئة	النقاط (y،x)
39 – 30	4	34.5	(4 ، 34.5)
49 – 40	6	44.5	(6 ، 44.5)
59 – 50	8	54.5	(8 ، 54.5)
69 – 60	12	64.5	(12 ، 64.5)
79 – 70	11	74.5	(11 ، 74.5)
89 – 80	9	84.5	(9 ، 84.5)



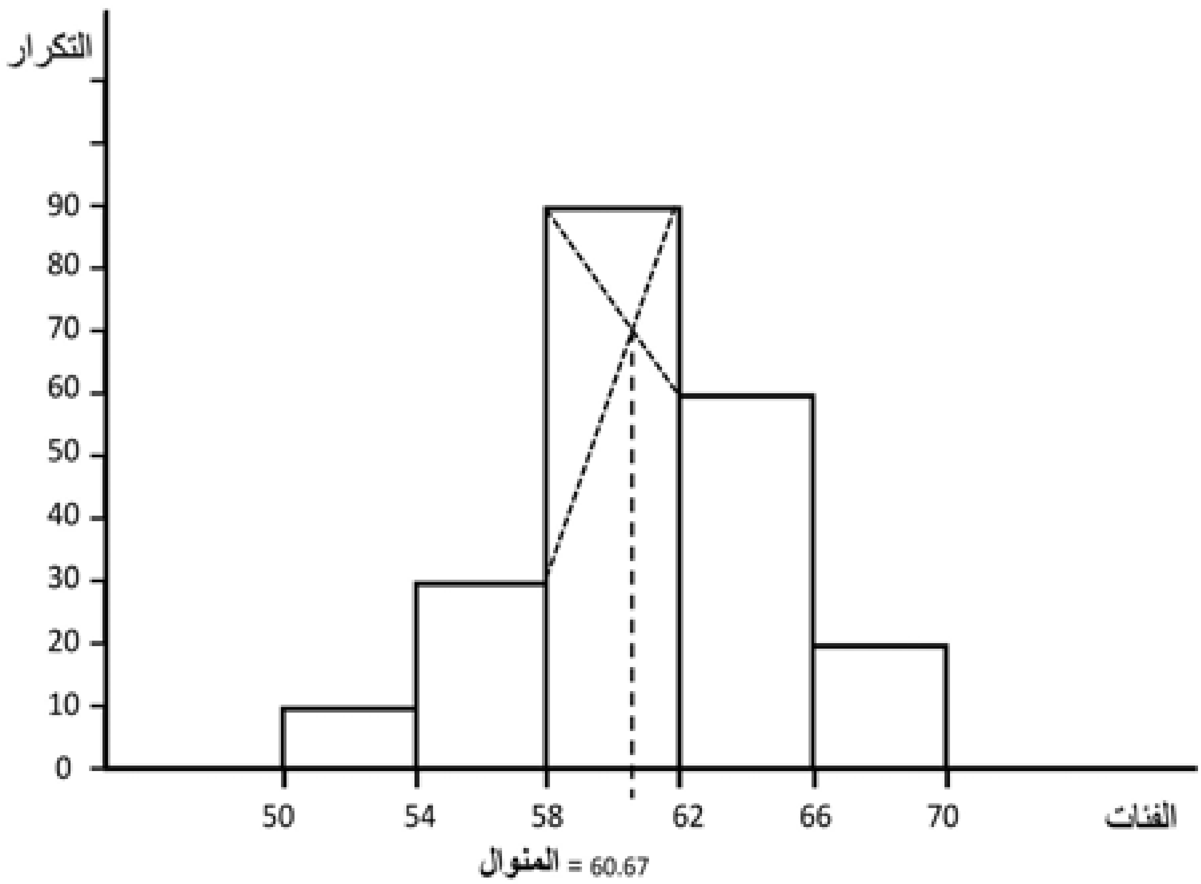
شكل (4 – 4)

مثال (4-20) :

احسب المنوال بيانياً باستخدام المدرج التكراري لبيانات المثال(4-17).

الحل :

نرسم المدرج التكراري ثم نحدد مستطيل الفترة المنوالية ومستطيل الفترة السابقة لها ومستطيل الفترة اللاحقة لها وبتوصيل الحد الأدنى للقاعدة العليا لمستطيل الفترة المنوالية بالحد الأدني للقاعدة العليا لمستطيل الفترة اللاحقة ، وبتوصيل الحد الأعلى لمستطيل الفترة المنوالية بالحد الأعلى للقاعدة العليا لمستطيل الفترة السابقة لها ومن نقطة تقاطع المستقيمين نسقط عموداً على المحور الأفقي فتكون نقطة التقاء العمود مع المحور الأفقي هي قيمة المنوال كما هو واضح في شكل (4-5).



شكل (4 – 5)

خواص المنوال :

1- أسهل مقاييس النزعة المركزية في حسابه .
2- لا يتأثر بوجود قيم متطرفة .
3- يمكن حسابه في حالة التوزيعات التكرارية المفتوحة بشرط ألا تكون الفئة المفتوحة هي الفئة المنوالية .
4- يمكن إيجاد المنوال للبيانات النوعية وذلك كما هو واضح في مثال (4 – 16) .
5- ليس له معنى إذا كانت البيانات قليلة العدد وقد لا يوجد أصلاً ، أما في حالة البيانات كثيرة العدد فله معنى معقول وله أهمية كبيرة وخاصة في عملية التسويق ، فمثلاً شركات تسويق الأحذية في مدينة ما لا تهتم بالوسط الحسابي أو بالوسيط بل تهتم بالمقياس الأكثر شيوعاً وهو المنوال .
6- يمكن إيجاد المنوال بيانياً .
7- قد لا يكون للبيانات منوالٌ وقد تحتوي على منوالين أو أكثر .
8- يتأثر كثيراً بطريقة اختيار الفئات التكرارية للتوزيع ، فإذا غيرنا تقسيم الفئات لنفس التوزيع فيحدث تغيراً في التكرارات، وفي الغالب يحدث تغيراً في موقع الفئة المنوالية ، ولذلك نحصل على قيم مختلفة للمنوال .

تمارين (4)

1- ما المقصود بخاصية النزعة المركزية ؟ وما أهم مقاييسها ؟ مع تعريف كل مقياس من هذه المقاييس .

2- البيانات التالية تمثل درجات امتحان في مادة اللغة الإنجليزية لتسعة طلبة :

43 ، 56 ، 30 ، 42 ، 64 ، 60 ، 30 ، 70 ، 82

أ – احسب الوسط الحسابي لهذه الدرجات .

ب – احسب الوسيط .

ج – أوجد المنوال .

3- فيما يلي عدد القطع المنتجة شهرياً من قبل 10 عاملين :

14 ، 18 ، 17 ، 14 ، 17 ، 15 ، 14 ، 17 ، 15 ، 12

أ – احسب الوسط الحسابي .

ب – احسب الوسيط .

ج – أوجد المنوال .

4- إذا كان الوسط الحسابي لعشرة قيم هو 62، وإذا كان مجموع انحرافات 9 قيم منها عن الوسط الحسابي هو 5 ، فما هي القيمة العاشرة ؟

5- إذا كان لدينا البيانات التالية : 35 ، 25 ، 20 ، 33 ، 180 ، 21 ، 27

أوجد قيمة المتوسط المناسب لهذه البيانات ، مع ذكر لماذا يفضل هذا المتوسط عن المتوسطين الآخرين ؟

6- بعد رصيد درجات 10 طلبة في مادة الرياضيات وجدنا أن الوسط الحسابي لهذه الدرجات = 65 ، ثم انتبهنا أن هناك خطأ في تسجيل درجات 3 طلبة ، وذلك كما يلي :

الدرجة المسجلة	الدرجة الصحيحة
65	56
57	75
28	29

فأوجد الوسط الحسابي بعد إجراء عملية التصحيح .

7- علل مايلي :

أ. لا نستطيع حساب الوسيط للبيانات النوعية إلا إذا كان عددها فردياً .

ب. لا نستطيع حساب الوسط الحسابي للجداول التكرارية المفتوحة.

8- ما مقياس النزعة المركزية المناسب في كل حالة من الحالات التالية :

أ. بيانات نوعية لا يمكن ترتيبها ترتيباً تصاعدياً أو تنازلياً .

ب. بيانات مبوبة في جدول تكراري فئته الأخيرة مفتوحة ، مع العلم بأن هذه الفئة المفتوحة تكرارها أقل من نصف التكرارات وهو أكبر تكرار في الجدول .

جـ. بيانات تتكون من قيم لها نفس التكرار وتحتوي على قيم متطرفة .

9- إذا كانت القيم التالية مرتبة ترتيباً تصاعدياً : 12 ، 20 ، س ، 30 ، 33 ، 40

فما هي القيمة المجهولة (س) إذا علمت أن قيمة الوسيط = 28 ؟

10- فيما يلي جدول التوزيع التكراري لعدد أطفال العائلات القاطنة في حي من إحدى المدن:

عدد الأطفال (الفئة)	0 - 2	3-5	6-8	9-11	12-14
عدد العائلات (التكرار)	25	45	60	40	30

أ – احسب الوسط الحسابي   ب – احسب الوسيط .

ج – أوجد قيمة المنوال .

