صف توزيعات القياسات

Describing Distribution of Measurements

مقدمة : Introduction

اشتقت كلمة الإحصاء من اللفظ اللاتيني " ستاتوس – Status" بمعنى الدولة . وقد استعمل علم الإحصاء قديما استعمالات مبكرة تضمنت تجميع البيانات والتخطيط ووصف مظاهرة متعددة للدولة.

في عام 1662 قام العالم " جون جرونت " بنشر معلومات إحصائية حول المواليد والوفيات ، ثم تلي عمل جرونت بكثير من الدراسات حول الوفيات ونسب ومعدلات الأمراض وأحجام السكان والدخول ونسب البطالة .

تعتمد المجتمعات الكبيرة والحكومات على الدراسات الإحصائية كموجة أو مرشد في عمليات الدراسة وأخذ قرارات مستقبلية معينة وإصدار توجيهات خاصة ببعض المشاكل ’ على سبيل المثال ، تقدير حجم البطالة وتقدير حجم التضخم وفهرسة الاستهلاك ومعدلات المواليد والوفيات وأسباب الضعف في العملية التعليمية والاقتصادية .

1-2 الهدف من علم الاحصاء The Objective of Statistics

1- يحتوى علم الإحصاء على عملية تجميع البيانات والمعلومات (Sampling) من محتوى كبير للبيانات يسمى بالمجتمع (Population).
2- المجتمع (Population) هو المجموعة الكاملة من العناصر موضع الدراسة (يقوم الإحصائي – الباحث بتعيينها على حسب الخاصية المراد دراستها).

مثال : (مجموعة الطلبة في فرقة دراسية معينة – مجموع الدرجات طلاب الفرقة الأولى – مجموعة طالبات فرقة معينة- ...).

3- الاستقراء أو الاستدلال الإحصائي (Statistical Inference) حول خاصية معينة عن المجتمع يمكن الحصول عليها من البيانات والمعلومات الموجودة داخل العينة (Sample) الممثلة للمجتمع .

4- العينة (Sample) هي مجموعة جزئية من المجتمع موضع الدراسة .

5- الاستقرار أو الاستدلال الإحصائي يحتوى إما على تقدير (Estimate) أو قرار (أحكام – تعميمات) (Decision) حول خاصية معينة عن المجتمع.

6- المعلم (Parameter) هو قياس عددي يوضح خاصية معينة عن المجتمع ويعرف باسم (إحصائي مجتمع).

7- الإحصائي (Statistic) هو قياس عددي يوضح خاصية معينة عن العينة ويعرف باسم (إحصائي عينة).

<u>مثال</u> :

في تجربة لمعرفة أثر استخدام الحاسب الألى وتطبيقاته على عملية تدريس مناهج اللغة الانجليزية في المراحل الابتدائي . أجريت بعض الاختبارات علا طلاب عشرة مدارس في هذه المرحلة وتم تسجيل نتائج هذه الاختبارات.

القياسات (نتائج الاختبارات) التي تم تسجيلها عن طلاب المدارس العشرة تمثل عينة (Sample) أخذت من مجتمع (Population) الدراسة (المدارس التي تستخدم الحاسب الألى في عملية تدريس اللغة الانجليزية في المرحلة الابتدائية).

البيانات الموجودة في هذه العينة يمكن استخدامها لعمل استقراء أو استدلال حول خاصية معينة عن المجتمع موضوع الدراسة.

8-علم الإحصاء (Statistics) مرتبط بنظرية المعلومات (Theory of Information وتطبيقاتها في عملية اتخاذ القرارات (الاستقراء – الاستدلال الإحصائي) مستخدما البيانات المتوفرة بالعينة.

2-2 عناصر المشكلة الإحصائية

The Elements of a Statistical Problem

علمنا من البند السابق أن الهدف من علم الإحصاء هو عملية الاستدلال حول خاصية معينة عن المجتمع وهذه العملية تعتمد على ثلاثة عناصر سنسميها بعناصر المشكلة الإحصائية.

1. الخطوة الأولى في المشكلة الإحصائية:

أ) هي عملية الدراسة حول أكثر الطرق اقتصادا للحصول على كمية معينة من البيانات والمعلومات وتسمى هذه الطريقة (بعملية اختيار العينة أو المعاينة (Sampling) procedure (technique) أو تخطيط أو تصميم التجربة (Design of the experiment).

ب)تكلفة الحصول على الكمية المطلوبة من البيانات تتغير كثيرا تبعا للطريقة المتبعة لجمع هذه البيانات.

ج) من أهم وسائل جمع البيانات (الاستمارات البحثية – أسلوب الملاحظة – أسلوب الوسائل الموضوعية).

د) وتعتبر أهم مصادر جمع البيانات هي:

- المصادر الميدانية (ميدان الظاهرة موضع الدراسة)
- المصادر التاريخية (هيئة بيانات – مؤلفات ودوريات علمية - ...)

