الفصل الثاني

تلخيص البيانات

Summarizing Data

عند الحصول على مشاهدات عن واحد أو أكثر من المتغيرات التي تم أخذها من مجتمع أو عينه نكون حصلنا على البيانات. وكما نعلم فإن هناك متغيرات كمية ومتغيرات وصفية ومن ثم لدينا صور مختلفة للبيانات. والمتغيرات الكمية عبارة عن نتيجة لقياس معين أو عملية عدّ معينة وبذلك فإن الأرقام التي تم جمعها كمشاهدات لتلك المتغيرات يتم تصنيفها على شكل بيانات نسبية للمشاهدات التي تم الحصول عليها بطريقة القياس في حين أنه يتم تصنيفها كبيانات فترة للمشاهدات التي تم الحصول عليها بطريقة العدّ.

من جهة أخرى فإن المتغيرات الوصفية تعبّر عن صفات معينة للعينة أو المجتمع مثل العجول، والعجول الفحول، والثيران، والأبقار، والتي تعبّر عن بيانات أسمية كونها عبارة عن أسماء يتم استخدامها من قبل الباحثين للتصنيف على شكل مجموعات بناء على الأسماء، أو تعبّر عن بيانات ترتيبية يتم قياسها بمعيار ترتيبي تتكون من مستويات أو فئات حيث يتم ترتيبها إما تصاعدياً وإما تنازلياً من الأفضل للأسوأ أو العكس.

بالإضافة لما ذكر، فإنه يتم تصنيف البيانات بناءً على مصادرها حيث تصنّف البيانات إلى بيانات أولية عند الحصول عليها عن طريق التجارب، أو استمارات الاستبيان، ومصادر النشر التي حصلت على البيانات مثل التعدادات الزراعية في حين تصنف إلى بيانات ثانوية عند الحصول عليها من مصادر لم تقم بجمع البيانات حتى ولو قامت تلك المصادر بنشر تلك البيانات.

إن الهدف من جمع البيانات يتمثل في استخدامها لاتخاذ قرارات تجاه مشكلة معيّنة من خلال معالجة تلك البيانات وتحليلها إحصائيا ومن ثم الحصول على معلومات عن المشكلة والتي يمكن استخدامها في اتخاذ القرار والذي يتوقع إن يكون ذلك القرار أفضل عند مقارنته بحالة غياب تلك المعلومات.

وبصفة عامة يتم جمع البيانات الإحصائية الخام بطريقة عشوائية كما في الجدول رقم (٢,١) حيث لا يوجد ترتيب معين لتلك البيانات ولذلك فإنه من الصعوبة الحصول على أي قيم معينة أو استدلالات من تلك البيانات في صورتها الخام لذا يتم ترتيب البيانات للمساعدة في تفسيرها.

الجدول رقم (٢,١). إنتاجية نسيج القطن الكتاني (باوند/أيكر) لخمسة و سبعون مزرعة في سهول الأراضي السوداء.

٢٥٥	٣٧٣	٢٤٢	٢٥٧	٣٠٥
٢٨٥	٣٥٨	٢٧٩	٢٦١	٣١٢
٢٨٨	٢٩٧	٣٠٣	٢٩٠	٢٨٣
٢١٥	٢٩٩	٢٧٥	٢٦٠	٢٨٤
٣١١	٢١٧	٢٩٥	٢٦٠	٣١٦
٣١٤	٢٩٠	٢٩٤	٢٨٠	٣٢٦
٢٤٩	٢٥٦	٢٧٦	٢٩٢	٢٨٦
٣٣٣	٢٧٤	٢٥٠	٢٩٥	٣١٨

تابع الجدول رقم (٢,١).

٢٩٢	٢٨٣	٢٧٤	٣٣٦	٣٢٥
٢٩٠	٣٠٩	٢٧٢	٢٩٦	٣٤٦
٢٥٤	٢٣٥	٢٦٨	٢٩٩	٣٥٢
٣٣٤	٢٥١	٣٦٧	٣٠٩	٣١٥
٢٢٨	٢٥٩	٣٥٤	٢٨٣	٣٠٦
٢٣٤	٢٥٨	٣٦٥	٢٨١	٣١٢
٣٤٢	٢٦٨	٢٧٨	٢٩١	٢٨٨

يتم ترتيب البيانات تصاعدياً من أقل قيمه حتى أكبر قيمه ثم حساب المدى بطرح أكبر قيمة من أقل قيمة. في مثال القطن كانت أكبر قيمة ٣٧٣ وأقل قيمة ٢١٥ ولذلك فإن المدى يساوي ١٥٨ رطلاً من الكتان. من ناحية أخرى فإن ترتيب البيانات الجدول رقم (٢,٢) يشير إلى توزيع الوحدات بين أكبر قيمة وأصغر قيمة ومدى ميل تلك البيانات للتجمع حول قيمة معيّنة مثل ٢٩٠ في مثال القطن.

الجدول رقم (٢,٢). إنتاجية نسيج القطن الكتاني (باوند / أيكر) لخمسة و سبعون مزرعة من مزارع سهول الأراضي السوداء مرتبة تصاعدياً.

٢١٥	٢١٧	٢٢٨	٢٣٤	٢٣٥
٢٤٢	٢٤٩	٢٥٠	٢٥١	٢٥٤
٢٥٥	٢٥٦	٢٥٧	٢٥٨	٢٥٩
٢٦٠	٢٦٠	٢٦١	٢٦٨	٢٦٨
٢٧٢	٢٧٤	٢٧٤	٢٧٥	٢٧٦
٢٧٨	٢٧٩	٢٨٠	٢٨١	٢٨٣
٢٨٣	٢٨٣	٢٨٤	٢٨٥	٢٨٦
٢٨٨	٢٨٨	٢٩٠	٢٩٠	٢٩٠

تابع الجدول رقم (٢,٢).

٢٩١	٢٩٢	٢٩٢	٢٩٤	٢٩٥
٢٩٥	٢٩٦	٢٩٧	٢٩٩	٢٩٩
٣٠٣	٣٠٥	٣٠٦	٣٠٩	٣٠٩
٣١١	٣١٢	٣١٢	٣١٤	٣١٥
٣١٦	٣١٨	٣٢٥	٣٢٦	٣٣٣
٣٣٤	٣٣٦	٣٤٢	٣٤٦	٣٥٢
٣٥٤	٣٥٨	٣٦٥	٣٦٧	٣٧٣

يمكن أيضاً تلخيص البيانات في شكل توزيع تكراري كما هو موضح في جدول رقم (٢,٣) حيث يتم إنشاء التوزيع بتقسيم البيانات إلى عدد محدد من الفئات لكل فئة طول معين أو عدد من الوحدات يعبر عن فترة. فبالنسبة لبيانات القطن في جدول رقم (٢,٢) والذي يحتوي على البيانات المرتبة تم تقسيم البيانات لثمان فئات طول كل فئة ٢٠ رطلاً من النسيج الكتاني. ويشير عمود التكرارات إلى عدد المشاهدات التي تقع في الفئة ، فمثلاً القيم في الفئة الأولى يمكن أن تصل إلى القيمة ٢٣٥ ولكن لا تساويها. وإذا كانت القيمة ٢٣٥ من بين البيانات فإنها تصبح أحد القيم للفئة الثانية والتي تعتبر القيمة ٢٣٥ حد أدنى لها. من جهة أخرى فإن إجمالي عدد الفئات يجب أن يساوي العدد الإجمالي للبيانات والذي يساوي ٧٥ مزرعة في مثال القطن.

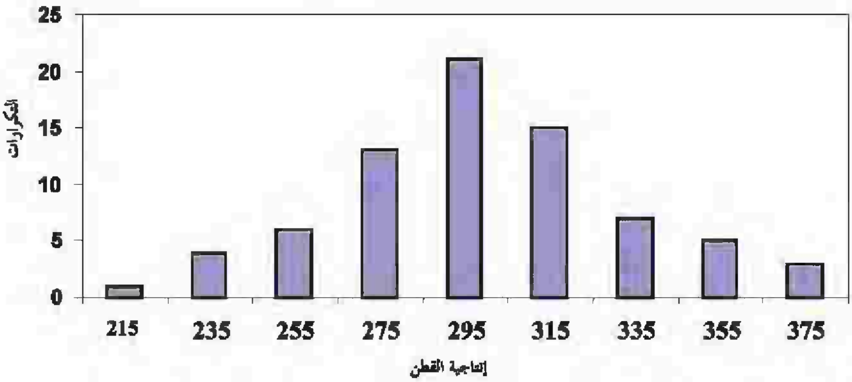
ليس بالضرورة أن يكون طول الفئات متساوي في جميع التوزيعات التكرارية وخاصة عند التعامل مع بيانات الدخل ؛ حيث إن استخدام فئات متساوية الطول يؤدي إلى أن بعض الفئات تكون خالية أو تحتوي على مشاهدات قليلة. وفي مثل هذه الحالات يفضل استخدام فئات بطول غير متساوي مما يساهم في تخفيض التوزيع

الجدول رقم (٢,٣). التوزيع التكراري لإنتاجية القطن (باوند/أيكر) لخمسة و سبعون مزرعة من مـــزارع سهول الأراضي السوداء.