11- الجدول التالي يوضح توزيع 500 عامل في أحد المصانع حسب دخولهم الأسبوعية :

الدخل الأسبوعي (بالدينار)	عدد العاملين
20 إلى أقل من 25	100
25 إلى أقل من 30	180
30 إلى أقل من 35	120
35 إلى أقل من 40	50
40 إلى أقل من 45	30
45 إلى أقل من 50	12
50 إلى أقل من 55	8

أ – احسب الوسط الحسابي .

ب – احسب الوسيط حسابياً وبيانياً .

ج – أوجد قيمة المنوال حسابياً وبيانياً .

12- إذا علمت أن الجدول التكراري التالي يمثل بيانات عن 234 مفردة، وأن الوسيط لهذه البيانات = 46 ، فأوجد تكرار الفئة الثالثة f_3 ، وتكرار الفئة الخامسة f_5 ، (حدد الفئة الوسيطية باستخدام قيمة الوسيط) .

الفئة	التكرار
10 إلى أقل من 20	12
20 إلى أقل من 30	30
30 إلى أقل من 40	f_3
40 إلى أقل من 50	65
50 إلى أقل من 60	f_5
60 إلى أقل من 70	25
70 إلى أقل من 80	18

13-أوجد مقياس النزعة المركزية المناسب لهذه البيانات النوعية التي تمثل مهنة أربع وعشرين أماً من الأمهات اللائي يترددن على العيادة الخاصة بالأمومة والطفولة ؟

مدرسة	ربة بيت	ربة بيت	مدرسة	عاملة	ربة بيت
طبيبة	ربة بيت	مهندسة	ربة بيت	عاملة	ربة بيت
مدرسة	موظفة	مدرسة	ربة بيت	موظفة	مدرسة
مدرسة	مدرسة	موظفة	موظفة	ربة بيت	عاملة

14. الجدول التكراري التالي يوضح توزيع 120 طالباً حسب أوزانهم :

الوزن (بالكيلوجرام)	عدد الطلبة
40 إلى أقل من 50	8
50 إلى أقل من 60	14
60 إلى أقل من 70	38
70 إلى أقل من 80	30
80 إلى أقل من 90	20
90 إلى أقل من 100	10

أ. احسب قيمة الوزن الذي أكثر من 6/10 من الأوزان .

ب. أوجد قيمة الوزن الذي أقل منه 85 % من الأوزان .

الفصل الخامس

مقاييس التشتت

مقاييس النزعة المركزية (المتوسطات) السابق دراستها ، تدلنا على القيمة التي تتجمع حولها القيم المشاهدة للظاهرة محل الدراسة، ولكن هذه المقاييس لا تعطينا أي فكرة عن درجة انتشار واختلاف القيم وتباعدها عن بعضها أو تباعدها عن القيمة المركزية لها أي عن متوسطاتها ، وخاصية تباعد قيم التوزيع واختلافها يطلق عليها خاصية التشتت ، وبالتالي فإن مقياس النزعة المركزية وحده لا يكفي لوصف مجموعة البيانات محل الدراسة ، فقد تتساوى المتوسطات لمجموعتين أو أكثر من البيانات بالرغم من اختلاف القيم في هذه المجموعات .

مثال (5-1) :

إذا كـان لدينا المجموعات التالية من البيانات :

المجموعة الأولى : 6 6 6 6 6

المجموعة الثانية : 4 5 6 7 8

المجموعة الثالثة : 0 1 6 10 13

نلاحظ أن المجموعات الثلاث لها نفس الوسط الحسابي والوسيط، فكلاهما في جميع المجموعات يساوي القيمة 6 ، وذلك بالرغم من وجود اختلاف واضح بين المجموعات الثلاث في انتشار القيم وتباعدها .

ففي المجموعة الأولى نلاحظ عدم وجود اختلاف أو تشتت بين القيم فكل القيم متساوية وتساوي قيمة الوسط الحسابي وهي القيمة 6 .

أما في المجموعة الثانية ، نلاحظ أن القيم تختلف عن بعضها بعضاً وعن وسطها الحسابي ولكن ليس اختلافاً كبيراً ، أي أن تشتت القيم داخل هذه المجموعة صغير .

بينما في المجموعة الثالثة ، نلاحظ أن القيم تختلف عن بعضها بعضاً وعن وسطها الحسابي اختلافاً كبيراً ، أي أن تشتت القيم داخل هذه المجموعة كبير .

نستنتج من ذلك أن تساوي متوسطات المجموعات لا يعني أن البيانات في هذه المجموعات متكافئة ، ولوصف البيانات محل الدراسة وصفاً جيداً يجب بالإضافة إلى تحديد القيمة التي تتجمع حولها القيم ، معرفة كيفية انتشار هذه القيم أي تشتتها ، ولقياس التشتت نستخدم مقاييس إحصائية يطلق عليها مقاييس التشتت وأهمها ما يلي :

- **المدى .**
- **الانحراف الربيعي .**
- **متوسط الانحرافات المطلقة .**
- **التباين .**
- **الانحراف المعياري .**

بالإضافة إلى **معامل الاختلاف** الذي يستخدم لمقارنة مجموعتين أو أكثر من البيانات من حيث التشتت .

وكلما زاد تشتت القيم داخل مجموعة البيانات ، زادت قيمة مقياس التشتت . مع العلم بأن كل مقاييس التشتت هي مقاييس معتمدة على الأرقام وبالتالي لا يمكن استخدامها إلا في حالة البيانات الكمية .

وسيتم دراسة كل من المدى والتباين ، والانحراف المعياري ومعامل الاختلاف لأهمية هذه المقاييس في التطبيقات العملية .

(5-1) المدى :

المدى هو الفرق بين أكبر قيمة وأصغر قيمة في البيانات ، فهو عبارة عن الفترة التي يتغير فيها المتغير محل الدراسة ويرمز له بالرمز R.

ويحسب المدى في حالة البيانات غير المبوبة كما يلي :

المدى = أكبر قيمة - أصغر قيمة

أما في حالة البيانات المبوبة نعتبر أكبر قيمة هي الحد الأعلى للفئة الأخيرة وأصغر قيمة هي الحد الأدنى للفئة الأولى مع مراعاة أن تكون فئات الجدول مرتبة ترتيباً تصاعدياً ، أي يحسب المدى في حالة البيانات المبوبة كما يلي :

المدى = الحد الأعلى للفئة الأخيرة - الحد الأدنى للفئة الأولى

إذا كان المدى صغيراً فيعنى ذلك أن البيانات منتشرة في فترة قصيرة أي قريبة من بعضها وتشتتها صغير ، أما إذا كان المدى كبيراً فيعني ذلك أن البيانات منتشرة في فترة طويلة أي متباعدة عن بعضها وتشتتها كبير .

مثال (5-2) :

باستخدام المدى قارن بين تشتت المجموعات الثلاث المذكورة في مثال (5 – 1) .

الحل :

مدى المجموعة الأولى = 6 – 6 = 0

مدى المجموعة الثانية = 8 - 4 = 4

مدى المجموعة الثالثة = 13 – 0 = 13

بما أن مدى المجموعة الأولى يساوي صفراً ، فيعني ذلك أنه لا يوجد اختلاف أو تشتت بين قيم هذه المجموعة ، أي أن كل القيم داخل هذه المجموعة متساوية .

ونلاحظ أن مدى المجموعة الثانية أصغر من مدى المجموعة الثالثة، وهذا يعني أن تشتت القيم داخل المجموعة الثانية أقل من تشتت القيم داخل المجموعة الثالثة .

مثال (5-3) :

قارن بين تشتت درجات مادة الإحصاء لثلاث مجموعات من الطلبة :

درجات المجموعة الأولى : 50 ، 62 ، 30 ، 70 ، 45 ، 80

درجات المجموعة الثانية : 0،20 ، 70، 55، 65 ،82، 35،77

درجات المجموعة الثالثة :49،54 ، 85، 50، 60 ،63، 75،65

الحل :

مدى المجموعة الأولى = 80 – 30 = 50 درجة

مدى المجموعة الثانية = 82 – 0 = 82 درجة

مدى المجموعة الثالثة = 85 – 49 = 36 درجة

لاحظ أن المجموعة الثالثة أقلهن تشتتاً ثم يليها المجموعة الأولى ، ثم يليها المجموعة الثانية .

مثال (5-4) :

احسب المدى للبيانات التالية التي توضح أوزان 100 طالب بالكيلو جرامات :

الوزن (بالكيلو جرام)	عدد الطلبة
60 إلى أقل من 63	5
63 إلى أقل من 66	15
66 إلى أقل من 69	40
69 إلى أقل من 72	28
72 إلى أقل من 75	12

الحلّ :

أكبر قيمة = الحد الأعلى للفئة الأخيرة = 75 كيلو جراماً

أصغر قيمة = الحد الأدنى للفئة الأولى = 60 كيلو جراماً .

المـــــدى هو :   كيلو جرام 15 = 60 - 75 =R

خــواص المدى :

1- المدى مقياس سهل في حسابه وبسيط في مفهومه ودلالته .

2- يهتم بقيمتين فقط في البيانات ويهمل بقية القيم .