2. الخطوة الثانية في المشكلة الإحصائية:

تحتوى هذه الخطوة على عملية تحليل وتعميم البيانات(Extraction of the information) الموجودة في العينة المختارة (الحصول على أكبر كم من المعلومات من خلال بيانات معينة).

3 . الخطوة الثالثة في المشكلة الإحصائية:

تحتوى على عملية استخدام نتائج التحليل السابق لعمل استقراء استدلال إحصائي عن المجتمع موضوع الدراسة.

4. بعض الاستقراءات أو الاستدلالات الإحصائية ، كتقدير خاصية معينة عن المجتمع ، تكون دقيقة وقريبة من الواقع وبالتالي جيدة. والبعض الأخر بعيد عن الواقع وبالتالي سيئ . وعلى هذا فإنه لابد من تعريف قياس لجودة (Measure of goodness) صانع الاستقراء وبالتالي للقرار الناتج عنه.

5. مقياس الجودة أو الثقة (Measure of goodness or reliability) لعملية الاستقراء أو الاستدلال الإحصائي تكون دائما ضرورية لتعيين أهمية هذه العملية وكذلك حجمها العملي أو التطبيقي . وبالتالي يمكن اعتبار عملية صناعة الاستقراء أو الاستدلال كبر وتوكول ذو خطوتين (Two steps procedure)

الأولى : تتمثل في عملية اختيار أفضل الطرق لصناعة الاستقراء أو الاستدلال ثم صناعة هذا الاستقراء أو الاستدلال.

والثانية: تتمثل في إعطاء مقياس لجودة هذا الاستقراء

3-2 طبيعة القياسات The Nature of Measurements

تصنف القياسات أو البيانات إلى نوعين أساسيين هما:

أ) القياسات النوعية Qualitative Measurements

ب) القياسات الكمية Qualitative Measurements

القياسات النوعية (أو الطبقية أو الوصفية)

Qualitative (or Categorical or Attribute) Measurements

وهى التي يمكن تقسيمها أو فصلها إلى طبقات أو أنواع من القياسات يمكن التمييز بينها ببعض الخصائص غير عددية مثل الجنس ومستوى الدراسة ونوع المنتج من شركة ما للإلكترونيات (راديو – تلفزيون - .....)

القياسات الكمية Qualitative Measurements

تحتوى على مجموعة من الأرقام تمثل مقدار القياس أو الخاصية لشيْ معين مقارنة بمجموعته.

ومثال ذلك: عدد أجهزة الراديو التي تنتجها شركة ما للإلكترونيات في أشهر مختلفة.

وتصنف القياسات الكمية إلى:

1-**قياسات كمية منفصلة** وهى التي تأخذ قيما محددة كعدد المواد الدراسية في الفرقة الأولى أو عدد أفراد الأسرة أو ترتيب أخ بين أخواته وهكذا ..

2- قياسات كمية متصلة وهى التي تأخذ أي قيم داخل مدى معين أو داخل فترة معينة كالعمر والطول والوزن وهكذا....

وهناك طريقة أخرى لتصنيف القياسات وهى استخدام المستويات الأربعة الآتية للقياس :

1.المستوى الاسمي (التصنيفي) للقياس

Nominal Level of Measurement

ويتميز هذا النوع بالقياسات التي تحتوى على الأسماء ، العناوين ، أو الأصناف فقط. وهذا المستوى لا يمكن ترتيب القياسات بأي طريقة.

أمثلة:

أ- تصنيف الأفلام على حسب نوعها (كوميدي – رومانسي – تاريخي).

ب-تصنيف الأعضاء الذين تم حضورهم الاقتراع للتصويت على موضوع ما (45 عضوا ديمقراطيا – 80 عضوا جمهوريا – 90 عضوا مستقلا)

ج- الجنس (ذكور – إناث)

ويتضح من الأمثلة السابقة أن هذا المستوى يفتقر إلى عملية الترتيب كذلك إلى الدلالة العددية وبالتالي لايمكن إجراء أي عملية حسابية علية ، فعلى سبيل المثال لايمكن حساب الوسط Average ل 12فيلما كوميديا و 15 فيلما رومانسيا و9 أفلام تاريخية.

رغم إمكانية ارتباط الأعداد بالتصنيفات أحيانا وخاصة عند استخدام الحاسبات الآلية في التحاليل الإحصائية، إلا أن هذا المستوى يظل يفتقر إلى الدلالة العددية والعمليات الحسابية، فعلى سبيل المثال، إذا حدد العدد صفر ليشير إلى نوع العضو (ديمقراطيا)، العدد واحد ليشير إلى نوع العضو (جمهوريا)،والعدد اثنين ليشير إلى نوع العضو (مستقلا)،

فان الأعداد هنا تستخدم للدلالة على النوع أو الصنف فقط وينقصها خاصية الدلالة العددية (المعنى الكمي)، بمعنى أنة لايمكن القول أن اثنين > واحد أو صفر < واحد وكذلك الوسط الحسابي $\frac{\text{صفر} + 1 + 2}{3} = 1$ ليس له أي دلالة إحصائية.