إنتاجية القطن	عدد المزارع
٢١٥ حتى ٢٣٥	٤
٢٣٥ حتى ٢٥٥	٦
٢٥٥ حتى ٢٧٥	١٣
٢٧٥ حتى ٢٩٥	٢١
٢٩٥ حتى ٣١٥	١٥
٣١٥ حتى ٣٣٥	٧
٣٣٥ حتى ٣٥٥	٥
٣٥٥ حتى ٣٧٥	٤
الإجمالي	٧٥

التكراري وبالتالي يساعد في التفسير. في بعض مجموعات البيانات يفضل أن تكون الفئة الأخيرة مفتوحة بدلاً من أن يكون للفئة حد أعلى نظراً لوجود عدد قليل من المشاهدات أكبر من الحد الأعلى المقترح للفئة، وبذلك يمكن كتابة الفئة الأخيرة ٣٥٥ أو أعلى أو أكبر من ٣٥٥. وبالرغم من ذلك إلا أن استخدام هذه الطريقة لها بعض العيوب المتمثلة في عدم القدرة على حساب بعض المتوسطات ومقاييس التشتت مثل المتوسط الحسابي ومتوسط المدى والانحراف المعياري والمدى من التوزيع التكراري.

ويمكن عرض التوزيع التكراري باستخدام الأعمدة البيانية الشكل رقم (٢,١) ويسمى هذا الشكل بالمدرج التكراري Histogram والميزة لاستخدام الرسم في عرض التوزيع التكراري هي إمكانية مشاهدة الشكل الحقيقي لذلك التوزيع ومن ثم يمكن تحديد هل الشكل متماثل أم ملتوي وكذلك مستوى التفرطح بالنسبة للشكل. ويتضح من بيانات القطن في الشكل رقم (٢,١) أنها تقريباً متماثلة وكذلك التوزيع نوعاً ما له



الشكل رقم (٢،١). المدرج التكراري لإنتاجية القطن الكتاني بالباوند/ إيكر مزارع سهول الأراضي السوداء.

قيمة (قيمة عظمى) أي أنه ليس مسطح ولا مدبب.

أما المضلع التكراري Frequency polygon فهو عبارة عن خط بياني يستخدم أحياناً لعرض البيانات حيث يتم عرض التكرارات على المحور الرأسي (الصادي Y) ومراكز الفئات على المحور الأفقي (السيني X) وتوصيل النقاط بخطوط مستقيمة ويتم تحديد مركز فئة تخّيلية (وهمية) قبل بداية البيانات ومركز فئة تخّيلية (وهمية) بعد نهاية البيانات بحيث يكون تكرارها يساوي الصفر وذلك بهدف توصيل المنحنى بالمحور الأفقي ليعطي مدلولاً مناسباً ويتم حساب مراكز الفئات بأخذ متوسط الحد الأعلى للفئة والحد الأدنى. ومثال على ذلك بالنسبة للفئة من ٢١٥ إلى أقل من ٢٣٥ يكون مركزها $(215+235)/2=225$ وهكذا بالنسبة لجميع الفئات (انظر الجدول رقم ٢،٤).

وتتميز المضلعات التكرارية بسهولة استخدامها مقارنة بالمدرجات التكرارية لعرض شكل التوزيع.ومن المعتاد استخدام المضلعات التكرارية لرسم التكرارات

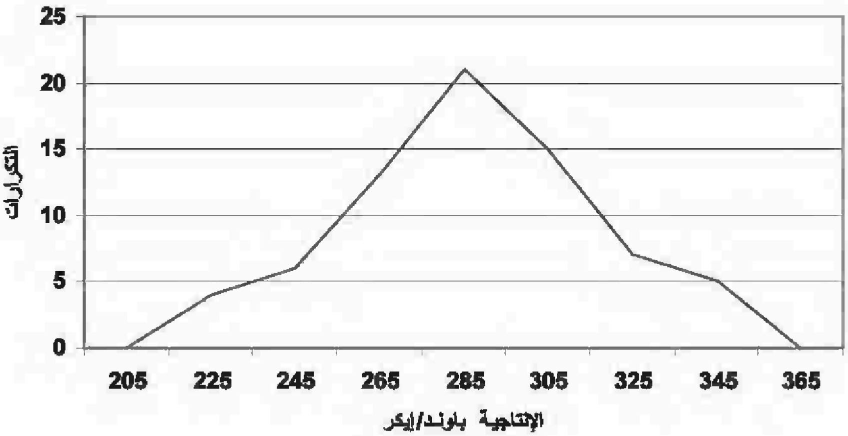
النسبية على نفس المنحنى لمقارنة التوزيعات خاصة إذا كان لدينا عدد مختلف من التكرارات في كل توزيع. يتحدّد التكرار النسبي بقسمة تكرار الفئة على مجموع التكرارات للتوزيع وضربه بـ ١٠٠ لتحويله لنسبة مئوية.ولتوضيح كيفية مقارنة توزيعين يمكن دراسة توزيعات الأجر الأسبوعي لسائقي القاطرات (الشاحنات) وعمال المزرعة كما هو موضح في الجدول رقم (٢,٥). وكما يلاحظ من البيانات بأن أجور العمال الزراعيين بشكل عام أقل ولكن عند رسم تلك البيانات باستخدام المضلع التكراري والموضح بالشكل رقم (٢,٢) نجد صعوبة في المقارنة لصغر العدد بالنسبة للعمال الزراعيين. ويوضح الشكل رقم (٢,٣) التكرارات النسبية للتوزيعين والشكل رقم (٢,٤) المضلعات التكرارية النسبية لهما ، ويلاحظ بشكل عام أن توزيع الأجر بالنسبة لسائقي الشاحنات أكبر من توزيع الأجر بالنسبة للعمال الزراعيين ولكنها تأخذ تقريباً نفس الشكل في المثال السابق.

الجدول رقم (٢,٤). التوزيع التكراري باستخدام مراكز الفئات لبيانات إنتاجية القطن في مزارع سهول الأراضي السوداء.

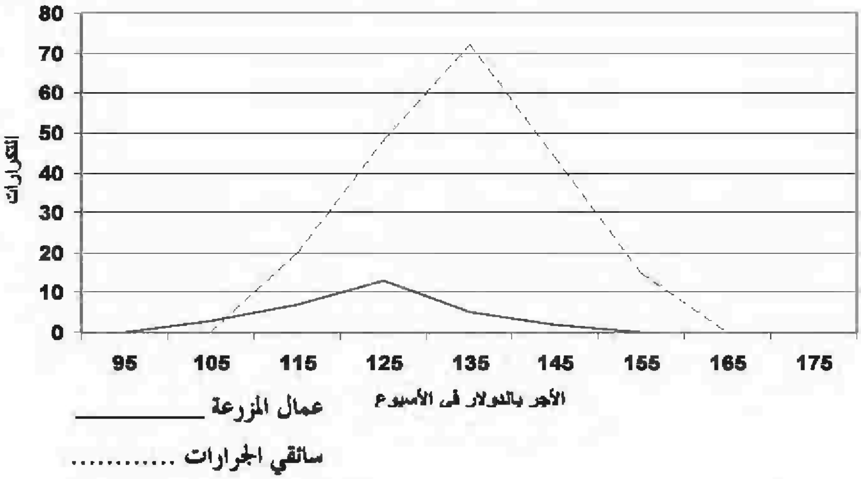
إنتاجية القطن	مراكز الفئات	عدد المزارع
فئة تخيلية	٢٠٥	٠
٢١٥ حتى ٢٣٥	٢٢٥	٤
٢٣٥ حتى ٢٥٥	٢٤٥	٦
٢٥٥ حتى ٢٧٥	٢٦٥	١٣
٢٧٥ حتى ٢٩٥	٢٨٥	٢١
٢٩٥ حتى ٣١٥	٣٠٥	١٥
٣١٥ حتى ٣٣٥	٣٢٥	٧
٣٣٥ حتى ٣٥٥	٣٤٥	٥
فئة تخيلية	٣٦٥	٠
الإجمالي		٧٥

الجدول رقم (٢,٥). التوزيعات التكرارية للأجر الأسبوعي لعمال المزرعة وسائقي الجرارات.

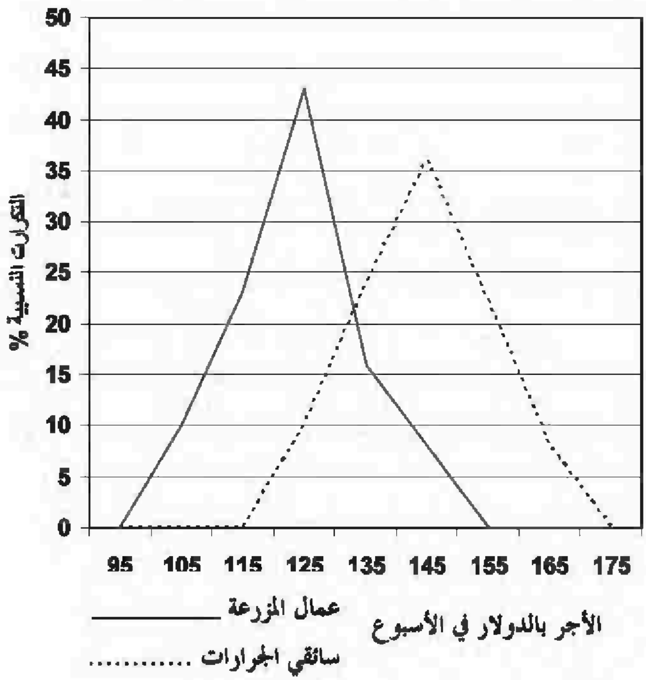
الأجر الأسبوعي	نقاط مراكز الفئة	التكرارات عمال المزرعة	التكرارات سائقي الجرارات	التكرارات النسبية عمال المزرعة	التكرارات النسبية سائقي الجرارات
١٠٠ حتى ١١٠	١٠٥	٣		١٠	٠
١١٠ حتى ١٢٠	١١٥	٧	٠	٢٣	٠
١٢٠ حتى ١٣٠	١٢٥	١٣	٢٠	٤٣	١٠
١٣٠ حتى ١٤٠	١٣٥	٥	٤٨	١٦	٢٤
١٤٠ حتى ١٥٠	١٤٥	٢	٧٢	٨	٣٦
١٥٠ حتى ١٦٠	١٥٥	٠	٤٤	٠	٢٢
١٦٠ حتى ١٧٠	١٦٥	٠	١٦	٠	٨
الإجمالي		٣٠	٢٠٠	١٠٠	١٠٠



الشكل رقم (٢,٢). المضلع التكراري لبيانات إنتاجية القطن في مزارع سهول الأراضي السوداء.