3- يعتمد في حسابه على القيمة الكبرى والقيمة الصغرى فقط ولذلك يعتبر مقياساً مضللاً ، لأنه عندما تكون القيمة الكبرى أو القيمة الصغرى أو كلتاهما قيماً شاذة ففي هذه الحالة يكون المدى كبيراً بينما قيم المجموعة تكون غير متباعدة ، فمثلاً إذا كان لدينا البيانات التالية :

10 ، 12 ، 8 ، 9 ، 11 ، 14 ، 55

فنجد أن : المدى هو $R = 55 - 8 = 47$

فالمدى كبير ، ويشير إلى وجود تشتت كبير في المجموعة في حين أن القيم متقاربة ، ولذلك فهو مقياس مضلل ، وسبب ذلك هو اعتماده على القيم المتطرفة ، فلو حذفنا القيمة المتطرفة وهي القيمة 55 فنجد أن قيمة المدى تساوي 14 – 8 = 6 وهي قيمة صغيرة وواقعية .

4- لا نستطيع حساب المدى في حالة التوزيعات التكرارية المفتوحة وذلك لأنه في هذه الحالة يكون الحد الأدنى للفئة الأولى أو الحد الأعلى للفئة الأخيرة أو كلاهما مجهولاً .

(2-5) التباين :

يُعرف التباين بأنه الوسط الحسابي لمربعات انحرافات القيم عن وسطها الحسابي ويرمز له بالرمز S^2 .ويحسب على النحو التالي :

أ – التباين في حالة البيانات غير المبوبة :

- نحسب الوسط الحسابي للبيانات
- نحسب انحراف كل قيمة عن الوسط الحسابي ،
  حيث : انحراف القيمة عن الوسط الحسابي = القيمة - الوسط الحسابي
- نوجد مربع انحراف كل قيمة عن الوسط الحسابي .
- نوجد مجموع مربعات انحرفات القيم عن الوسط الحسابي .
- نحسب قيمة التباين باستخدام القانون التالي :

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

مثال (5-5) :

احسب التباين للبيانات التي تمثل درجات 8 طلبة وهي : 7 ، 6 ، 10 ، 9 ، 8 ، 5 ، 5 ، 6

الحل : الوسط الحسابي هو :

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{56}{8} = 7$$

$(X_i - \overline{X})^2$	$(X_i - \overline{X})$	**القيم** (X_i)
0	0	7
1	1-	6
9	3	10
4	2	9
1	1	8
4	2-	5
4	2-	5
1	1-	6
24	$\sum_{i=1}^{n} (X_i - \overline{X})^2$	

التباين هو :

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{24}{7} = 3.43$$

ب – التباين في حالة البيانات المبوبة :

في حالة البيانات المعروضة في جداول تكرارية نتبع الخطوات التالية :

- نحسب مراكز الفئات .
- نحسب الوسط الحسابي للتوزيع .
- نحسب انحراف كل مركز عن الوسط الحسابي .
- نوجد مربع انحراف كل مركز عن الوسط الحسابي .
- يضرب مربع انحراف كل مركز عن الوسط الحسابي لفئة في تكرار هذه الفئة .
- نجمع حواصل الضرب التي تحصلنا عليها في الخطوة السابقة .
- نحسب قيمة التباين باستخدام القانون التالي :

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

حيث : x_i : مركز الفئة ، f_i : تكرار الفئة ، $\bar{x}$: الوسط الحسابي

مثال (6-5) :

احسب التباين للبيانات التالية التي تمثل توزيع 200 عامل على حسب الوقت الإضافي الذي يقضونه في العمل في المصنع شهرياً :

الوقت الإضافي (بالساعات)	عدد العاملين
0 إلى أقل من 10	20
10 إلى أقل من 20	80
20 إلى أقل من 30	50
30 إلى أقل من 40	40
40 إلى أقل من 50	10
المجموع	**200**

الحل :

جدول **(5 – 1)** يوضح العمليات الحسابية اللازمة لحساب التباين :

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} = \frac{4400}{200} = 22$$

جدول (5-1)

$(x_i - \overline{x})^2 f_i$	$(x_i - \overline{x})^2$	$(x_i - \overline{x})$	$x_i f_i$	مركز الفئة x_i	التكرار f_i	الفئة
5780	289	17-	100	5	20	0 إلى أقل من 10
3920	49	7-	1200	15	80	10 إلى أقل من 20
450	9	3	1250	25	50	20 إلى أقل من 30
6760	169	13	1400	35	40	30 إلى أقل من 40
5290	529	23	450	45	10	40 إلى أقل من 50
22200			4400		200	المجموع

التباين هو :

$$S^2 = \frac{\sum_{i=1}^{k} f_i \,(x_i - \overline{x})^2}{\sum_{i=1}^{k} f_i - 1}$$

$$= \frac{22200}{200-1} = \frac{22200}{199} = 111.56$$ (ساعة)^2

ونــلاحظ أن التباين وحداته هي مربع وحدات القياس الأصلية، وكثيراً ما تكون غير ذات معنى فمثلاً في هذا المثال التباين = **111.56** ساعة تربيع ، فهنا ساعة تربيع ليس لها أي معنى، وهذا هو عيب التباين ؛ ولذلك نُرجع الوحدات إلى أصلها بأخذ الجذر التربيعي للتباين ، ويسمى المقياس الجديد بالانحراف المعياري .

(3-5) الانحراف المعياري :

هو الجذر التربيعي الموجب للوسط الحسابي لمربعات انحرافات القيم عن وسطها الحسابي ، أي: هو الجذر التربيعي الموجب للتباين،ويرمز له بالرمز S ويحسب كالآتي :

أ – في حـالة البيانات غير المبوبة :

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

ب – في حالة البيانات المبوبة

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

ففي المثال **(5 – 5)** نجد أن الانحراف المعياري = $S = \sqrt{3.43} = 1.85$

وفي المثال **(5 – 6)** نجد أن الانحراف المعياري = $S = \sqrt{111.56} = 10.56$

إذا لم يكن الوسط الحسابي عدداً صحيحاً فإن حساب التباين ومن ثم الانحراف المعياري باستخدام الصيغ السابقة الذكر ، يصبح أمراً غير سهلٍ ، ولذلك اشتقت من الصيغة الأساسية للتباين والتي تعتمد على الانحرافات عن الوسط الحسابي ، صيغة أخرى تعتمد على القيم مباشرة ، وذلك لتسهيل العمليات الحسابية ، وبالطبع الصيغتان تعطيان نفس النتيجة تماماً ، وصيغ التباين التي تعتمد على القيم مباشرة هي :

أ – في حالة البيانات غير المبوبة :

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

ب – في حالة البيانات المبوبة :

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

حيث : X_i : مركز الفئة ، f_i : تكرار الفئة

يجب الانتباه بأن هناك فرقاً بين المقدار $\sum_{i=1}^{k} x_i^2 f_i$ والمقدار $\left(\sum_{i=1}^{k} x_i f_i\right)^2$.

مثال (5-7) :

احسب التباين والانحراف المعياري للبيانات المذكورة في مثال **(5 - 5)** ، وذلك باستخدام الصيغة الثانية للتباين (صيغة القيم مباشرة).

الحل :

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

الحسابات اللازمة لتطبيق هذه الصيغة موضحة فيما يلي :

$\sum X_i=56$	6	5	5	8	9	10	6	7	X_i
$\sum X_i^2=416$	36	25	25	64	81	100	36	49	X_i^2

التباين هو:

$$S^2 = \frac{1}{8-1}\left[416 - \frac{(56)^2}{8}\right]$$

$$= \frac{1}{7}\left[416 - 392\right] = \frac{24}{7} = 3.43$$

من ذلك تكون قيمة الانحراف المعياري هي : $S = \sqrt{3.43} = 1.85$

نلاحظ أنها نفس النتيجة التي تحصلنا عليها في مثال **(5 – 5)** تماماً.

مثال (5-8) :

احسب التباين والانحراف المعياري للبيانات المذكورة في مثال **(5 – 6)** وذلك باستخدام الصيغة الثانية للتباين **(صيغة القيم مباشرة)**

الحل :

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

الحسابات اللازمة لتطبيق هذه الصيغة موضحة في جدول**(5– 2).**

جدول (5 – 2)

الفئة	**التكرار** f_i	**مركز الفئة** (X_i)	$f_i X_i$	X_i^2	$f_i X_i^2$
0 إلى أقل من 10	20	5	100	25	500
10 إلى أقل من 20	80	15	1200	225	18000
20 إلى أقل من 30	50	25	1250	625	31250
30 إلى أقل من 40	40	35	1400	1225	49000
40 إلى أقل من 50	10	45	450	2025	20250
المجموع	200		4400		119000

التباين هو :

$$S^2 = \frac{1}{200-1}\left[119000 - \frac{(4400)^2}{200}\right]$$

$$= \frac{1}{199}\left[119000 - 96800\right] = 111.56$$

ومن ذلك نجد قيمة الانحراف المعياريهي :

$$S = \sqrt{111.56} = 10.56$$

نلاحظ أنها نفس النتيجة التي تحصلنا عليها في مثال **(6-5)** تماماً .

خواص الانحراف المعياري :

1- الانحراف المعياري هو أهم مقاييس التشتت وأكثرها استخداماً .
2- تدخل في حسابه جميع القيم المشاهدة دون إهمال أي قيمة .
3- يتميز بقابليته للمعالجات الجبرية .
4- لا يمكن حسابه للتوزيعات التكرارية المفتوحة .