1-المستوى الرتبي للقياس Ordinal Level of Measurement

ويتميز هذا النوع بأنه يحتوى على القياسات التي يمكن إجراء عمليات الترتيب عليها ، ولكن الفروق بين القياسات (الرتب) إما أن تكون لايمكن تعيينها (تحديد قيمتها) أو تكون لامعنى لها.

أمثلة :

أ- في عينة من منتج حجمها 36 ، ثم تصنيف "12" منتجا بحالة جيدة، "16" منتجا بحالة متوسطة ، "8" منتجات بحالة سيئة.

ب- تبعا للتقارير عن جودة أداء المهام في عمل معين تم تصنيف العامل س في المرتبة الثالثة ، العامل ص في المرتبة السابعة ، والعامل ع في المرتبة العاشرة .

فى المثال الأول لا يمكننا إعطاء مقياس معين للفرق بين "الجيد" و"المتوسط".

أما في المثال الثاني يمكننا تحديد الفرق بين الثالث والسابع ، لكن القيمة الناتجة "4" لا تعنى أي شيْ ، وليس لها أي دلالة عددية (بمعنى أن الفرق "4" بين 3،7 ليس بالضرورة أن يكون نفس الفرق "4" بين 7،11.

وعلى ذلك يمكن استخدام المعلومات الموجودة في هذا النوع من القياس للمقارنات النسبية (غير المطلقة) مع عدم التأكد من تساوى الفرق بين القياسات (الرتب) المتتالية.

1-المستوى الفئوي (أو الفترى) للقياس

Interval level of Measurement

ويشبه هذا المستوى إلى حد كبير المستوى الرتبى مع تمييزه بخاصية إضافية وهى إمكانية تحديد الفروق بين القياسات ومعرفة دلالتها ولا يوجد لهذا المقياس صفرا أو (نقطة بداية) محددة بل تكون دائما اختيارية أو افتراضية.

أمثلة :

أ) الأجسام التي درجة حرارتها 98.2°، 98.6° فهر نهايت . تتبع المستوى الفئوي للقياس، ونلاحظ أن :

1- هذه القيم يمكن ترتيبها،

2- ويمكن تحديد الفروق بينها،

3- وعدم وجود صفر مطلق أو نقطة بداية طبيعية لهذه القياسات، فالقيمة صفر° فهر نهايت تبدو كنقطة بداية لكنها اختيارية أو افتراضية وكذلك لا تعنى عدم وجود حرارة أي لا تعنى غياب الخاصية.

4- من الخطأ أن نقول أن درجة الحرارة 50° فهرنهايت هي ضعف درجة الحرارة 25° فهر نهايت.

ب)مجموعة السنوات 1000، 2000،17776،1944،1995 تتبع المستوى الفئوي للقياس، مع ملاحظة عدم وجود نقطة بداية حيث أن الزمن لا يبدأ من السنة صفر ولكنها قيمة افتراضية أو اختيارية.

ج) درجات الحرارة المئوية للغرف في مكان ما تتبع المستوى الفئوي للقياس.

نلاحظ أنة من الصعب وجود خصائص أو صفات في مجال التربية والعلوم الإنسانية تتبع المستوى الفئوي للقياس، فليس منطقيا أن نقول أن 8-6 = 10-8 (حيث 10،8،6 درجات ثلاثة طلاب في مادة ما).

1- **مستوى النسبة للقياس** Ratio Level of Measurement

ويعتبر هذا المستوى تطويرا للمستوى الفئوي حيث إنه يحتوى على نقطة بداية طبيعية (الصفر المطلق) والذي يعنى غياب الخاصية ، كما أن الفرق والنسب بين القياسات في هذا المستوى لها دلالة ومعنى .

أمثلة: أ) أوزان البلاستيك المهملة من بعض الشركات .

ب)المسافات التي تقطعها السيارات في اختبار لاستهلاك الوقود .

نلاحظ أن القيم في هذه الأمثلة :

1- يمكن ترتيبها ،

2- يمكن حساب الفروق بينها ،

3- وجود صفر مطلقا (نقطة بداية طبيعية) .

والذي يعطى معنى للنسبة بين القياسات ، فقيمة الوزن 200 كجم هي ضعف قيمة الوزن 100 كجم بينما درجة الحرارة **50** ° فهرنهايت ليست ضعف درجة الحرارة **25** °

فهرنهايت كما ذكرنا سابقا .

والجدول الأتي يعطى ملخصا للمستويات الأربعة للقياس .