الشكل رقم (٢,٣). المضلع التكراري للأجر الأسبوعي لعمال المزرعة وسائقي الجرارات.



الشكل رقم (٢,٤). المضلع التكراري النسبي للأجر الأسبوعي لعمال المزرعة وسائقي الجرارات.

في المثال السابق تم تحديد عدد الفئات للتوزيع التكراري ولكن بصفة عامة وكقاعدة يتم اختيار القيمة لعدد الفئات في المدى بين ٥ ، ١٥. حيث إن عدد الفئات أقل من ٥ لايعطي خصائص كافية يمكن من خلالها وصف البيانات وكذلك الحال بالنسبة لعدد فئات أكثر من ١٥ يعتبر كبيراً جداً. في بعض الأحيان يتم استخدام قاعدة سترجس Sturges's rule والموضحة بالمعادلة التالية :

$$K = 1+3.322(\log_{10} n)$$

حيث K : عدد الفئات.

n : عدد التكرارات الإجمالي.

وفي الغالب فإن قيمة K لن تكون رقم صحيح وعليه يجب التقريب لأعلى أو أقل رقم صحيح. ويمكن تحديد طول الفئة I بقسمة مدى البيانات (عبارة عن الفرق بين أكبر قيمة وأصغر قيمة) على عدد الفئات K. وبعد تحديد عدد الفئات وطول الفئة يجب التأكد من البيانات ؛ ومدى ملائمة الفئات المختارة لجميع البيانات حيث إنه في بعض الأحيان قد يكون هناك مشاهدة خارج حدود الفئات المحددة، والاعتبار الآخر الواجب ملاحظته في اختيار طول الفئة هو أن يعرض مركز الفئة نفس البيانات في الفئة. فمثلاً إذا كانت بيانات الأجر تنتهي بـ ٥ دولارات يجب اختيار مركز الفئة بحيث ينتهي بـ ٥ دولارات وذلك بتعديل الفئات ؛ حيث إن قاعدة سترجس تقريبية ويجب أن نقرر في الأخير ما يجب عمله.

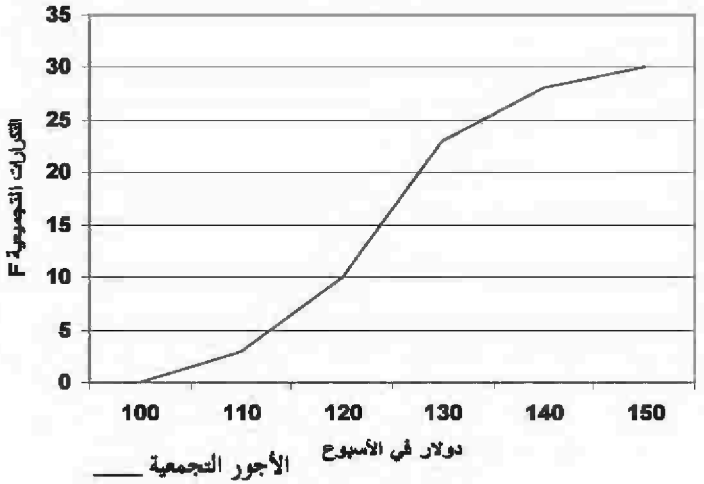
من جهة أخرى ولفهم محتوى البيانات يتم استخدام التوزيعات التكرارية المتجمعة حيث يتم تكوين توزيعات تكرارية صاعدة (أقل من) وتوزيعات تكرارية هابطة (أكثر من). ويتم تكوين التكرارات الصاعدة بجمع التكرارات بُدءاً بأقل فئة

وانتهاءً بأعلى فئة، وتكتب كلمة أقل من أمام الحد الأعلى للفئة ثم تحسب التكرارات المتجمعة الصاعدة بجمع التكرارات بدْءاً بأقل فئة و انتهاءً بأعلى فئة (الجدول رقم ٢,٦).

الجدول رقم (٢,٦). التوزيع التكراري التجميعي الصاعد لأجور العمال الزراعيين الأسبوعية.

الأجور	التكرارات التجميعية F
أقل من ١٠٠	٠
أقل من ١١٠	٣
أقل من ١٢٠	١٠
أقل من ١٣٠	٢٣
أقل من ١٤٠	٢٨
أقل من ١٥٠	٣٠

ويمكن رسم التكرار المتجمع الصاعد والذي يأخذ شكل حرف S حيث تُمثَّل الفئات (أقل من) على المحور السيني (الأفقي) وتُمثَّل التكرارات المتجمعة الصاعدة على المحور الصادي (الرأسي) (الشكل رقم ٢,٥). وهذا الشكل يساعد في تقسيم التوزيعات التكرارية إلى أجزاء متساوية مثل المئينيات أو العشيريات أو الربيعيات أو أي إحصائيات ترتيبية أخرى؛ حيث إن المئينيات تقسّم البيانات إلى ١٠٠ جزء متساوٍ وفي الربيعيات تقسم إلى أربعة أجزاء متساوية.

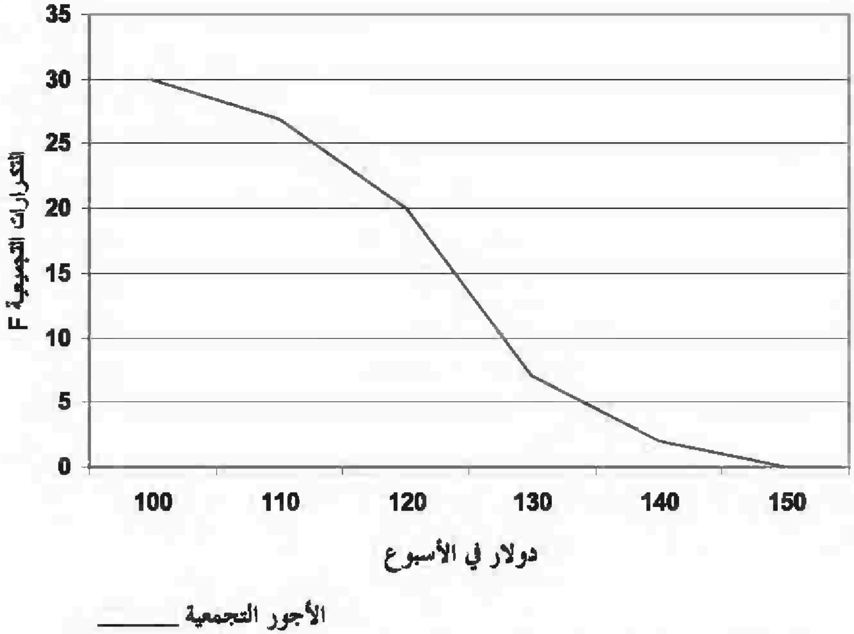


الشكل رقم (٢,٥). المضلع التكراري التجميعي الصاعد لأجور عمال المزرعة.

أما التكرار المتجمع الهابط (أكبر من) يتم البدء بالفئة الأعلى والانتهاء بالفئة الأقل حيث نبدأ بالفئة الأكبر من أقل فئة (الجدول رقم ٢,٧)، ويتم رسم التكرار المتجمع الهابط بحيث نضع الحدود الدنيا للفئات (أكبر من) على المحور السيني والتكرارات المتجمعة على المحور الرأسي ومن ثم نحصل على صورة عكسية للتكرار المتجمع الصاعد (الشكل رقم ٢,٦)، وعند رسم التكرار المتجمع الصاعد و التكرار المتجمع الهابط معاً على رسم واحد فإن الشكلين يتقاطعان في الوسيط.

الجدول رقم (٢,٧). التوزيع التكراري التجميعي الهابط لأجور العمال الزراعيين الأسبوعية.

الأجور	التكرارات التجميعية F
أكثر من ١٠٠	٣٠
أكثر من ١١٠	٢٧
أكثر من ١٢٠	٢٠
أكثر من ١٣٠	٧
أكثر من ١٤٠	٢
أكثر من ١٥٠	٠



الشكل رقم (٢,٦). المنحنى المتجمع الهابط للأجور التجميعية للعمال الزراعيين.