(4-5) معامل الاختلاف :

كل مقاييس التشتت السابقة تعتمد على وحدات القياس، وبالتالي لا يمكن استعمالها لمقارنة تشتت توزيعين مختلفين في وحدات القياس كمقارنة تشتت الأطوال بتشتت الأوزان مثلاً ، ولذلك يجب التعامل مع مقياس نسبي لا يعتمد على الوحدات المستعملة ويسمى هذا المقياس معامل الاختلاف ويحسب كما يلي :

$$\text{معامل الاختلاف} = \frac{\text{الانحراف المعياري}}{\text{الوسط الحسابي}} \times 100$$

وبالطبع كلما زاد التشتت أخذ معامل الاختلاف نسبة أكبر .

مثال (5-9) :

قارن بين تشتت أطوال الطلبة وأوزانهم ، وذلك باستخدام البيانات التالية ، التي تمثل أطوال وأوزان **100** طالب .

الأطوال بالسنتيمتر	عدد الطلبة
110 إلى أقل من 120	12
120 إلى أقل من 130	15
130 إلى أقل من 140	25
140 إلى أقل من 150	30
150 إلى أقل من 160	10
160 إلى أقل من 170	8

الأوزان (كجم)	عدد الطلبة
60 إلى أقل من 63	6
63 إلى أقل من 66	16
66 إلى أقل من 69	40
69 إلى أقل من 72	28
72 إلى أقل من 75	10

الحل :

نحسب معامل الاختلاف لكل من التوزيعين ثم نقوم بالمقارنة ، ولحساب معامل الاختلاف لكل توزيع يلزمنا حساب الوسط الحسابي والانحراف المعياري لكل منهما ، وذلك كما يلي :

أولاً – توزيع الأطوال :

نوضح العمليات الحسابية التي تلزمنا لحساب الوسط الحسابي والانحراف المعياري في الجدول التالي .

جدول (5 – 3)

الفئة	التكرار f_i	مركز الفئة (X_i)	$f_i X_i$	$X_i - \overline{X}$	$f_i (X_i - \overline{X})^2$
110 إلى أقل من 120	12	115	1380	-23.5	6627.00
120 إلى أقل من 130	15	125	1875	-13.5	2733.75
130 إلى أقل من 140	25	135	3375	-3.5	306.25
140 إلى أقل من 150	30	145	4350	6.5	1267.50
150 إلى أقل من 160	10	155	1550	16.5	2722.50
160 إلى أقل من 170	8	165	1320	26.5	5618.00
المجموع	100		13850		19275.00

الوسط الحسابي هو :

$$\overline{X} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} = \frac{13850}{100} = 138.50$$

قيمة التباين هي :

$$S^2 = \frac{1}{\sum_{i=1}^{k} f_i - 1}\left[\sum_{i=1}^{k} f_i \, (x_i - \overline{x})^2\right]$$

$$= \frac{1}{100-1} \, 19275 = 194.7$$

وتكون قيمة الانحراف المعياري هي :

$$S = \sqrt{194.7} = 13.95$$

$$\text{معامل الاختلاف} = \frac{\text{الانحراف المعياري}}{\text{الوسط الحسابي}} \times 100$$

$$\text{معامل الاختلاف للأطوال} = \frac{13.95}{138.5} \times 100 = 10.07\ \%$$

ثانياً – توزيع الأوزان :

الجدول التالي يوضح لنا العمليات الحسابية اللازمة للحصول على الوسط الحسابي والانحراف المعياري للأوزان :

جدول (5 – 4)

الفئة	التكرار f_i	مركز الفئة (X_i)	$f_i X_i$	$x_i - \overline{x}$	$(x_i - \overline{x})^2 f_i$
60 إلى أقل من 63	6	61.5	369	- 6.6	261.36
63 إلى أقل من 66	16	64.5	1032	- 3.6	207.36
66 إلى أقل من 69	40	67.5	2700	- 0.6	14.4
69 إلى أقل من 72	28	70.5	1974	2.4	161.28
72 إلى أقل من 75	10	73.5	735	5.4	291.6
المجموع	100		6810		936

قيمة الوسط الحسابي هي :

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} = 68.1$$

قيمة التباين هي :

$$S^2 = \frac{1}{\sum_{i=1}^{k} f_i - 1} \left[\sum_{i=1}^{k} f_i \, (x_i - \overline{x})^2 \right]$$

$$= \frac{936}{99} = 9.45$$

وتكون قيمة الانحراف المعياري هي :

$$S = \sqrt{9.45} = 3.07$$

$$\text{معامل الاختلاف} = \frac{\text{الانحراف المعياري}}{\text{الوسط الحسابي}} \times 100$$

$$\text{معامل الاختلاف للأوزان} = \frac{3.07}{68.1} \times 100 = \%4.51$$

بما أن معامل الاختلاف للأطوال أكبر من معامل الاختلاف للأوزان ، إذن تشتت الأطوال أكبر من تشتت الأوزان .

تمارين (5)

1- عرف خاصية التشتت ، مع ذكر أهم مقاييسها .

2- إذا كان لدينا مجموعتان من البيانات :

المجموعة الأولى : 21 15 23 16 14 12 25

المجموعة الثانية : 30 42 64 60 28 70 82 56

أ – لكل مجموعة من هاتين المجموعتين ، احسب ما يلي :

المدى ، التباين، الانحراف المعياري .

ب – قارن بين تشتت المجموعتين باستخدام معامل الاختلاف .

3- يوضح الجدول التالي التوزيع التكراري لدرجات 100 طالب :

الدرجة	19 - 0	39 - 20	59 - 40	79 - 60	99 - 80
عدد الطلبة	5	15	35	40	5

فاحسب :

أ – المدى . ج – الانحراف المعياري .

ب – التباين . د – معامل الاختلاف .

4- يوضح الجدول التكراري التالي توزيع 40 عاملاً على حسب عدد أيام الغياب فيها خلال سنة .

عدد أيام الغياب	**عدد العاملين**
2 – 0	10
5 – 3	15
8 – 6	8
11 – 9	5
14-12	2

فاحسب :

أ – المدى . ب – التباين .

ج – الانحراف المعياري . د – معامل الاختلاف .

5- الجدول التالي يبين أوزان 50 طالباً بالكيلو جرامات :

الوزن (بالكيلو جرام)	عدد الطلبة
55 إلى أقل من 65	8
65 إلى أقل من 75	10
75 إلى أقل من 85	20
85 إلى أقل من 95	8
95 إلى أقل من 105	4

فاحسب :

أ – المدى ب – التباين.

ج – الانحراف المعياري . د – معامل الاختلاف .

6- يوضح جدول التوزيع التكراري التالي توزيع 40 طالباً على حسب عدد الساعات التي يقضيها الطالب في المذاكرة شهرياً .

ساعات المذاكرة	عدد الطلبة
24 إلى أقل من 40	3
40 إلى أقل من 56	5
56 إلى أقل من 72	10
72 إلى أقل من 88	12
88 إلى أقل من 104	5
104 إلى أقل من 120	5

فاحسب :

أ – المدى ب – التباين .

ج – الانحراف المعياري . د – معامل الاختلاف .

7- إذا كانت القيم التالية مرتبة ترتيباً تصاعدياً :

x_1 ، 30 ، 25 ، 22 ، x_5

وعلمت أن المدى = 18 ، والوسط الحسابي = 27 ، فما قيمة x_5 ، x_1 .

8- علّل ما يلي :

- لا نستطيع حساب المدى للجداول التكرارية المفتوحة .
- يعتبر المدى مقياساً مضللاً في حالة وجود قيم متطرفة .
- التباين لا يمكن أن يكون مقدراً سالباً .
- الانحراف المعياري أكثر استعمالاً في التحليل الإحصائي .
- لا نستطيع مقارنة تشتت الأطوال والأوزان باستخدام الانحراف المعياري .

9- إذا علمت أن معامل الاختلاف = 25% وأن التباين =25 ، وذلك للبيانات التالية :

x، 15 ، 25 ، 18 ، 10

فاحسب القيمة المجهولة x .

10- توضح البيانات التالية التوزيعين التكراريين لمائة عامل على حسب مرتباتهم وعلى حسب عدد أطفالهم . فقارن بين تشتت هذين التوزيعين .

المرتبات (بالدينار)	عدد العاملين
100 إلى أقل من 150	10
150 إلى أقل من 200	20
200 إلى أقل من 250	30
250 إلى أقل من 300	25
300 إلى أقل من 350	15

عدد الأطفال	عدد العاملين
2 – 0	15
5 – 3	25
8 – 6	35
11 – 9	20
14-12	5

الفصل السادس

الارتبــاط والانحــدار

1.6 الارتباط :

الارتباط يبحث في العلاقة بين المتغيرات (الظواهر) من حيث قوتها واتجاهها ، وبالتالي فالارتباط معيار لقياس قوة واتجاه العلاقة بين متغيرين أو أكثر . ووفقا لما تقدم فإن الارتباط قد يكون قويا أو ضعيفا أو معدوما تبعا للعلاقة ، وأيضا قد يكون الارتباط موجبا (طرديا) أو سالبا (عكسيا) حسب اتجاه العلاقة بين المتغيرات.