المستوى	ملخص	مثال
الاسمي	أصناف أو أنواع فقط . البيانات لايمكن ترتيبها بأي طريقة .	توزيع الناخبين : 45 ديمقراطيا 80 جمهوريا 90 مستقلا (أصناف فقط)
الرتبى	أصناف أو أنواع يمكن ترتيبها . الفروق إما أن تكون لا يمكن إيجادها أو تكون لا معنى لها .	توزيع الناخبين : 45 ناخبا ضعيف الدخل 80 ناخبا متوسط الدخل 90 ناخبا مرتفع الدخل (يمكن الترتيب على حسب ضعيف – متوسط - مرتفع)
الفترى أو الفئوي	الفروق يمكن الحصول عليها لكن لا يوجد صفر طبيعي أو (نقطة بداية) . النسب لا معنى لها .	درجة حرارة مجموعة من القضبان الحديدية : **45** ° فهر نهيت **80** ° فهر نهيت **90** ° فهر نهيت (**90** ° فهر نهيت ليست ضعف **45** ° فهر نهيت)
النسبي	يشبه الفترى ، لكن يوجد به صفر طبيعي (نقطة بداية) . النسب لها معنى .	أطول القضبان الحديدية : 45 سم 80 سم 90 سم (90 سم ضعف 45 سم)

4-2 طرق اخذ العينة (المعاينة)

Sampling Procedure (Sampling technique)

1- يتضح من الخطوة الأولى في البند 2.2 (عناصر المشكلة الإحصائية) ان معظم الأخطاء التي تظهر عن استخدام الإحصاء في تحليل البيانات وبالتالي فان عملية الاستقراء أو الاستدلال تنتج عن عملية تجميع البيانات بطريقة غير مناسبة.

2- عملية اخذ العينة (المعاينة) عادة تتطلب الكثير من الوقت، الجهد، والمال أكثر من عملية تحليل بيانات هذه العينة. ويمكن تقليل هذه الأشياء بعملية التخطيط الجيد والحذر لعملية المعاينة.

3- في عملية المعاينة يمكن مراعاة <u>الخطوات الأربعة الآتية:</u>

<u>أولا</u>:

يجب التأكد أن حجم العينة كاف للغرض المطلوب.

وهناك اعتقاد غير صحيح أن العينات الكبيرة تكون أفضل لكن هذه العينات الكبيرة تكون لا قيمة لها إذا أخذت بطريقة غير مناسبة وبدون دراسة مع كبر حجمها.

<u>ثانيا</u> :

في حالة حصولك على قياسات عن خاصية معينة (كالطول مثلا) مباشرة عن طريق السؤال **تأكد من حصولك على هذه القياسات** حيث أن عملية الأسئلة دائما ما ينتج عنها **خللا في القياسات لوجود نتائج غير متكافئة وبها الكثير من الانحرافات**.

<u>ثالثا</u> :

تحديد الطريقة المستخدمة للحصول على القياسات.

نجد أن أكثر هذه الطرق هي البريد – التليفون والمقابلات الشخصية على الرغم من وجود طرق كثيرة أخرى يمكن استخدامها.

فاستخدام البريد في الحصول على القياسات ينتج عنه بطء في الحصول على المعلومات وكذلك المقابلات الشخصية التي تستغرق الكثير من الوقت والتكلفة ، ولكنها تكون مطلوبة وضرورية إذا كانت القياسات أو المعلومة المطلوبة مفصلة أو معقدة ، أما المقابلات التليفونية فهي أكثر كفاءة واقل تكلفة بشكل نسبى.

رابعا :

تأكد من الطريقة التي تم استخدامها للحصول على العينة لكي تنتج عنها عينة ممثلة لمجتمع الدراسة.

وهناك نوعان من العينات هما:

أولا : العينات الاحتمالية Probabilistic Samples

ويتم اختيار هذا النوع باستخدام (مبادئ نظرية الاحتمالات) ومن أهم أنواعها:

(أ)- العينة العشوائية Random Sample

وفيها يتساوى جميع أفراد المجتمع في فرصة اختيارهم داخل العينة. وتعتبر العينة العشوائية عينة ممثلة لمجتمع الدراسة في جميع خصائصه ونادرا ما يحدث اختلاف بين (إحصائيات المجتمع) و (إحصائيات العينة) ويتم ذلك في حالة وجود انحرافات داخل العينة وذلك لعدم الحذر عند اخذ العينة وعدم اختيار الطريقة المناسبة للمعاينة.

مثال : استخدام الحاسب الألى في عملية تخليق أرقام التليفونات.

(ب)- العينة الطبقية Stratified Sample

ويتم في هذا النوع تقسيم المجتمع إلى مجتمعين جزئيين (طبقيين) على الأقل بشرط اشتراكهم فى نفس الخاصية، ثم يتم اخذ العينة من كل مجتمع جزئي (طبقة).

1- يجب اختيار العينة من كل من مجتمع جزئي (طبقة) بنسبة وجودها في مجتمع الدراسة.
2- من الخصائص التي يمكن تقسيم المجتمع عن طريقها إلى طبقات (مجتمعات جزئية) ، الجنس (ذكر– أنثى)، المستوى الدراسي (ابتدائي – إعدادي – ثانوي- .....)، الحالة الاجتماعية (متزوج – أعزب) ، الحالة الاقتصادية ، .... الخ.