المتوسطات Averages

يعرّف المتوسط بأنه رقم يستخدم للتعبير عن القيمة المركزية لمجموعة من البيانات أو التوزيع. وفي الغالب تستخدم المتوسطات بكثرة مقارنة بالمقاييس الإحصائية الأخرى ويتم حسابها للبيانات الخام (غير المبوبة) و البيانات المبوبة والتي تم تلخيصها في جداول تكرارية. والطريقة الحسابية المستخدمة للنوعين من البيانات سابقاً تختلف في الغالب عن بعضها ؛ نظراً لأننا نحتاج لصيغ رياضية أخرى عند التعامل مع التوزيعات التكرارية.

على الرغم من أن حساب المتوسطات من العينات لا تختلف عنها في المجتمع ولكن بشكل عام يتم استخدام رموز مختلفة للتعبير عن المتوسطات بالنسبة للمجتمع والعينة، وذلك لمنع اللبس الممكن حدوثه حول نوعية البيانات التي تتم معالجتها. وكما هو معلوم أن العينة مجموعة جزئية من المجتمع في حين أن المجتمع يحتوي جميع البيانات في المجموعة. وفي الغالب تستخدم الأحرف الإغريقية في المعادلات عند التعامل مع بيانات المجتمع.

الوسط الحسابي The Arithmetic Mean

البيانات غير المبوبة Ungrouped Data

يعتبر الوسط الحسابي من أكثر المقاييس استخداماً لقياس النزعة المركزية ويمكن حساب المتوسط، μ، للمجتمع بجمع قيم المشاهدات، X_i ، وقسمتها على عددها N كما في المعادلة رقم (2.1):

$$\mu = \frac{\sum_{1}^{N} x_i}{N} \qquad (2.1)$$

مثال : أوجد المتوسط الحسابي للبيانات التالية :

7, 3, 2 , 8

$$\sum_{1}^{4} X_i = 7+3+2+8 = 20$$

$$N = 4$$

$$\mu = \frac{\sum_{1}^{N} x_i}{N} = 20/4 = 5$$

وفي حالة أخذ عينة من هذا المجتمع ولنفرض أن حجمها يساوي ٢ يمكن حساب المتوسط الحسابي للعينة باستخدام المعادلة رقم (2.2):

$$\bar{X} = \frac{\sum_{1}^{n} X_i}{n} \qquad (2.2)$$

وتعتمد قيمة متوسط العينة على البيانات التي تم اختيارها من المجتمع سابقاً، ففي حالة اختيار أول مشاهدتين (٧، ٣) فإن المتوسط يكون حاصل جمعها ١٠ مقسوماً على عددها ٢ $\bar{X} = \frac{10}{2} = 5$ والذي يعبر تماماً عن متوسط المجتمع.

ولكن في حالة اختيارات أخرى للعينة من المجتمع فإن متوسط العينة ليس بالضرورة أن يساوي متوسط المجتمع فقد يكون أكبر أو أصغر منه وعليه فإن قيمة متوسط العينة تعتمد على عناصر العينة المختارة من المجتمع ولذلك فإنها من المحتمل أن تساوي أو لا تساوي متوسط المجتمع μ.

وكلما كان حجم العينة المختارة من المجتمع كبير (n) فإن تأثير القيم المتطرفة في المجتمع سيكون قليلاً وبذلك يقترب متوسط العينة من متوسط المجتمع.

في المثال السابق عند اختيار العينة فإن جميع عناصر المجتمع كان لها نفس الوزن

ويساوي ١ ولكن في بعض الأحيان قد نضطر لاستخدام أوزان للعنصر تختلف عن الواحد ويرمز له بـ w_i وفي هذه الحالة فإننا نحسب المتوسط المرجح والذي يمكن حسابه باستخدام المعادلة التالية:

$$\bar{X} = \frac{\sum_{1}^{n} W_i X_i}{\sum_{1}^{n} W_i} \qquad (2.3)$$

مثال: البيانات الموضحة في الجدول رقم (٢,٨) توضح الأجر بالساعة وعدد العمال المهاجرين(الموسمين) لمجموعة من محاصيل الخضار والمطلوب حساب المتوسط المرجح.

الجدول رقم (٢,٨). أجر الساعة وعدد العمال الموسميين حسب نوع المحصول.

المحصول	الأجر بالساعة، x	عدد العمال، w	wx
خيار	٤,٥	٩٥٠	٤٢٧٥
شمام	٤,٧٥	٦٠٠	٢٨٥٠
بصل	٥,٢٥	١٠٢٠	٥٣٥٥
الإجمالي	ــ	٢٥٧٠	١٢٤٨٠

يمكن حساب المتوسط المرجح بقسمة المجموع الكلي لحاصل ضرب عدد العمال للمحصول في أجر الساعة لجميع المحاصيل على مجموع العمال باستخدام المعادلة رقم (2.3) $\bar{X} = 12480/2570 = \$4.86$ وبذلك يكون المتوسط المرجح للأجر ٤,٨٦ دولاراً للساعة وهو يكون أدق نظراً لأن عدد العمال غير متساوٍ في كل محصول وعند حساب المتوسط غير المرجح فإنه سيعطى نتيجة مختلفة. وفي الحالات التي تكون

فيها الأوزان كسور عشرية مجموعها يساوي الواحد الصحيح فإن الصيغة الرياضية السابقة يمكن تعديلها بحيث يمكن حذف المقام ؛ نظراً لأن أي قيمة تقسم على الواحد لا تتغير.

خصائص الوسط الحسابي

يوجد خاصيتين للوسط الحسابي هما :

١- مجموع انحرافات القيم عن الوسط الحسابي يساوي صفر.

ويمكن حساب ذلك من المعادلة التالية :

$$x = (X - \overline{X}) \tag{2.4}$$

$$\sum x = 0$$

٢- مجموع مربعات انحرافات القيم عن متوسطها أقل ما يمكن.

فإذا كانت $M = \sum X / n$ عليه فإن $\sum (X - M)^2$ سوف تكون أقل قيمة ممكنة مقارنة بأي قيمة محسوبة بطريقة أخرى لـ M.

البيانات المبوبة Grouped Data

عند تبويب البيانات في شكل توزيعات تكرارية فإن قيم المشاهدات المفردة لن تكون متاحة في الجدول ؛ لأنها على شكل فترات ولذلك يتم التعامل مع التوزيع على شكل فترات ويتم استخدام مركز الفئة كقيمة تقريبية لجميع المشاهدات داخل الفئة. وحيث إن عدد التكرارات يختلف بين الفئات يتم استخدام التكرارات كأوزان في حساب الوسط باستخدام المعادلة التالية :

$$\overline{X} = \frac{\sum_{1}^{n} f_i x_i}{\sum_{1}^{n} f_i} \tag{2.5}$$

حيث f_i تعبر عن التكرار للفئة i

X_i مركز الفئة i

ولنفترض أننا نرغب في حساب متوسط إنتاجية القطن للتوزيع التكراري في الجدول التالي.

الجدول رقم (٢,٩). حساب المتوسط الحسابي لبيانات إنتاجية القطن باستخدام التوزيع التكراري.

إنتاجية القطن	التكرارات، f	مراكز الفئات، X	FX
٢١٥ حتى ٢٣٥	٤	٢٢٥	٩٠٠
٢٣٥ حتى ٢٥٥	٦	٢٤٥	١٤٧٠
٢٥٥ حتى ٢٧٥	١٣	٢٦٥	٣٤٤٥
٢٧٥ حتى ٢٩٥	٢١	٢٨٥	٥٩٨٥
٢٩٥ حتى ٣١٥	١٥	٣٠٥	٤٥٧٥
٣١٥ حتى ٣٣٥	٧	٣٢٥	٢٢٧٥
٣٣٥ حتى ٣٥٥	٥	٣٤٥	١٧٢٥
٣٥٥ حتى ٣٧٥	٤	٣٦٥	١٤٦٠
الإجمالي	٧٥		٢١٨٣٥

في هذه الحالة فإن الوسط الحسابي للتوزيع التكراري يساوي $\overline{X} = 21835/75 = 291.1$ أو ٢٩١ رطلاً للأيكر بينما عند حساب الوسط الحسابي للبيانات الأصلية(انظر الجدول رقم ٢,١) نجد أن المتوسط الحسابي $\overline{X} = 21807/75 = 290.8$ ولذلك فإن استخدام البيانات المبوبة أعطى قيمة أعلى للمتوسط بحوالي ٠,٣ رطل/أيكر.

ومن النتائج السابقة يتضح أن استخدام صيغة التوزيع التكراري بشكل عام قد لا تؤثر على دقة النتائج ؛ لأن عملية إعداد التوزيع التكراري تجعل مركز الفئة تقريب مناسب

للبيانات في الفئة وهذا في الحقيقة يعتبر من الأخبار الجيّدة؛ نظراً لأن كثير من مصادر البيانات الثانوية في الغالب تكون توزيعات تكرارية ولذلك عند الحاجة لحساب بعض المؤشرات الإحصائية مثل المتوسط يمكن الاعتماد على تلك البيانات بدرجة من الثقة.