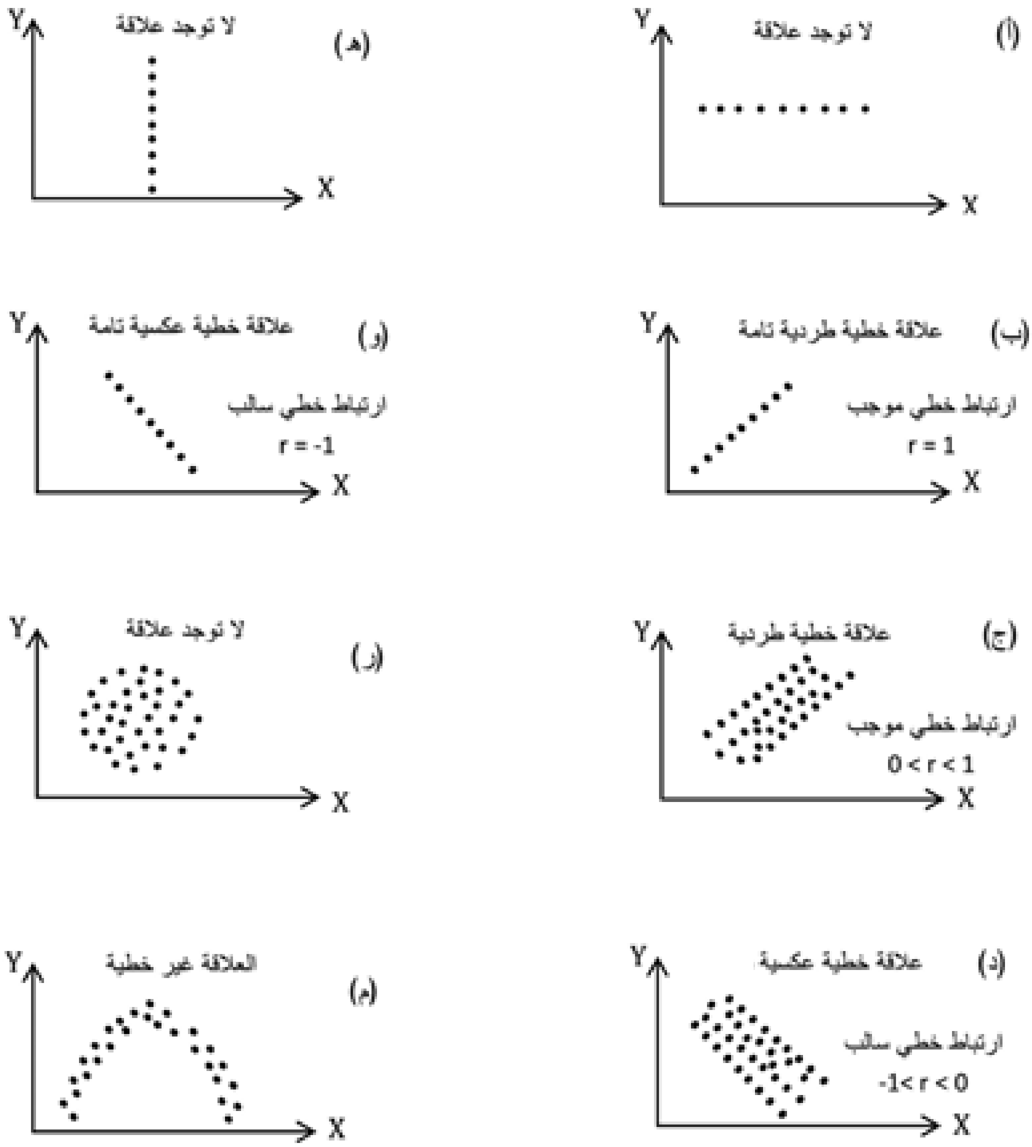
1.1.6 الارتباط البسيط :

وهو يبحث في قوة واتجاه العلاقة بين متغيرين (ظاهرتين) فقط مثل الطول والوزن أو درجات مقررين أو عدد الساعات التي يقضيها طالب في المذاكرة والدرجة المتحصل عليها في الامتحان النهائي . والارتباط البسيط قد يكون خطيا "بسيطا" حيث يفترض أن العلاقة بين المتغيرين خطية ، أو غير خطية بفرض أن العلاقة بين المتغيرين غير خطية .

والارتباط البسيط سيكون طرديا (موجبا) إذا تغير أحد المتغيرين (الظاهرتين) في اتجاه ما أدى إلى تغير المتغير الآخر (الظاهرة الأخرى) في الاتجاه نفسه مثل الطول والوزن للأطفال، كمية الأملاح بالجسم وضغط الدم ، كمية الإنتاج وعدد العمال ، ويكون هذا الارتباط عكسيا (سالبا) إذا تغير أحد المتغيرين في اتجاه ما أدى إلى تغير المتغير الآخر في الاتجاه المخالف مثل العرض والسعر لسلعة ما ، عدد أيام الغياب والدرجات بالامتحان .

فالارتباط البسيط يكون خطيا إذا كنت العلاقة التي تربط المتغيرين خطية (في صورة خط مستقيم) .

إن الارتباط بين متغيرين من حيث القوة والضعف، طردي أوعكسي خطي أوغير خطي، يمكن الاستدلال عليه باستخدام الأشكال الانتشارية ، والشكل الانتشاري يبين انتشار قيم المتغيرين على محوري الإحداثيات، ومن خلال تفحص هذا الشكل يمكن الحصول على فكرة عامة حول العلاقة ونوعيتها وبالتالي حول الارتباط بين المتغيرين ، وهناك أنواع مختلفة من الأشكال الانتشارية من أمثلتها:


(أ)
لا توجد علاقة
Y
X
(هـ)
لا توجد علاقة
Y
X
(ب)
علاقة خطية طردية تامة
ارتباط خطي موجب
r = 1
Y
X
(و)
علاقة خطية عكسية تامة
ارتباط خطي سالب
r = -1
Y
X
(ج)
علاقة خطية طردية
ارتباط خطي موجب
0 < r < 1
Y
X
(ز)
لا توجد علاقة
Y
X
(د)
علاقة خطية عكسية
ارتباط خطي سالب
-1< r < 0
Y
X
(م)
العلاقة غير خطية
Y
X

2.1.6 الارتباط الخطي البسيط :

إن الارتباط الخطي البسيط هو أكثر أنواع الارتباط استخداما وذلك نظراً لأن معظم العلاقات غير الخطية بين متغيرين يمكن تقريبها بشكل فرضي إلى علاقة خطية ، وعلى وجه العموم الارتباط الخطي البسيط يقيس قوة واتجاه العلاقة الخطية بين المتغيرين ، وعادة ما يقاس بما يسمى بمعامل الارتباط ويرمز له بالـــرمز (r) وهناك عدة أنواع لمعاملات الارتباط تختلف باختلاف الظواهر، ومن أشهر هذه المعاملات للظواهر الكمية :

أ. معامل ارتباط بيرسون :

ويعرّف على أنه متوسط حاصل ضرب القيم المعيارية للظاهرتين ومن خواص هذا المعامل :

1. إذا كانت العلاقة الخطية بين المتغيرين منعدمة فإن معامل الارتباط = 0 .
2. إذا كانت العلاقة الخطية بين المتغيرين علاقة طردية تامة فإن معامل الارتباط = 1، وإذا كانت عكسية تامة فإن معامل الارتباط = 1- .
3. تتراوح قيمة معامل الارتباط بين 1+ ، و 1- .
4. كلما اقترب معامل الارتباط من الصفر دل على ضعف العلاقة بينهما، وكلما اقترب من ∓ 1 دل على قوة العلاقة .

1) معامل الارتباط الخطي البسيط (معامل ارتباط بيرسون) :

لقد تعرضنا لهذا النوع من البيانات فيما سبق ، فإذا كان المتغيران هما (Y , X) فإن معامل ارتباط بيرسون يرمز له بالرمز r ويعرّف كالآتي :

$$r = \frac{1}{n}\left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{s_x\, s_y}\right]$$

حيث $\bar{x}$ متوسط الظاهرة x ، $\bar{y}$ متوسط الظاهرة y ، s_x الانحراف المعياري للظاهرة X
s_y الانحراف المعياري للظاهرة Y .

وحساب معامل الارتباط باستخدام هذه العلاقة غالبا ما يكون صعب الحساب ، وعليه فقد وجدت صيغة أخرى أكثر سهولة في التطبيق وهي :

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

وهذه الصيغة يطلق عليها الصيغة المباشرة ، ويمكن توضيح استخدامها لبيانات المثال التالي :

مثال (6-1) :

أوجد معامل الارتباط بين المتغيرين X , y مبينا نوعه ودرجته حسب بيانات الجدول التالي وذلك باستخدام الصيغة المباشرة :

52	51	50	49	48	X
31	29	33	29	28	y

الحلّ :

لإيجاد معامل الارتباط بالصيغة المباشرة ، نكوّن الجدول التالي :

y^2	x^2	x y	y	x	ر.م
784	2304	1344	28	48	1
841	2401	1421	29	49	2
1089	2500	1650	33	50	3
841	2601	1479	29	51	4
961	2704	1612	31	52	5
4516	12510	7506	150	250	**المجموع**

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

$$= \frac{5(7506) - (150)\ (250)}{\sqrt{[\ 5(12510) - (250)^2\ ]\ [\ 5(4516) - (150\ )^2]}}$$

$$= \frac{30}{\sqrt{(50)\ (80)}} = \frac{30}{63.25} = 0.47$$

وهذا يعنى أن الارتباط طردي ضعيف .