(ج)- العينة المنتظمة Systematic Sample

ويتم في هذا النوع اختيار نقطة بداية ثم اختيار عناصر العينة بشكل دوري بداية من هذه النقطة كأن نختار كل خامس طالب إذا كان طول الدورة خمسة أو ثامن كتاب إذا كان طول الدورة ثمانية، وهكذا.

1- من أهم عيوب هذا النوع هو اختيار نقطة البداية التي تؤثر في تحديد قيم العناصر في العينة. ولهذا لا ننصح باستخدام هذا النوع في حالة ما إذا كانت عناصر المجتمع مرتبة بشكل دوري.
2- ومثال ذلك اختيار عينة من طلاب الفرقة الأولى بكلية التربية بحيث يتم اختيار ثالث شخص أثناء جلوس الطلبة في ترتيب (طالب ثم طالبة).

<u>(د)- العينة العنقودية</u> **Cluster Sample**

ويتم في هذا النوع تقسيم المجتمع إلى قطاعات أو أقسام تسمى عناقيد (ليست ذات خصائص مشتركة كما في حالة العينة الطبقية) ثم يتم الاختيار عشوائيا لعدد من هذه العناقيد ثم اختيار كل أعضاء هذه العناقيد في العينة.

1- هذا النوع من العينات يوفر الكثير من الجهد والتكاليف وخاصة إذا كان المجتمع مترامي الإطراف ومنتشرا على منطقة واسعة . ويكون أدق كلما تماثلت العناقيد المختارة.
2- خطأ المعاينة في هذا النوع من العينات يكون اكبر منه في العينات العشوائية والعينات الطبقية.
3- يستخدم هذا النوع كثيرا في الهيئات والحكومية وكذلك هيئات الأبحاث الخاصة.

<u>أمثلة</u> :

أ- الكليات تشكل عناقيد.

ب- الفرق من نفس المستوى في الكلية الواحدة تشكل عناقيد.

ج- سكان الأحياء المختلفة داخل مدينة أو قرية واحدة يشكلون عناقيد.

<u>(هـ)- العينة الجاهزة (الميسرة)</u> **Convenience Sample**

وفيها يتم استخدام النتائج المتاحة من قبل بسهولة وبسرعة.

1- في بعض الحالات تكون العينات الجاهزة جيدة وفى بعض الحالات تحتوى على انحرافات.

مثال :

من السهل أو الأكثر ملاءمة للمعلم أن يختار مجموعة الطلبة الذين يستخدمون يدهم اليسرى لتقييم الطلاب في الفصل ، على الرغم من أن هذه العينة ليست عشوائية ولكنها لا تحوى أي انحرافات.

2- استخدام إحدى الشركات للبيانات سابقة التجهيز من قبل إحدى الهيئات أو المؤسسات.

3- صغر حجم العينة من هذا النوع قد يعيبها ويؤدى إلى ظهور الانحرافات بها. كعملية الاستناد إلى عدد صغير جدا من استبيان تم طرحة (مثال: استخدام 5000تم إعادتهم من 100000) .

ثانيا : العينات غير الاحتمالية Non-probabilistic Samples

1- وهى العينات التي يتم أخذها تحت شروط ومواصفات أو معايير يراها الباحث لتحقيق غرض معين في التجربة.

وبذلك فإن هذا النوع لا يتبع نظرية الاحتمالات في الاختيار.

2- هذا النوع من العينات يكون أكثر فائدة من ناحية الحصول علية حيث أنة لا يحتاج إلى الجهد والتكاليف والوقت ، كما في الأنواع السابقة ، إلا إنه من الصعب تعميم نتائجه على المجتمع.

3- تنقسم العينات الاحتمالية إلى قسمين أساسين هما :

(أ)- العينة العرضية Accidental Sample

ويتم اخذ مثل هذه العينات عن طريق الصدفة . ولا يمكن تعميم نتائجها على مجتمع الدراسة .

مثال ذلك ، أن يقوم الباحث بتوزيع استبيان خاص به على العاملين أثناء مشاهدتهم لمباراة كرة القدم أو أثناء تناولهم وجبة الإفطار.

(ب)- العينة الهادفة (القصدية) **Purposeful Sample**

1- وفى هذا النوع يقوم الباحث باختيار أغراض العينة تحت شروط معينة لتحقيق الهدف أو الغرض من التجربة أو الدراسة.

2- ومثال ذلك ، إذا أراد باحث الكتابة عن حدث معين لم يعاصره كقيام احد الثورات أو استقلال احد الشعوب فإنه لابد من اختيار أفراد عينته من أشخاص قد عاصروا هذه الفترة وعلى قدر من الوعي والموضوعية ليحصل منهم على البيانات أو المعلومات التي يراها.