متوسط المدى The Midrange

متوسط المدى أو المركز عبارة عن الوسط الحسابي لأعلى قيمة وأقل قيمة للبيانات ولذلك عند ترتيب البيانات تصاعدياً أو تنازلياً فإنه عبارة عن متوسط القيمة الأولى والقيمة الأخيرة.

$$MR = \frac{X_1 + X_n}{2} \qquad (2.6)$$

حيث X_1 أقل قيمه ، و X_n أعلى قيمه

ولبيانات إنتاجية القطن الموضحة في جدول رقم (٢،٢) فإن متوسط المدى يساوي:

$$MR = (215+373)/2 = 294$$

ويستخدم متوسط المدى في الغالب لحساب المتوسط اليومي لدرجات الحرارة أو متوسط أسعار الأسهم؛ نظراً لأننا نهتم بأعلى قيمة وأقل قيمة لذلك النوع من البيانات.ولكن بصفة عامة فإن متوسط المدى غير ملائم عند تقدير متوسط عدد السكان؛ نظراً لأنها تعتمد على قيمتين تتغير من عينة لعينة ولذلك فإن قيمته سوف تختلف كثيراً من عينة لأخرى.

الوسيط The Median

يختلف الوسيط عن المتوسط الحسابي؛ حيث إن المتوسط الحسابي يعتمد على جميع قيم المشاهدات في حين أن الوسيط يعبّر عن مكان المتوسط. والوسيط لا يتأثر

بالقيم المتطرفة للبيانات ويعتبر ملائماً للاستخدام في الحالات التي نصادف فيها بعض القيم المتطرفة، كما في حالة بيانات الدخل والتعليم.

يحسب الوسيط للبيانات غير المبوبة بترتيب البيانات تصاعدياً أو تنازلياً ويكون الوسيط هو القيمة التي في وسط البيانات إذا كان عدد البيانات فردياً أما إذا كان عدد البيانات زوجياً فإن الوسيط عبارة عن الوسط الحسابي للقيمتين التي في الوسط بعد ترتيب البيانات.

مثال: احسب الوسيط للبيانات التالية ٧، ٣، ٢، ٨

لإيجاد الوسيط ترتب البيانات تصاعدياً ٢، ٣، ٧، ٨ وكون عدد البيانات زوجياً نأخذ متوسط القيم التي في الوسط ٣، ٧ ولذلك فإن الوسيط $5 = \frac{7+3}{2}$.

الجدير بالذكر أن استخدام الوسيط في البيانات القليلة (أقل من ٦ مشاهدات) لا يعطي نتائج دقيقة؛ لأنه يعبّر عن مركز البيانات ولكن في حالة كان عدد المشاهدات أكثر من ٢٠ مشاهدة يمكن الاعتماد على الوسيط كمقياس حيث على الأقل تكون نصف المشاهدات أقل من الوسيط والنصف الآخر أعلى من الوسيط، ويكون الوسيط في حالة التوزيعات المدببة جداً extremely peaked أكثر المقاييس ثقة كتقدير لمتوسط المجتمع.

في حالة البيانات المبوبة يتم حساب الوسيط باستخدام الصيغة الرياضية الموضحة في المعادلة رقم (2.7):

$$\mathrm{Md} = L + \frac{\frac{n}{2} - F}{f} I \qquad (2.7)$$

حيث:

L : الحد الأدنى لفئة الوسيط.

n : مجموع التكرارات.

F : قيمة التكرار المتجمع الصاعد للفئة التي قبل فئة الوسيط.

f : تكرار فئة الوسيط.

I : طول فئة الوسيط.

ولتحديد فئة الوسيط يتم حساب التكرار المتجمع الصاعد ومن ثم حساب رتبة الوسيط والتي تساوي $\frac{n}{2}$ حيث كما نعلم فإن الوسيط عبارة عن القيمة التي تقع في وسط البيانات. ويمكن تحديد الفئة التي تحتوي على الوسيط بعد حساب رتبة الوسيط باستخدام التكرار المتجمع الصاعد ؛ حيث إن الحد الأدنى لتلك الفئة هو L وعدد تكرارات تلك الفئة هو f وطول تلك الفئة هو I . أما قيمة F فإنها عبارة عن قيمة التكرار المتجمع الصاعد للفئة التي تسبق الوسيط ومن ثم يتم التعويض بتلك القيم في المعادلة السابقة لحساب قيمة الوسيط.

مثال: لبيانات الأجر لعمال المزرعة الموضحة في الجدول رقم (٢،١٠) احسب قيمة الوسيط.

من تلك البيانات نلاحظ أن مجموع التكرارات يساوي٣٠ وبذلك يمكن تحديد رتبة الوسيط والتي تساوي $\frac{30}{2} = 15$. ولتحديد فئة الوسيط يمكن مقارنة رتبة الوسيط بقيم التكرارات المتجمعة الصاعدة F في الجدول حيث تقع بين ١٠ و ٢٣، ويشير التكرار المتجمع الصاعد ١٠ إلى أن الفئة ١١٠ إلى أقل من ١٢٠ تحتوي على ٧ قيم تبدأ برقم ٤ وتنتهي بالرقم ١٠ وبصفة مشابهه فإن التكرار المتجمع الصاعد ٢٣ يعني أن الفئة من ١٢٠ إلى أقل من ١٣٠ تحتوي على ثلاث عشرة قيمة تبدأ بالرقم ١١ وتنتهي بالرقم ٢٣ ولذلك فإن الرقم ١٥ يقع في الفئة من ١٢٠ إلى أقل من ١٣٠ والتي تعبّر عن الفئة التي تحتوي على الوسيط أو فئة الوسيط ولذلك فإن القيم هي $L = 120$, f=13 , I= 10 , F=10

وبالتعويض بتلك القيم في المعادلة رقم (2.7) يمكن حساب الوسيط كالتالي :

$$Md = 120 + \left(\frac{15 - 10}{13}\right)(10) = 120 + 3.8 = 123.8$$

الجدول رقم (٢،١٠). التوزيع التكراري للأجر الأسبوعي للعمال الزراعيين المستخدم في حساب الوسيط.

الأجر اليومي	التكرارات، f	التكرارات التجميعية، F
من ١٠٠ إلى ١١٠	٣	٣
من ١١٠ إلى ١٢٠	٧	١٠
من ١٢٠ إلى ١٣٠	١٣	٢٣
من ١٣٠ إلى ١٤٠	٥	٢٨
من ١٤٠ إلى ١٥٠	٢	٣٠
الإجمالي	٣٠	

المنوال The Mode

يعبّر المنوال عن متوسط آخر حيث يُحدَّد بالقيمة الأكثر تكراراً في مجموعة البيانات. للبيانات غير المبوبة، يتم تحديد المنوال بالبحث حيث يتم ببساطة فحص البيانات قيد الدراسة ومن ثم تحديد القيمة الأكثر تكرار فيها والتي تعبّر عن المنوال.وفي حالة البيانات غير المبوبة قد لا يوجد منوال إذا كانت جميع القيم تظهر لمرة واحدة وقد يوجد أكثر من منوال بالبيانات في حالة تكرار القيم المتشابهة نفس العدد من المرات. وتجدر الإشارة إلى أن المنوال في حالة البيانات غير المبوبة غير حقيقي ؛ نظراً لأن قيمته تتغير في حالة إعادة العينة المسحوبة. ويستخدم المنوال في الحالات التي تتطلب معرفة للحالات المعتادة أو الأكثر شيوعاً مثل برامج التلفزيون المرغوبة، البرامج الأكثر مشاهدة من المشاهدين، أو الاختلافات الأكثر شيوعاً لمحصول الذرة المزروعة ... إلخ.

في حالة البيانات المبوبة أو للتوزيعات التكرارية بشكل عام فإن المنوال عبارة عن القيمة التي تقع عند أعلى نقطة للتوزيع. والمنوال البسيط عبارة عن مركز الفئة التي تحتوي أكثر تكرارات.

ويتضح من الجدول رقم (٢,١٠) أن المنوال البسيط يساوي ١٢٥ ؛ لأن الفئة من (١٢٠ إلى ١٣٠) تحتوي على أعلى تكرار ١٣. ويمكن استخدام الصيغة الرياضية التالية لحساب المنوال معادلة رقم (2.8) .

$$Mode = L + \frac{d_1}{d_1 + d_2} I \tag{2.8}$$

حيث :

L : الحد الأدنى لفئة المنوال.

d_1 : الفرق الأول: عبارة عن الفرق بين تكرار فئة المنوال وتكرار الفئة السابقة.

d_2 : الفرق الثاني: عبارة عن الفرق بين تكرار فئة المنوال وتكرار الفئة التالية.

I : طول فئة المنوال.

ويمكن تحديد فئة المنوال بالبحث ؛ حيث إنها تحتوي على أعلى تكرار.

وفي حالة المثال السابق الموضح في جدول رقم (٢,١٠) نجد أن الحد الأدنى لفئة المنوال يساوي $L = 120$ والفرق الأول له يساوي٦ ،(١٣- ٧)، والثاني له يساوي ٨ ، (١٣-٥)، وطول فئة المنوال I يساوي ١٠.

وبالتعويض في المعادلة (2.8) نحصل على:

$$Mode = 120 + \left(\frac{6}{6+8}\right)(10) = 120 + 4.3 = 124.3$$

وهذه القيمة مساوية تقريباً لقيمة المنوال البسيط والتي تساوي ١٢٥. عندما يكون التوزيع ثنائي القمة Bimodal يفضل تجزئة البيانات إلى مجموعتين وتحليل كل مجموعة على حدة؛ لأن إهمال بعض البيانات وعدم إدخالها في التحليل قد يسبب تكتلاً للبيانات في المجموعتين.