مثال (6-2) :

أوجد معامل الارتباط بين المتغيرين X , y مبينا نوعه ودرجته حسب بيانات الجدول التالي :

5	4	3	2	1	X
2	3	1	1	3	y

الحلّ :

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

لتحديد معامل الارتباط نكوّن الجدول الآتي :

y^2	X^2	x y	y	x	ر.م
9	1	3	3	1	1
1	4	2	1	2	2
1	9	3	1	3	3
9	16	12	3	4	4
4	25	10	2	5	5
24	55	30	10	15	**المجموع**

$$r = \frac{(5)\,(30) - (15)\,(10)}{\sqrt{[\,5\,(55) - (15)^2\,]\;[\,5(24) - (10)^2]}}$$

$$= \frac{150 - 150}{\sqrt{[\;(275) - (225)\;]\;[\;(120) - (100\;)]}} = \frac{0}{\sqrt{1000}} = 0$$

وهذا يعني أن الارتباط منعدم بين المتغيرين.

ب. معامل ارتباط الرتب لسبيرمان

لقد درسنا معامل ارتباط بيرسون والذي يستخدم لتحديد قوة واتجاه العلاقة بين ظاهرتين كميتين ، وحيث أن الظواهر قد تكون كمية أو وصفية . فإذا أردنا دراسة العلاقة بين ظاهرتين وصفيتين فإنه لا يمكن استخدام معامل ارتباط بيرسون ، وعليه فقد وجدت طريقة أخرى يمكن استخدامها في مثل هذه الحالات وهي تعتمد على رتب الظاهرتين ، وبالتالي فإن ناتج هذه الطريقة يسمى بمعامل ارتباط الرتب لسبيرمان وسوف نرمز له بالرمز (r) حيث :

$$r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

حيث n عدد الرتب (البيانات المزدوجة) , d_i هو الفرق بين رتبتين متناظرتين ، $\sum_{i=1}^{n} d_i^2$ هو مجموع مربعات تلك الفروق ، وسوف نتعرض لطريقة حسابه والتي تكون على النحو التالي :

أولا : إذا كانت البيانات مفردة :

لحساب معامل ارتباط الرتب يجب اتّباع الخطوات التالية :

1. إعطاء رتب لكل مفردة من مفردات الظاهرتين (قيم أو صفات) وبالطريقة نفسها (تصاعديا معا أو تنازليا معا) .

2. نجد الفرق بين رتب الظاهرتين للمفردة نفسها ، أي نجد d_i ، ثم نجد مربع الفرق (d_i^2) ثم نجد مجموع مربعات الفروق ($\sum_{i=1}^{n} d_i^2$) .

3. نستخدم القانون $r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$ وذلك لتحديد معامل ارتباط الرتب .

مثال (3-6) :

إذا كانت تقديرات 5 طلبة في مادتين كما هو مبين أدناه ، فاحسب معامل الارتباط بين هاتين الظاهرتين .

الطالب	1	2	3	4	5
تقدير المادة الأولى	جيد	ممتاز	جيد جداً	مقبول	ضعيف
تقدير المادة الثانية	مقبول	جيد جداً	جيد	ممتاز	ضعيف

الحل :

لاحظ أن البيانات وصفية ، وبالتالي معامل الارتباط المناسب لها هو معامل ارتباط الرتب على النحو التالي :

الطالب	تقدير المادة الأولى	تقدير المادة الثانية	رتب المادة الأولى	رتب المادة الثانية	الفروق d	مربعات الفروق d^2
1	جيد	مقبول	3	2	1	$1^2=1$
2	ممتاز	جيد جدا	5	4	1	$1^2=1$
3	جيد جدا	جيد	4	3	1	$1^2=1$
4	مقبول	ممتاز	2	5	-3	$-3^2=9$
5	ضعيف	ضعيف	1	1	0	$0^2=0$
المجموع			15	15	0	12

وعليه :

$$r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6\,(12)}{5(25-1)} = 1 - \frac{6(12)}{5(24)} = 1 - \frac{3}{5} = 0.4$$

وهو ارتباط طردي ضعيف.

مثال (4-6):

البيانات التالية تمثل درجات 8 طلبة في مقررين والمطلوب تحديد معامل ارتباط الرتب

المقرر	1	2	3	4	5	6	7	8
درجة المقرر الأول	30	17	35	28	42	25	19	34
درجة المقرر الثاني	35	31	40	46	50	32	33	42

الحل :

المطلوب هو تحديد معامل ارتباط الرتب ، وبالتالي يجب تحديد رتب القيم ، كما هو بالجدول التالي :

الطالب	درجة المقرر الأول	درجة المقرر الثاني	رتب درجات المقررالأولى	رتب درجات المقررالثاني	الفروق d	مربعات الفروق d^2
1	30	35	5	4	1	$1^2=1$
2	17	31	1	1	0	$0^2=0$
3	35	40	7	5	2	$2^2=4$
4	28	46	4	7	-3	$-3^2=9$
5	42	50	8	8	0	$0^2=0$
6	25	32	3	2	1	$1^2=1$
7	19	33	2	3	-1	$(-1)^2=1$
8	34	42	6	6	0	$0^2=0$
المجموع			36	36	0	16

$$r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6(16)}{8(64-1)} = 1 - \frac{6(16)}{8(63)} = 1 - 0.19 = 0.81$$

وهو ارتباط قوي ، أي هناك علاقة قوية طردية بين رتب درجات المقررين .

ثانياً : إذا كانت البيانات مكررة :

إن حساب معامل ارتباط الرتب مبنى على أساس أن يكون لكل صفة رتبة ولكل رتبة صفة ، فإذا وجد تكرار لبعض الصفات فإنه لا يمكن حساب معامل ارتباط الرتب إلا بعد تحديد الصفات المكررة ثم تعطى رتباً (تصاعدية أو تنازلية) على اعتبار أنها مختلفة ، وبحساب المتوسط الحسابي لرتب الصفات المتكررة والذي يمكن اعتباره الرتبة الموحدة لتلك البيانات (الصفات) المكررة . ثم نتبع الخطوات السابقة لحساب معامل ارتباط الرتب لسبيرمان وذلك باستخدام الرتب المعدلة ، ولتوضيح ذلك نتبع المثال التالي :

مثال (6-5):

ذكر أحد المختصين بأنّ الارتباط بين عدد الأطفال والمستوى التعليمي لرب الأسرة ارتباط طردي ، فهل تؤيد رأيه بناءً على البيانات الآتية :

الأسرة	1	2	3	4	5	6	7	8	9
عدد الأطفال	4	3	2	7	3	4	7	3	5
المستوى التعليمي	يقرأ ويكتب	شهادة عليا	شهادة متوسطة	يقرأ ويكتب	شهادة متوسطة	شهادة متوسطة	أمي	شهادة متوسطة	أمي

لاحظ أنه لا يمكن حساب معامل ارتباط الرتب لسبيرمان لوجود بيانات مكررة إلا بعد تعديل الرتب للصفات المكررة أي الحصول على الرتب المعدلة وذلك باتّباع الآتي :

كيفية حساب الرتب المعدلة للبيانات المكررة (النجوم تبين الصفات المتكررة)

أولاً (لعدد الأطفال)

1. أسر لديها 4 أطفال مكررة مرتين، الرتب المعدلة هي $\frac{(6+5)}{2} = 5.5$.

2. أسر لديها 3 أطفال مكررة ثلاث مرات، الرتب المعدلة هي $\frac{(4+3+2)}{3} = 3$.

3 .أسر لديها طفلان (غير متكرر)، أي: أسرة واحدة تأخذ رتبتها الأصلية وهي 1 .

4. أسر لديها 7 أطفال مكررة مرتين، الرتب المعدلة هي $\frac{(9+8)}{2} = 8.5$.

ثانياً : (للمستوى التعليمي)

1. يقرأ ويكتب مكررة مرتين، الرتب المعدلة هي : $\frac{(4+3)}{2} = 3.5$.

2. شهادة عليا (غير متكرر)، أي: أسرة واحدة تأخذ رتبتها الأصلية وهي 9 .

3. شهادة متوسطة مكرره أربع مرات، الرتب المعدلة: $\frac{(8+7+6+5)}{4} = 6.5$.