والجدول التالي يعطى ملخصا لأنواع العينات السابقة :

أولا العينات الاحتمالية :-

1- العينة العشوائية (الممثلة)	جميع عناصر المجتمع لها نفس الفرصة في الظهور في العينة. استخدام الحاسب الألى لتخليق أرقام التليفون.
2- العينة الطبقية	يقسم المجتمع إلى طبقتين على الأقل ثم نختار العينة من كل منهما.
3- العينة المنتظمة	نختار نقطة بداية من المجتمع ثم نختار العنصر الموجود على بعد ثابت من هذه النقطة.
4- العينة العنقودية	يقسم المجتمع إلى مساحات أو أجزاء ثم نختار عشوائيا بعض هذه المساحات ، ثم نختار جميع عناصرها بالعينة.
5- العينة الجاهزة (الميسرة)	استخدام البيانات والنتائج سابقة التحضير أو التجهيز.

ثانيا العينات غير الاحتمالية :-

1- العينة العرضية	يتم اختيارها عن طريق الصدفة.
2- العينة الهادفة (القصدية)	يتم اختيار أفراد العينة تحت شروط معينة لتحقيق الهدف من التجربة.

5-2 لماذا نصف مجموعات القياسات ؟

Why Describe Sets of Measurements?

1- بالرجوع إلى الهدف من علم الإحصاء ، صناعة الاستدلال الإحصائي أو الاستقراء الإحصائي ، عن خاصية معينة في مجتمع الدراسة مستخدمين البيانات الموجودة بالعينة المختارة.

2- لكن نقوم بعمل جملة استدلالية "Inferential Statement"

فإننا نحتاج إلى طريقة لوصف هذا المجتمع.

3- **طرق الوصف تنقسم إلى قسمين :**

أ) الطرق البيانية (الرسوم) Graphical Methods

ب) الطرق العددية (الحسابية) Numerical Methods

4- عند تحليل مجموعة كبيرة من القياسات ، فإننا نقوم أولا بترتيب وتلخيص هذه البيانات مستخدمين الجداول والأشكال البيانية.

5- الجدول التكراري يستخدم لسرد أو تلخيص القياسات داخل فترات (أو فصول) تبعا لعدد مرات (أو تكرار) وقوع البيان داخل كل فترة (أو فصل).

6- الحد الأدنى للفترة (أو الفصل) هو أصغر عدد تحتويه الفترة (أو الفصل).

7- الحد الأعلى للفترة (أو الفصل) هو اكبر عدد تحتويه الفترة (أو الفصل).

6-2 الطرق البيانية **A Graphical Method**

1- تستخدم الطريقة البيانية لعرض مجموعة القياسات في صورة أشكال أو رسومات بيانية تعطى الدارس أو الباحث وصفا مرئيا كافيا عن هذه المجموعة من القياسات.

2- سنقوم بشرح مفهوم الطريقة البيانية من خلال هذا المثال :

مثال: في اختبار لقياس درجة الذكاء يحتوى على 30 فقرة رصد درجات 25 طالبا على النحو التالي :

25	27	23	29	25
28	22	25	22	23
30	24	17	24	28
24	21	23	17	19
23	19	26	20	15

3- **أولا** : نوجه اكبر درجة (Highest Score) وهى 30 وكذلك اقل درجة (Lowest Score) وهى 15 واللتان تعطيان مؤشرا إلى أن مدى هذه القياسات (Range) هو 15 (اكبر درجة – اقل درجة).

4- لمعرفة كيفية توزيع درجات الطلبة بين الحد الأعلى للدرجات (أكبر درجة) والحد الأدنى للدرجات (اقل درجة) سنقوم بتقسيم هذه الفترة إلى فترات جزئية متساوية Equal Sub – Intervals أو فصول.

5- الفترة من 15 إلى 30 يمكن تقسيمها عادة إلى عدد (من 5 إلى 20) فترة جزئية أو فصل حسب عدد القياسات المتاحة . وعند تحديد عدد الفترات الجزئية أو الفصول يجب مراعاة أن (خطأ التجميع) يزداد بنقصان عدد الفترات الجزئية وبالمقابل فإن زيادة عدد الفترات الجزئية يعنى الاقتراب من التوزيع الأصلي للقياسات.

6- بفرض إننا نريد الحصول على 7 فترات جزئية أو فصول ، فإنه لابد من اختيار طول مناسب لهذه الفترات الجزئية ، Sub – Intervals وذلك بتقسيم مدى هذه القياسات على 7 (عدد الفترات الجزئية المراد الحصول عليها).

7- العدد الصحيح 2 (ناتج عملية القسمة) هو أقرب عدد صحيح يمكن استخدامة كطول للفترة الجزئية محققا شرط اختيار اكبر عدد من الفترات الجزئية أو الفصول.