خصائص الوسط والوسيط والمنوال

يمكن استخدام المتوسطات الثلاثة معاً لتحديد التماثل النسبي أو الالتواء للتوزيع. فعندما يكون التوزيع ذو شكل متماثل تماماً فإن قيمة الوسط والوسيط والمنوال تكون متساوية. أما إذا كان التوزيع له ذيل باتجاه اليمين، أو موجب الالتواء، فإن قيمة المتوسط الحسابي تكون الأكبر والمنوال يكون الأصغر قيمة والوسيط يقع في حوالي ثلثي المسافة بين الوسط والمنوال ويكون أقرب للوسط. ويرجع سبب ارتفاع قيمة المتوسط في هذا التوزيع لتأثره بالقيم الشاذة الكبيرة. أما الوسيط فهو يتأثر بمواقع تلك القيم دون أن يتأثر بحجمها. في حالة التوزيع ذي الذيل الأطول باتجاه اليسار، أو سالب الالتواء، تكون قيمة المنوال هي الأكبر وقيمة المتوسط هي الأصغر ويكون الوسيط بين القيمتين بموقع ثلثي المسافة باتجاه الوسط. وعليه فإن معرفة أي قيمتين من تلك القيم الثلاث تمكننا من الحكم على شكل التوزيع هل هو متماثل أو ملتوٍ باتجاه اليمين أو ملتوٍ باتجاه اليسار.

في بعض الأحيان قد يكون التوزيع بدون منوال مثل التوزيعات المفرطحة جداً أو التوزيعات المستطيلة والتوزيعات المتماثلة.

ويمتاز الوسط الحسابي عن المتوسطين الآخرين بإمكانية استخدامه في الحسابات الجبرية ولهذا السبب فإنه من أكثر المقاييس استخداماً، فعلى سبيل المثال يمكن حساب

المتوسط العام لمجموعة من متوسطات العينات بأخذ المتوسط للمتوسطات ثم ترجيحه بعدد المشاهدات في كل عينة.

ومن ناحية أخرى فإن من الصعوبة بل من المستحيل في بعض الأحيان حساب المتوسط عندما يحتوي التوزيع التكراري على فئات مفتوحة في حين أن الوسيط والمنوال لا يتأثر في هذه الحالة حيث يمكن حسابها.

مقاييس التشتت Measures of Dispersion

يعرف التشتت كممثل أو مقياس لمدى صحة المتوسط حيث يبيّن القيمة التي تتوزع بها البيانات سواء في حالة البيانات الخام أو التوزيعات التكرارية حيث تشير القيمة إلى مدى بعد البيانات أو تركزها حول متوسطها. وكلما كان مقدار التشتت للبيانات صغير بالنسبة لمتوسطها دلَّ ذلك على تمثيل المتوسط للبيانات الخام ولذلك يكون لدينا ثقة في استخدام المتوسط كرقم يمثل التوزيع. ولهذا السبب عند استخدام المتوسط كممثل للبيانات يجب أن يضاف له أحد مقاييس التشتت لتمكين المشاهدين من معرفة مدى تمثيل المتوسط للبيانات. فمثلاً في بيانات إنتاجية الذرة إذا كانت الإنتاجية لجميع المزارع تساوي ١٠٠بوشل فإن المتوسط سيكون ١٠٠ سواء أكان الوسط الحسابي أم الوسيط أو المنوال. ويكون التشتت في هذه الحالة مساوٍ للصفر ؛ لأن المتوسط يمثل البيانات تماماً. ولكن في حالة اختلاف أحد القيم عن ١٠٠ فإنه سيكون هناك تشتت أو انتشار في التوزيع. وكلما كان الاختلاف بين بيانات الإنتاجية كبير كلما كان انتشارها في التوزيع كبير وبالتي تقل إمكانية استخدام المتوسط كرقم يمثل مجموعة البيانات.

المدى The Range

المدى ، R ، أحد مقاييس التشتت ويحسب بإيجاد الفرق بين القيمة الأولى والقيمة الأخيرة في البيانات بعد ترتيبها تصاعدياً أو تنازلياً كما في المعادلة رقم (2.9). ويمكن استخدامه مع المتوسط أو الوسيط أو منتصف المدى ولكن لا يمكن استخدامه مع المنوال. وفي الحقيقة ليس للمنوال مقياس تشتت معين مرتبط به ولهذا السبب فإنه أقل المتوسطات استخداماً، إلا أن الميزة الوحيدة للمنوال هي استخدامه في الحالات غير الرقمية.ويقاس المدى باستخدام المعادلة رقم (2.9).

$$(R) = X_n - X_1 \tag{2.9}$$

وبالإضافة إلى عرض المدى لأعلى قيمة وأقل قيمة في البيانات، فإنه يستخدم أيضا لمعرفة مدى البيانات نفسها. ويمكن استخدام المدى لقياس تشتت البيانات في حالة العينات الصغيرة ($n \leq 12$) المختارة من التوزيع الطبيعي. ولكن في حالة العينات الكبيرة ($n \geq 30$) فإن المدى يعطي تقدير متحيز لتشتت المجتمع. وفي الحقيقة فإن المدى ليس الاختيار الأول لقياس التشتت ؛ نظراً لاعتماده على قيمتين متطرفتين من مجموعة البيانات ولذلك فإنه يتجه لعدم الاتساق في حالة المعاينة المتكررة من نفس المجتمع.

الانحراف الربيعي The Quartile Deviation

الانحراف الربيعي (QD) أحد مقاييس التشتت ويستخدم مع الوسيط حيث يعبّر عن تشتت البيانات في منتصف وسط التوزيع ويمكن حسابه باستخدام المعادلة رقم (2.10).

$$QD = \frac{Q_3 - Q_1}{2} \qquad (2.10)$$

حيث يمثل نصف الفرق بين الربيعي الأول Q_1 والربيعي الثالث Q_3 . وكما نعلم فإنه لحساب الربيعيات تقسم البيانات لأربعة أجزاء متساوية ولعمل ذلك نحتاج لثلاثة مقاييس: الربيعي الأول Q_1، الربيعي الثاني أو الوسيط والربيعي الثالث Q_3. وبصفة أخرى فإن الربيعي الأول يعبّر عن الوسيط للنصف الأول من البيانات والربيعي الثالث يعبّر عن الوسيط للنصف الثاني من البيانات ولإيجاد الربيعيات يتم ترتيب البيانات تصاعدياً ثم إيجاد الوسيط Md ، بعد ذلك نوجد الوسيط للنصف الأول من البيانات من X_1 إلى Md ونسميه Q_1 وهو يمثل الربيعي الأول، وبنفس الطريقة نوجد الوسيط للنصف الثاني من البيانات من الوسيط Md إلى Xn ونسميه Q_3 وهو يمثل الربيعي الثالث. بعد إيجاد الربيعيان نحسب الانحراف الربيعي ونستخدمه مع الوسيط.

مثال: البيانات التالية توضح عدد سنوات التعليم لعينة صغيرة من البالغين في الريف Rural ٨ ، ١٢ ، ٦ ، ١٤ ، ١٠

والمطلوب حساب الانحراف الربيعي.

الحل:

أولاً يتم ترتيب البيانات تصاعدياً ٦ ، ٨ ، ١٠ ، ١٢ ، ١٤ ثم نوجد الوسيط والذي يساوي ١٠ والربيعي الأول Q_1= ٧ والربيعي الثالث Q_3 = ١٣ ولذلك فإن الانحراف الربيعي يساوي:

$$QD = (13 - 7)/2 = 3$$

ويتشابه الانحراف الربيعي إلى حد ما مع المدى؛ نظراً لأنه يقيس الفرق بين قيمتين ولكن تلك القيمتين في منتصف وسط التوزيع. ونظراً لأن جزءاً من البيانات

الواقع بين الربيعي الثالث ونهاية البيانات لا يدخل في حساب الانحراف الربيعي فإن استخدام هذا المقياس يعطي نتائج غير دقيقه عندما يكون هناك تشتت كبير في هذا الجزء من البيانات. من جهة أخري، فإنه في حالة حساب المتوسط للبيانات ذات النهايات المفتوحة فإن الانحراف الربيعي يعتبر المقياس المناسب لتشتت هذه البيانات.

في حالة التوزيعات التكرارية (البيانات المبوبة) يمكن حساب الربيعي الأول Q_1، الربيعي الثالث Q_3 باستخدام صيغ رياضية مشابهة للتي استخدمت لحساب الوسيط والموضحة في المعادلات رقم (2.11) و (2.12).

$$Q_1 = L + \frac{\frac{n}{4} - F}{f} I \qquad (2.11)$$

$$Q_3 = L + \frac{\frac{3n}{4} - F}{f} I \qquad (2.12)$$

وهي نفس الصيغة المستخدمة لحساب الوسيط ماعدا أنه يتم تحديد الفئة التي تحتوي على الربيعي الأول بمقارنه $\frac{n}{4}$ مع التكرار المتجمع الصاعد F ، وتحديد الفئة التي تحتوي على الربيعي الثالث بمقارنة $\frac{3n}{4}$ مع التكرار المتجمع الصاعد F ، وبتحديد الفئات لكل ربيعي فإن L ، f ، I يمكن إيجادها و F هي عبارة عن التكرار المتجمع الصاعد للفئة السابقة لها.

الانحراف المعياري The Standard Deviation

يستخدم الانحراف المعياري كمقياس للتشتت مع الوسط الحسابي . وتعتمد قيمته على جميع المشاهدات في البيانات . ويسمى في بعض الأحيان بجذر متوسط مربعات الانحرافات ؛ نظراً لأنه يتم حسابه بأخذ الجذر التربيعي للوسط الحسابي لمربعات الانحرافات عن المتوسط.