4. أمي مكررة مرتين ،الرتب المعدلة هي: $\frac{(2+1)}{2} = 1.5$

عدد الأطفال	المستوى التعليمي	رتب عدد الأطفال	رتب المستوى التعليمي	الرتب المعدلة لعدد الأطفال	الرتب المعدلة للمستوى التعليمي	الفروق d	مربعات الفروق d^2
4**	**يقرأ ويكتب	5	3	5.5	3.5	2	$2^2=4$
3*	شهادة عليا	2	9	3	9	-6	$-6^2=36$
2	***شهادة متوسطة	1	5	1	6.5	-5.5	$-5.5^2=30.25$
7***	**يقرأ ويكتب	8	4	8.5	3.5	5	25
3*	***شهادة متوسطة	3	6	3	6.5	-3.5	$-3.5^2=12.25$
4**	***شهادة متوسطة	6	7	5.5	6.5	-1	$-1^2=1$
7***	*أمي	9	1	8.5	1.5	7	$-7^2=49$
3*	***شهادة متوسطة	4	8	3	6.5	-3.5	$-3.5^2=12.25$
5	*أمي	7	2	7	1.5	5.5	30.25
المجموع		45	45	45	45	0	200

$$r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6(200)}{9(81-1)} = 1 - \frac{6(200)}{9(80)} = 1 - 1.67 = -0.67$$

ونلاحظ بأن الارتباط عكسي قوي وهذا لا يؤيد رأي الأختصاصي

2.6 الانحدار :

عند دراستنا لمعامل الارتباط ذكرنا أنه مقياس لقوة واتجاه العلاقة بين المتغيرات (الظواهر) المدروسة ، إلا أنه لا يتعرض إلى تحديد هذه العلاقة . وعليه وجد موضوع آخر يبحث في كيفية الحصول على الصيغة الرياضية للعلاقة التقديرية بين الظواهر، وهو ما يسمّى بالانحدار . فالانحدار هو الدراسة الخاصة بتحديد العلاقة بين المتغيرات . إن تحديد الصيغة الرياضية للعلاقة بين الظواهر إن وجدت تساعد في تقدير قيم بعض الظواهر إذا عرفت قيم الظواهر الأخرى ، وعند دراسة الانحدار يجب التمييز بين المتغيرات (الظواهر) المستقلة والمتغيرات (الظواهر) التابعة .

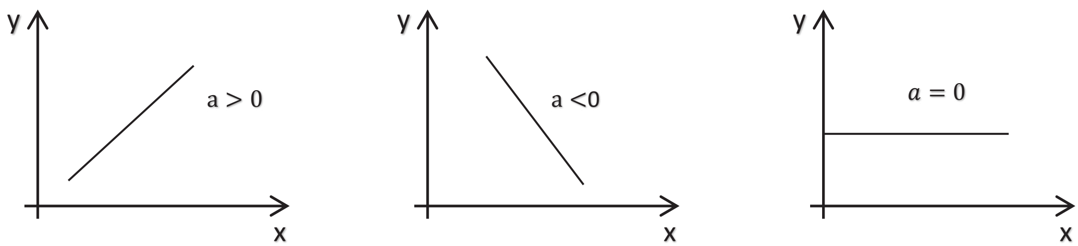
إن العلاقة بين المتغيرات يمكن أن تأخذ أشكالا متعددة وفقا لعدد المتغيرات من جهة ودرجة أو نوع العلاقة من جهة أخرى ، ومن أبسط هذه الصور هي العلاقة الخطية من الدرجة الأولى، وهي علاقة بين متغيرين فقط أحدهما تابع والآخر مستقل، ويسمى الانحدار في هذه الحالة بالانحدار الخطي البسيط .

وقد تكون العلاقة بين المتغيرين غير خطية (من درجة أعلى من الدرجة الأولى ،أو آسية) وفي هذه الحالة يسمى بالانحدار غير الخطي . وأما إذا كنا نبحث في العلاقة بين أكثر من متغيرين فإننا نتحدث على الانحدار المتعدد .

وعادة ما يستخدم الشكل الانتشاري في تحديد صورة العلاقة التقريبية ، فبواسطته يمكن اختيار المعادلة الرياضية المناسبة بين متغيرين من المعادلات المختلفة المتاحة (كمعادلة الخط المستقيم ، معادلة المنحنى الآسي معادلة القطع المكافئ ،...) لذلك يفضل من الناحية العملية رسم الشكل الانتشاري قبل الخوض في عملية التحليل، ومن بين المعادلات المختلفة المتاحة مايلي :

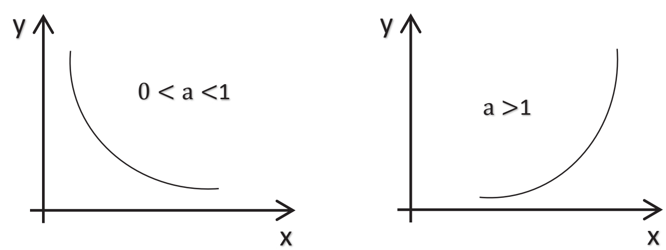
1 - معادلة الخط المستقيم:

$$Y = a\,x + b$$



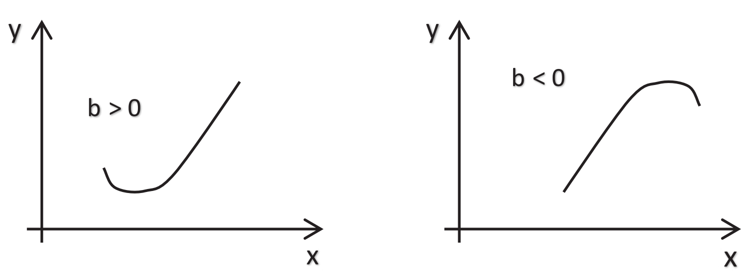
$$Y = a^{x}\,b$$

2 – معادلة المنحنى الأسي:



3 - معادلة منحنى الدرجة الثانية:

$$Y = a\,x^{2} + b\,x + c$$



ونظراً لسهولة التعامل مع الانحدار الخطي البسيط من جهة، وإن معظم العلاقات تربط بين متغيرين يمكن تقريبهما بشكل فرضي بعلاقة خطية بسيطة من جهة أخرى، لذا فسوف نتعرض للانحدار الخطي البسيط .

1.2.6 الانحدار الخطي البسيط :

الانحدار الخطي البسيط يهتم بدراسة العلاقة بين متغيرين أحدهما مستقل (X) والآخر تابع (y) مع افتراض أن العلاقة المناسبة بينهما في الصورة الخطية $y=ax+b$ حيث b , a ثابتان يجب تحديدها ، وبالتالي يتحدد شكل العلاقة التقديرية في صورة تامة . وهناك عدة طرق لتحديد هذه العلاقة منها التمهيد باليد (أو بمجرد النظر) وطريقة المربعات الصغرى . فعند اتّباع الطريقة الأولى يستخدم الشكل الانتشاري للمساعدة في تحديد العلاقة المناسبة بين المتغيرين . إلا أن هذه الطريقة لا يمكن الاعتماد عليها؛ لأنها تعطي علاقات مختلفة للبيانات نفسها باختلاف الباحثين ومهاراتهم .

وللخروج من هذه المشكلة وجدت طريقة المربعات الصغرى لتحديد العلاقة الخطية بين المتغيرين والتي سوف نتعرض لها باختصار .

طريقة المربعات الصغرى :

عند اتّباع طريقة التمهيد باليد لتحديد الخط المستقيم الذي يناسب بيانات الشكل الانتشاري أمر يتوقف على الخبرة والمهارة وبالتالي فإن هذا قد يختلف من شخص إلى آخر . ولذلك وجدت طريقة إحصائية يمكن استخدامها لتحديد أفضل خط مستقيم يمكن استخدامه كعلاقة بين متغيرين وهذه الطريقة تسمى طريقة المربعات الصغرى . وهي على النحو التالي:

$\hat{y} = \hat{a}\,X + \hat{b}$ وقد سميت هذه العلاقة بمعادلة انحدار y على X التقديرية .

حيث $\hat{y}$ القيمة التقديرية للمتغير التابع y و $\hat{a}$ القيمة التقديرية للمعلمة a (ميل معادلة انحدار y على x أى مقدار التغير في y عندما تتغير x بمقدار وحدة واحدة) و $\hat{b}$ القيمة التقديرية للمعلمة b (قيمة y عندما $x=0$ أي الجزء المقطوع من المحور الرأسي)

$$\hat{a} = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

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

حيث :

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

مثال (6-6) :

البيانات التالية تمثل الدخل والإنفاق الشهري بالدينار لعشر أسر .