البيانات غير المبوبة *Ungrouped Data*

يتم حساب الانحراف المعياري للبيانات غير المبوبة للمجتمع والذي يرمز له بالرمز σ (سيجما) باستخدام المعادلة رقم (2.13) :

$$\sigma = \sqrt{\frac{\Sigma (X - \mu)^2}{N}} \tag{2.13}$$

ويكون التباين σ^2 هو مربع الانحراف المعياري. ولكن من الناحية الرياضية يتم إيجاد التباين أولاً ثم حساب الانحراف المعياري بأخذ الجذر التربيعي للتباين إذا كان الهدف حساب الانحراف المعياري. ويمتاز الانحراف المعياري بأنه أكثر مقاييس التشتت استخداماً ؛ نظراً لأن الوسط الحسابي أكثر مقاييس النزعة المركزية استخداماً والذي يستخدم معه الانحراف المعياري إضافة إلى سهولة حساب التباين والتعامل معه رياضياً واستخدام تلك الحسابات في عمليات حسابية أخرى.

ويرمز لتباين العينة بـ S^2 ، وهو عبارة عن تقدير لتباين المجتمع σ^2 ، ويحسب من بيانات العينة المسحوبة من المجتمع. وكما هو معلوم فإنه في المعاينة المتكررة فإن تباين العينة يكون متحيز وأقل من تباين المجتمع بقيمة ثابته تساوي $(n-1)/n$. وبناء

على ذلك ، فإن معظم الإحصائيين عدّل معادلة تباين العينة بالقسمة على $(n-1)$ بدلاً من n حتى يتم التخلص من أثر التحيز ويحسب تباين العينة باستخدام المعادلة رقم (2.14) والتي يمكن صياغتها كالتالي :

$$S = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n-1}} \qquad (2.14)$$

مثال :

افترض أنه في إحدى مؤسسات الأعمال الزراعية الكبيرة تم اختيار سجل الغياب بسبب المرض في منتصف السنة لعدد من الموظفين وتسجيل عدد أيام غيابهم عن العمل فكانت على النحو التالي: ٧ ، ١٤ ، ٨ ، ٥ ، ١٥ ، ١١

المطلوب حساب المتوسط الحسابي والانحراف المعياري لهذه البيانات .

لحساب الانحراف المعياري للعينة يتم :

١- حساب المتوسط الحسابي للعينة $\bar{X}$ بإيجاد حاصل جمع البيانات Σx ثم قسمتها على عددها :

$$\bar{X} = \frac{\sum x}{n} = \frac{7+14+8+5+15+11}{6} = \frac{60}{6} = 10$$

٢- نحسب الانحراف للبيانات عن المتوسط $(X-\bar{X})$ لجميع المشاهدات مع ملاحظة أن مجموعها يساوي صفر:

$$\Sigma(X-\bar{X}) = 0$$

٣- نحسب مربع الانحراف $(X-\bar{X})^2$ لجميع المشاهدات ثم نوجد مجموعها $\Sigma(X-\bar{X})^2$ ويوضح الجدول رقم (٢,١١) نتائج تلك الحسابات.

الجدول رقم (٢,١١). بيانات توضح طريقة حساب الانحراف المعياري للعينة.

عدد أيام الغياب x	$(X - \overline{X})$	$(X - \overline{X})^2$
٧	- ٣	٩
١٤	٤	١٦
٨	- ٢	٤
٥	- ٥	٢٥
١٥	٥	٢٥
١١	١	١
٦٠	٠	٨٠

ولحساب التباين للعينة يتم التعويض في المعادلة التالية :

$$S^2 = \frac{\Sigma (x - \overline{x})^2}{n - 1} = \frac{80}{6 - 1} = \frac{80}{5} = 16$$

ثم إيجاد الانحراف المعياري $S = \sqrt{S^2} = \sqrt{16} = 4$

وكما هو ملاحظ سهولة العمليات الحسابية في هذا المثال ؛ نظراً لأن قيم الانحراف عن المتوسط عبارة عن أعداد صحيحة حيث تكون الحسابات أكثر تعقيدا في حالة وجود قيم كسرية.

وعليه فإنه يمكن استخدام صيغة رياضية أخرى لحساب الانحراف المعياري (المعادلة رقم (2.15)) بحيث لا نستخدم الانحراف عن المتوسط في الحسابات .

$$S = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n - 1}} \qquad (2.15)$$

ولحساب الانحراف المعياري S في المثال السابق يجب حساب قيم X^2 للمتغير كالتالي :

X^2	X
٤٩	٧
١٩٦	١٤
٦٤	٨
٢٥	٥
٢٢٥	١٥
١٢١	١١
٦٨٠	٦٠

وبالتعويض في المعادلة (2.15) بقيم $\sum x^2$, $\sum x$, n ثم حل المعادلة :

$$S = \sqrt{\frac{680 - \frac{(60)^2}{6}}{6 - 1}} = 4$$

وهي نفس النتيجة التي تم الحصول عليها باستخدام الصيغة الرياضية السابقة (2.14).

البيانات المبوبة Grouped Data

لحساب الانحراف المعياري للبيانات المبوبة في شكل التوزيعات التكرارية يتم استخدام التكرارات f في الحساب كأوزان. لذا يتم تعديل جميع الصيغ بحيث تشمل الأوزان. وتوضح المعادلات رقم (2.16) و (2.17) الصيغ الرياضية لحساب الانحراف المعياري لبيانات العينة المبوبة على شكل توزيعات تكرارية حيث يتم استخدام الفروق في المعادلة رقم (2.16) و التجميع في المعادلة رقم (2.17). ولتوضيح كيفية حساب S بمثال نعود مرة أخرى لبيانات أجور العمال الزراعيين الموضحة في الجدول رقم (٢,١٢). ويلاحظ أن الجدول إضافة لبيانات التوزيع التكراري الأساسية يحتوي على أعمدة

تشمل X ، fX ، X^2 و fX^2 والتي نحتاج لها للتعويض في المعادلات.

$$s = \sqrt{\frac{\sum f\ (X - \bar{X})^2}{\sum f - 1}} \tag{2.16}$$

$$s = \sqrt{\frac{\sum f\ X^2 - \frac{\left(\sum fX\right)^2}{\sum f}}{\sum f - 1}} \tag{2.17}$$

ويجب أن نضرب كل قيمة لـ X بتكرارها f ثم نقوم بتجميع تلك القيم للحصول على المجموع $\sum fX$ ، وأيضا نربع قيم X ثم نضربها بالتكرارات للحصول على $f X^2$ ثم نقوم بتجميع تلك القيم للحصول على المجموع $\sum f X^2$. والقيمة الأخيرة التي نحتاجها هي $\sum f$ والتي يتم الحصول عليها بتجميع عمود التكرارات. وبالتعويض عن القيم في المعادلة أعلاه نحصل على:

$$s = \left[\left(461950 - 3710^2 / 30\right)/(30 - 1)\right]^{0.5} = [108.5]^{0.5} = 10.4$$

الجدول رقم (٢,١٢). حسابات الانحراف المعياري للتوزيع التكراري لأجور المزرعة.

الأجور الأسبوعية	التكرارات f	مراكز الفئات X	X^2	fX	fX^2
١٠٠ حتى ١١٠	٣	١٠٥	١١٠٢٥	٣١٥	٣٣٠٧٥
١١٠ حتى ١٢٠	٧	١١٥	١٣٢٢٥	٨٠٥	٩٢٥٧٥
١٢٠ حتى ١٣٠	١٣	١٢٥	١٥٦٢٥	١٦٢٥	٢٠٣١٢٥
١٣٠ حتى ١٤٠	٥	١٣٥	١٨٢٢٥	٦٧٥	٩١١٢٥
١٤٠ حتى ١٥٠	٢	١٤٥	٢١٠٢٥	٢٩٠	٤٢٠٥٠
المجموع	٣٠	-	-	٣٧١٠	٤٦١٩٥٠

خصائص الانحراف المعياري S و المتوسط $\overline{X}$

يوجد خاصيتان للمتوسط والانحراف المعياري جديرة بالملاحظة خاصة عندما يجب أن نعمل تعديل للبيانات. فمثلاً لو رغبنا إضافة رقم ثابت لكل عنصر في مجموعة البيانات ماذا سيحدث لقيمة S ؟ وفقاً للخاصية الأولى فإن قيمة المتوسط تزيد بمقدار مساوٍ لقيمة الثابت في حين أن قيمة الانحراف المعياري لاتتغير بإضافة رقم ثابت لكل قيمة[١]. الخاصية الثانية تنص على أنه في حالة ضرب كل قيمة من القيم برقم ثابت فإن المتوسط يساوي حاصل ضرب القيمة لذلك الرقم في المتوسط والانحراف المعياري يساوي حاصل ضرب القيمة المطلقة للرقم الثابت في الانحراف المعياري[٢] في حين أن التباين يساوي حاصل ضرب مربع ذلك الرقم في التباين. ونجد أن تلك الخصائص مفيدة ؛ نظراً لعدم الحاجة لحسابات حقيقية للمتوسط أو التباين للبيانات المعدلة. حيث يتم استخدام الخاصية للمتوسط أو الانحراف المعياري التي تم حسابها من البيانات الأصلية للحصول على القيم المطلوبة للبيانات المعدلة.

ويمكن أيضاً معايرة البيانات بتحويل المتوسط لقيمة تساوي الصفر والانحراف المعياري لقيمة تساوي الواحد. وللحصول على متوسط يساوي الصفر نطرح المتوسط من كل قيمة من القيم في مجموعة البيانات على حدة. ولذلك يكون متوسط البيانات المعدلة مساوٍ للصفر ولكن الانحراف المعياري كما هو دون تغيير. ولتغيير الانحراف المعياري ليصبح مساوٍ للواحد نقسم كل قيمة من البيانات على الانحراف المعياري S (نضربها بـ 1/S). وهذه العملية الحسابية تقسم المتوسط على S ، ولكن قسمة الصفر على رقم ما عدا الصفر نحصل على قيمة تساوي الصفر. ويشكل عام يتم التعبير عن البيانات المعيارية للمجتمع بالرمز z والذي يساوي $Z_i = (X_i - \mu)/\sigma$.

معامل الاختلاف Coefficient of Variation

عندما نرغب في مقياس للاختلافات الموجودة في مجموعة البيانات فإننا نستخدم الانحراف المعياري مع المتوسط. ويتم تفضيل الانحراف المعياري ؛ لأن له نفس وحدات المتوسط بخلاف التباين الذي يعبر عن وحدات مربعة. ولكن في الغالب يتم حساب معامل الاختلاف لتحديد الاختلاف النسبي في مجموعة البيانات.وهو عبارة عن نسبة الانحراف المعياري إلى المتوسط مضروباً في ١٠٠ ، أي $V = (S/\bar{X})(100)$. ولذلك فإن معامل الاختلاف يعبر عن مدى كبر الانحراف المعياري مقارنة بالمتوسط على شكل نسبي. فإذا كانت قيمة V تساوي ١٠٠ فإنه يشير إلى تساوي الانحراف المعياري والمتوسط. وفي هذه الحالة فإن الاختلاف في البيانات كبير والمتوسط في هذه الحالة ليس المقياس المناسب لمركز التوزيع. وكلما كانت قيمة معامل الاختلاف V صغيرة كان المتوسط ممثل جيد لمجموعة البيانات. وكقاعدة فإننا نستخدم الحذر عندما تكون قيمة معامل الاختلاف V أعلى من ٥٠ ٪ حيال عرض البيانات باستخدام المتوسط.

من جهة أخرى لا يمكن مقارنة الانحرافات المعيارية لتوزيعين ؛ نظراً لارتباط قيم S بقيم المتوسطات لتلك التوزيعات.أي ، أنه لا يوجد تفسير للانحراف المعياري بدون المتوسط (ربما يستثنى من ذلك بعض الحالات غير المعتادة عندما تكون المتوسطات متساوية). وبناء على ذلك يتم مقارنة قيم V للتوزيعات بدلاً من ذلك. وكلما كانت قيمة V للتوزيع قليلة كان التوزيع أقل اختلاف.

ملاحظات ختامية

(١) إذا كان $s^2 = \sum\left(X - \bar{X}\right)^2/(n-1)$ وتم إضافة ثابت k لـ X بحيث يكون لدينا X+k فإن :

$\overline{X} = \sum(X+k)/n = \sum X/n + nk/n = \overline{X} + k$ وبنفس الطريقة يمكن التعويض في التباين كالتالي:

$$s^2 = \sum\left[(X+k)-(\overline{X}+k)\right]^2/(n-1) = \sum\left[X+k-\overline{X}-k\right]^2/(n-1) = \sum\left(X-\overline{X}\right)^2/(n-1)$$

وهي نفس الصيغة الأساسية للتباين.

(٢) إذا كان لدينا kX فإن المتوسط يكون:

$\overline{X} = \sum kX/n = k\sum X/n = k\overline{X}$ والتباين يمكن التعبير عنه بـ:

$$s^2 = \sum\left[kX - k\overline{X}\right]^2/(n-1) = \sum[k^2X^2 - 2kX\overline{X} + k^2\overline{X}^2]/(n-1) =$$

$$\sum\left[k^2\left(X^2 - 2X\overline{X} + \overline{X}^2\right)\right]/(n-1) = k^2\left[\sum X^2 - 2X\overline{X} + \overline{X}^2\right]/(n-1) = k^2\sum(X-\overline{X})^2/(n-1) = k^2S^2$$

وبذلك يكون الانحراف المعياري مساوي kS .

تمارين Exercises

١- لبيانات العينة التالية (-١٠ ، ٢ ، ٣ ، ٢ ، -٤ ، ٢ ، ٥) أوجد ما يلي:

أ) المتوسط الحسابي.

ب) الوسيط.

ج) المنوال.

د) متوسط المدى.

هـ) المدى.

و) الانحراف المعياري.

٢- للبيانات في التمرين ١ احسب الربيعي الأول والثالث ثم الانحراف الربيعي.

٣- إذا أعطيت مجموعة الأرقام التالية: ٨ ، ٥ ، ٢ ، ٦ ، ٤ ، ٥ المطلوب:

أ) أوجد المتوسط والوسيط والمنوال لهذا المجتمع.

ب) أوجد التباين والانحراف المعياري.

ج) احسب المدى ومتوسط المدى.

٤- البيانات التالية توضح الراتب السنوي عند التعيين لعدد ١٦ خريج حديث من كلية الزراعة:

٢٦٥٠٠دولار	١٩٩٠٠دولار	٣١٢٠٠دولار	٣١٤٠٠دولار
٢٠٤٠٠دولار	٢١٤٠٠دولار	٢١٨٠٠دولار	٢٥٥٠٠دولار
٢٤٦٠٠دولار	٢٢٦٠٠دولار	٢٤٨٠٠دولار	٢٧٠٠٠دولار
٢٣٦٠٠دولار	٢٨٤٠٠دولار	٢٣٤٠٠دولار	٢٩١٠٠دولار

أ) احسب المتوسط الحسابي والوسيط لهذه البيانات.

ب) أستخدم فئات بطول ٢٥٠٠دولار لإيجاد التوزيع التكراري للبيانات. ثم أوجد المدرج التكراري لهذا التوزيع وارسم المضلع التكراري بتوصيل مراكز الفئات (إبداء بالفترة ١٩٠٠٠دولار-٢١٥٠٠دولار).

ج) أوجد منحى التكرار النسبي المتجمع الصاعد ثم ارسمه. كم نسبة الخريجين الذين رواتبهم أقل من ٢٤٠٠٠دولار؟ أكثر من ٢٩٠٠٠دولار.

د) استخدم التوزيع التكراري الذي تحصلت عليه في الجزء ب لحساب المتوسط الحسابي والانحراف المعياري. كم نسبة الرواتب الواقعة في المدى واحد انحراف معياري للمتوسط؟ اثنين انحراف معياري ؟

هـ) كوّن التوزيع التكراري للجزء ب ثم احسب الوسيط ، الانحراف الربيعي والمنوال.

٥- البيانات التالية توضح التوزيع التكراري لكمية السعرات الحرارية في حليب يحتوي على نسبة دهون ٤٪ لعينة مأخوذة من ١٠٠ بقرة :

السعرات الحرارية	التكرار
من ٩٠ إلى أقل من ١٠٠	١٢
من ١٠٠ إلى أقل من ١١٠	٥٥
من ١١٠ إلى أقل من ١٢٠	٢٥
من ١٢٠ إلى أقل من ١٣٠	٨
الإجمالي	١٠٠

أوجد:

أ) المتوسط.

ب) الانحراف المعياري.

ج) الوسيط

د) المنوال.

هـ) ارسم المدرج التكراري للتوزيع التكراري ثم علق على تماثله.

٦- كيف تكون العلاقة بين المتوسط والوسيط والمنوال في حالة التوزيعات التكرارية المتماثلة؟ وكيف تكون تلك العلاقة في حالة التوزيعات التكرارية الملتوية لليمين؟ الملتوية لليسار؟ ارسم عدة توزيعات لتوضيح إجابتك.

استخدام برنامج أكسل أو أي برنامج حاسوبي مناسب لحل التمرين التالي:

٧- البيانات التالية توضح الوزن المكتسب خلال ٢٠٥ أيام لعدد ٢٤ من عجول التغذية نوع البرانقس brangus السوداء .

٤٨٢	٥٠٥	٤٦٧	٥٢١
٥٥٠	٥٣٤	٤٨٥	٥٤٢
٤٧٠	٥١١	٤٩٠	٥١٧
٥٤٥	٤٧٦	٤٨٧	٥٣٠
٤٩٦	٥٠٤	٥٥٨	٥٣٦
٤٦٣	٤٧٠	٥٥٣	٥١٢

احسب:

أ) المتوسط والانحراف المعياري للبيانات الخام باستخدام الأدوات ثم تحليل البيانات ثم أمر الإحصاء الوصفي.

ب) التوزيع التكراري بفترات مداها ٢٠ رطلاً مبتدئاً بـ ٤٦٠ ومنتهياً بـ ٥٦٠ باستخدام أدوات، تحليل البيانات، أمر المدرج التكراري.

ملاحظة: في العمود القريب أنشئ القيم ٤٦٠، ٤٨٠، ٥٠٠، ٥٢٠، ٥٦٠، قبل الضغط على أدوات في قائمة الأوامر. أدخل مدى هذا العمود في جدول تعريف المدرج التكراري الذي يظهر لك. أيضاً أشر على الصندوق الذي أمام الرسم؛ وذلك لرسم البيانات.