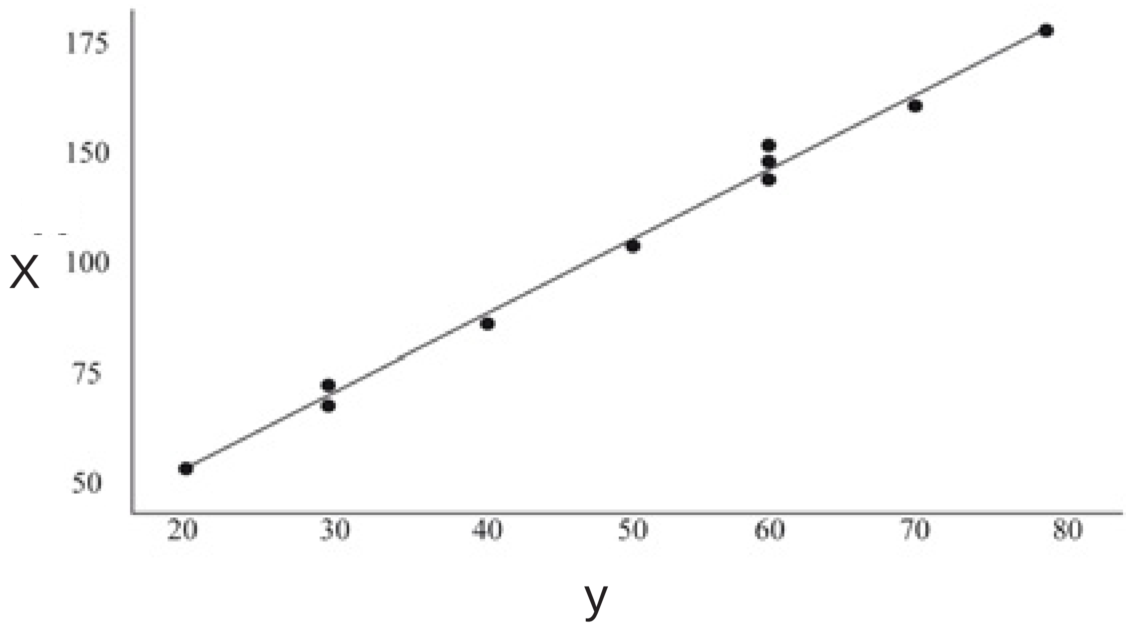
الدخل (X)	73	50	128	170	87	108	135	69	148	132
الإنفاق (y)	30	20	60	80	40	50	60	30	70	60

والمطلوب :

1. تحديد العلاقة المناسبة من خلال الشكل الانتشاري .
2. تقدير معادلة انحدار الإنفاق (y) على الدخل (X) المقدرة باستخدام طريقة المربعات الصغرى .
3. تحديد (تقدير) الإنفاق عندما يكون الدخل 110 دينارات .
4. تحديد معامل الارتباط بين الدخل والإنفاق .

الحلّ :

لتحديد العلاقة المناسبة بين المتغيرين X , y نرسم الشكل الانتشاري كما يلي :



1. من الشكل نلاحظ أن العلاقة الخطية $\hat{y} = \hat{a}x + \hat{b}$ هي العلاقة بين المتغيرين y، X .

2.لتحديد معاملات العلاقة الخطية $\hat{b}$ ، $\hat{a}$ نتبع الآتي :

ر.م	x	Y	X Y	X^2	Y^2
1	73	30	2190	5329	900
2	50	20	1000	2500	400
3	128	60	7680	16384	3600
4	170	80	13600	28900	6400
5	87	40	3480	7569	1600
6	108	50	5400	11664	2500
7	135	60	8100	18225	3600
8	69	30	2070	4761	900
9	148	70	10360	21904	4900
10	132	60	7920	17424	3600
المجموع	1100	500	61800	134660	28400

$$\hat{a} = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$= \frac{10\,(61800) - (1100)\,(500)}{10\,(134660) - (1100)^2}$$

$$= \frac{618000 - 550000}{1346600 - 1210000} = \frac{68000}{136600} = 0.4978$$

لا حظ أن $\hat{a}$ هي معامل معادلة الانحدار المقدرة . أي أنه كلما تغيرت (زادت) X (الدخل) بدينار واحد تغيرت (زادت) y (الإنفاق) بمقدار 0.4978 دينارًا.

$$\hat{b} = \bar{y} - \hat{a}\,\bar{x}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1100}{10} = 110 \quad ، \quad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{500}{10} = 50$$

$$\hat{b} = 50 - 0.4978\,(110) = 50 - 54.758 = -4.758$$

وحيث أن $\hat{b}$ هي الجزء المقطوع في معادلة الانحدار المقدرة $\hat{y}$ أي أن قيمة y (الإنفاق) المقدرة = 4.758 - عندما يكون الدخل (X) = صفر ، وبالتالي فإن معادلة انحدار y على X التقديرية هي :

$$\hat{y} = 0.4978\,x - 4.758$$

3. الإنفاق عندما يكون الدخل $x = 110$ دنانير

$$\hat{y} = 0.4978(110) - 4.758 = 50$$

4. معامل الارتباط

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

$$= \frac{(10)(61800) - (500)(1100)}{\sqrt{[10(134660) - (1100)^2]\ [10(28400) - (500)^2]}}$$

$$= \frac{618000 - 550000}{\sqrt{(1346600 - 1210000)\ (284000 - 250000)}}$$

$$= \frac{68000}{\sqrt{(136600)\ (34000)}} = \frac{68000}{68149.835} = 0.9978$$

معامل الارتباط قريب من الواحد وفي هذه الحالة الارتباط طردي قوي وهذا دليل على قوة العلاقة الخطية بين الدخل (X) والإنفاق (y) .

تمارين (6)

1. اذكر مثالا لظاهرتين (X,y) بحيث يكون معامل ارتباط بيرسون = 1 ومعامل ارتباط الرتب لسبيرمان = 1 .

2. البيانات التالية تمثل قيما لمتغيرين X . y .

X	2	4	5	6	8	10
y	18	12	10	8	7	5

أوجد :

معامل ارتباط بيرسون .

معامل ارتباط الرتب لسبيرمان . أيهما أفضل مع التعليق ؟

3. إذا كانت $\hat{Y}=0.6X+0.3$ معادلة خط انحدار y على X وكان الوسط الحسابي لقيم x يساوي 7 فما قيمة الوسط الحسابي لقيم y ؟

أ. 2.8 ب. 4.5 جـ. 0.9 د. 0.0

4. أقوى معامل ارتباط خطي بين متغيرين من المعاملات التالية هو

أ. 0.3 ب. -0.8 جـ. 0.9 د. -0.95

5. أوجد معامل الارتباط المناسب للبيانات الآتية والتي تمثل تقديرات أحد الطلبة في خمسة امتحانات لمادتين :

المادة الأولى	ضعيف	ممتاز	مقبول	جيد جدا	جيد
المادة الثانية	ضعيف جدا	جيد جدا	ضعيف	جيد	مقبول

6. اذكر مثالا لظاهرتين يكون فيه :

أ- معامل الارتباط قوياً.

ب- معامل الارتباط ضعيفًا.

ت- معامل ارتباط مساويا الصفر .

7. تقدم 10 أشخاص لشغل وظيفة معينة ، فجرى امتحانهم من قبل لجنتين ، حيث أعطي الأشخاص رتبًا فكانت على النحو التالي :

الشخص	1	2	3	4	5	6	7	8	9	10
رتب اللجنة الأولى	2	5	3	6	4	1	9	8	10	7
رتب اللجنة الثانية	2	5	4	7	9	1	10	6	8	3

فهل هناك علاقة بين رتب اللجنتين ؟

8. تم فحص ضغط الدم ومستوى السكر في الدم لمجموعة من الأفراد فكان على النحو التالي :

ضغط الدم	عادي	منخفض	متوسط	عالي	عادي	منخفض	عالي	متوسط	عادي
مستوى السكر	متوسط	مرتفع	عادي	منخفض	عادي	مرتفع	منخفض	متوسط	متوسط

فهل هناك ارتباط بين الرتب .

9. لقد تعرضنا إلى معامل ارتباط بيرسون ومعامل ارتباط الرتب . فما هو الفرق بينهما من حيث الاستخدام والدقة ، وأيهما أفضل ؟

10. إذا كان معامل ارتباط الرتب لسبيرمان = 1 فما معنى ذلك وهل هذا يعنى أن معامل ارتباط بيرسون = 1؟ ولماذا ؟.

11. في دراسة الارتباط بين المتغيرين y , X تم الحصول على البيانات التالية :

$\sum y^2=84$ ، $\sum x^2=52933$ ، $\sum Xy=1858$

$n=8$ ، $\sum y=22$ ، $\sum x=649$

أوجـد معادلة انحدار y على x

12. الجدول التالي يحتوى العمر (x) وضغط الدم (y) لعشرة أشخاص .

العمر (X)	56	42	72	46	48	56	40	42	68	60
ضغط الدم(Y)	149	125	161	149	128	152	146	140	152	155

فـهل هناك ارتباط :

1) بين القيم ؟ 2) بين الرتب ؟

13. اذكر نوع الارتباط وقوته بين الظاهرتين للحالات الآتية :

أ) عدد الطلبة في الفصل ومستوى الاستيعاب .
ب) السرعة والمسافة المقطوعة عند استعمال الفرامل فجأة .
ج) سنوات الخبرة ومستوى الدقة في العمل .
د) أنواع الأمراض والعمر .
هـ) مستوى الوعي الصحي ودرجة انتشار المرض .
و) مستوى الثقافة والغش .
ز) مستوى الوعي العام وعدد حوادث المرور .
ح) السعر والجودة لسلعة ما.
ط) الجنس والدقة في العمل .
ي) الغياب وكمية الإنتاج .

14. البيانات الآتية عن المتغيرين X , y :

X	3	1	5	2	4
Y	3	2	6	4	5

أوجـد :

أ- الشكل الانتشاري مع التعليق عليه .
ب- معامل الارتباط مع التعليق على الناتج .
ج- معادلة انحدار Y على X .

15. لتحديد العلاقة بين الوزن والطول جمعت البيانات الآتية :

العمر(X)	50	55	60	70	75
الوزن (Y)	150	160	165	160	165

أوجـد :

أ- الشكل الانتشاري مع التعليق عليه .
ب- معامل الارتباط مع التعليق على الناتج .
ج- معادلة انحدار y على x .
د- الوزن عندما يكون العمر 68 سنة .