8- نحدد النقط الحدية Boundary Points للفترات الجزئية على النحو التالي :

14.5, 16.5, 18.5, 20.5, 22.5, 24.5, 26.5, 28.5, and 30.5,

ونتأكد من إنه لا يوجد قياس أو مفردة من البيانات تقع على أي نقطة حدية ، مما يؤكد أن كل قياس أو مفردة يقع في فترة جزئية واحدة أو فصل واحد فقط. وتسمى النقاط المنصفة لهذه الفترات الجزئية بمراكز الفترات (Class midpoints) أو (Class marks) أي أن :

$$\text{مركز الفترة} = \frac{\text{حدها الادنى} + \text{حدها الاعلى}}{2}$$

9- نقوم بالمسح التدريجي لدرجات الطلاب ووضع إشارة (Tally) أمام الفترة التي تنتمي إليها كل درجة . ثم يتم ترجمة هذه الإشارات إلى أرقام تعبر عن تكرار كل فترة (أو فصل) (Class frequencies) وتوضع في عمود خاص بها في الجدول التكراري ، كما يمكن ترجمة هذه الإشارات إلى نسب تكرارية أو احتمالات (Relative frequencies) توضع أيضا في عمود خاص بها في الجدول التكراري كما هو موضح بالجدول التالي :

الفترة (الفصل)	حدود الفترة	التكرارات المطلقة		التكرار النسبي	التكرار التراكمي	التكرار التراكمي النسبي
		إشارة Tally	التكرار f_i			
1	14.5-16.5		1	1/25	1	1/25
2	16.5-18.5		2	2/25	3	3/25
3	18.5-20.5		3	3/25	6	5/25
4	20.5-22.5		3	3/25	9	9/25
5	22.5-24.5		7	7/25	16	16/25
6	24.5-26.5		4	4/25	20	20/25
7	26.5-28.5		3	3/25	23	23/25
8	28.5-30.5		2	2/25	25	25/25=1

10- (أ) العدد f_i من القياسات والذي يقع داخل الفترة الجزئية (أو الفصل) رقم i يسمى تكرار الفترة الجزئية (أو الفصل) رقم i (i-th Class Frquency)

(ب) النسبة $(\frac{f_i}{25})$ والتي تقع داخل الفترة الجزئية (أو الفصل) رقم i يسمى بالتكرار النسبي للفترة الجزئية (أو الفصل) رقم i ،

(Relative frequency in the i-th class)

(ج) إذا أعطيت n من القياسات فان التكرار النسبي للفترة الجزئية (أو الفصل) رقم i يعطى الشكل $(\frac{f_i}{n})$.

(د) للتأكد من صحة عملية الجدولة ، تذكر دائما أنة لأي عدد k من الفترات الجزئية أو الفصول

$$(i) \quad \sum_{i=1}^{k} f_i = n \qquad and \qquad (ii) \quad \sum_{i=1}^{k} \frac{f_i}{n} = 1$$

(هـ) بفرض وجود عينة حجمها n تحتوى على m من القيم العددية المختلفة

$$X_1, X_{2,}\ ....,\ X_m \quad (m \le n)$$

بتكرارات نسبية مناظرة

$$\tilde{f}_1, \tilde{f}_2, \tilde{f}_3, \tilde{f}_4, ..., \tilde{f}_m,$$

فإنه يمكن تعريف دالة التكرار (Frequency Function) لهذه العينة بالشكل التالي:

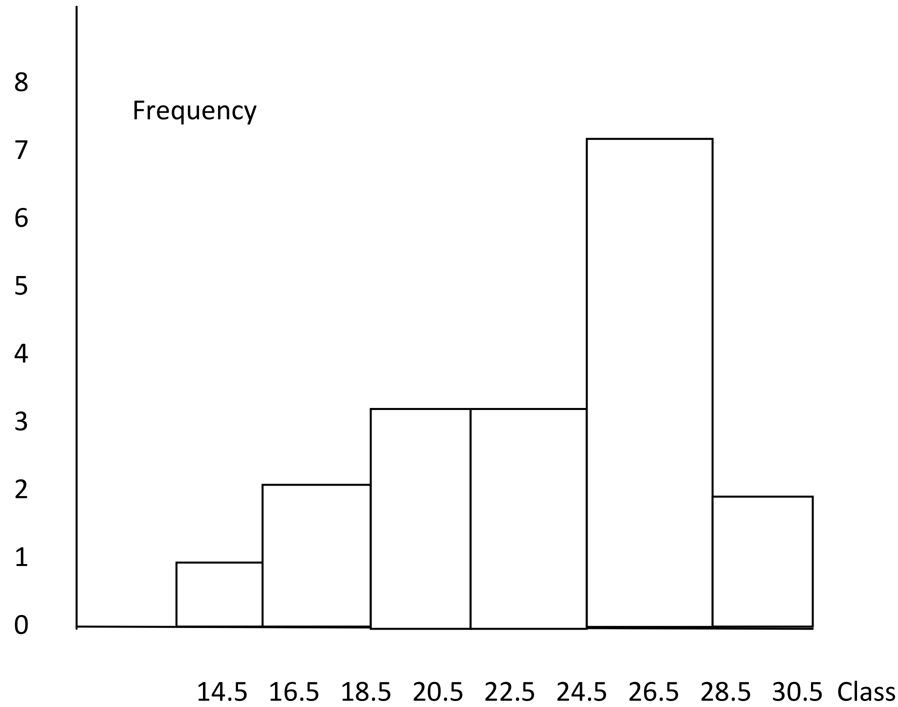
والتي تكافئ

<u>8-2 المدرجات التكرارية وشكل القياسات :</u>

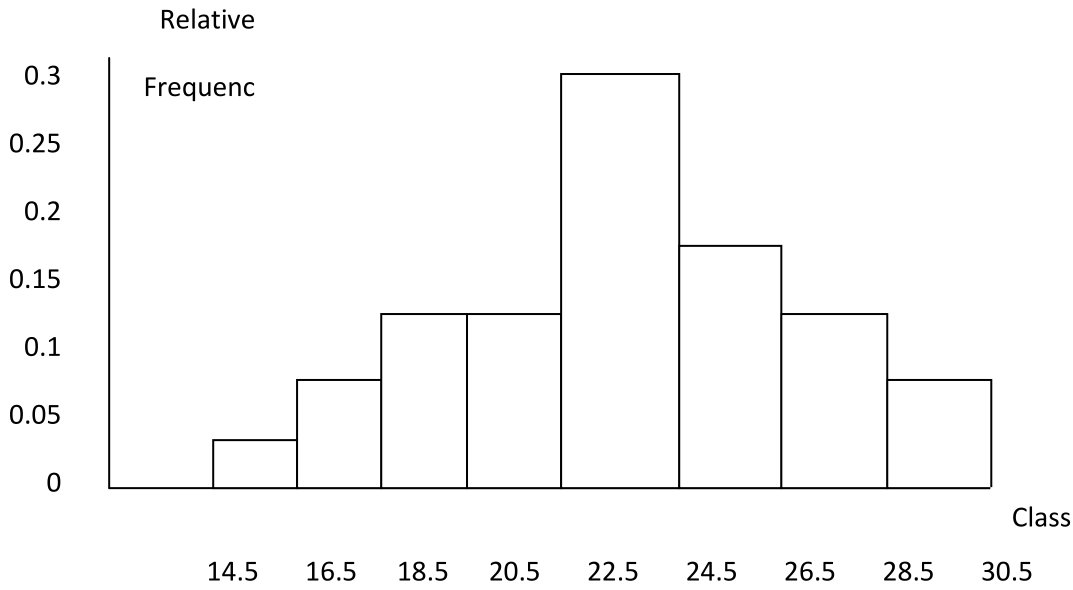
Histograms and The Shape of Measurements

-1 بعد إتمام عملية جدولة البيانات ، يمكننا استخدام المدرج التكراري (رسم التكرار ضد الفترات الجزئية) أو المدرج التكراري النسبي التكراري (رسم التكرار النسبي ضد الفترات الجزئية) لوصف البيانات ، نلاحظ أن المدرجين التكراريين متطابقان في ما عدا مقياس الرسم .

-2 أ) بدراسة المدرج التكراري للبيانات في مثال 6.2 (2) والموضح في الشكل التالي :



ب)بدراسة المدرج التكراري النسبي لنفس البيانات في مثال 6.2 (2) والموضح في الشكل التالي:



ج) يتضح من الدراسة المطلوبة في الفقرتين السابقين (أ ، ب) أن المدرجين متطابقان في ماعدا مقياس الرسم .

3- باستخدام المدرج التكراري النسبي في الفقرة (ب) افترض الأسئلة الآتية :

أ- ما هي نسبة الطلبة الذين حصلوا على درجات أقل من 20.5 ؟

6/25 or 24 %

ب- ما هي نسبة الطلبة الذين حصلوا على درجات أكبر من 26.5 ؟

5/25 or 20

ج- ما هي نسبة الطلبة الذين حصلوا على درجات بين 20.5 و 26.5؟

14/25 or 56 %

4-المدرج التكراري النسبي عادة يعرف باسم التوزيع التكراري لأنة (يعرض طريقة توزيع البيانات) على المحور الأفقي .

5-المستطيلات الموجودة أعلى الفترات الجزئية (أو الفصول) في المدرج التكراري النسبي تشير إلى :

أ- المستطيل أعلى الفترة الجزئية رقم i يمثل نسبة البيانات التي تقع داخل هذه الفترة الجزئية .

ب- المستطيل أعلى الفترة الجزئية رقم i يمثل أيضا احتمال أن القيمة المختارة عشوائيا من العينة تقع داخل هذه الفترة الجزئية .

6- مثال : أكمل العبارات الآتية مستخدما المدرج التكراري النسبي الموجود في الفقرة 2 (ب) :

أ- احتمال وقوع القيمة المختارة عشوائيا من البيانات داخل الفترة 20.5 إلى 24.5 هو .... (7/25)

ب-احتمال وقوع القيمة المختارة عشوائيا من البيانات داخل الفترة 26.5 إلى 28.5 هو .... (3/25)

ج- احتمال أن القيمة المختارة عشوائيا من البيانات تكون اكبر من 18.5 هو .... (22/25)

د- احتمال أن القيمة المختارة عشوائيا من البيانات تكون اقل من 24.5 هو .... (16/25)

هـ- احتمال أن القيمة المختارة عشوائيا من البيانات تكون اكبر من 18.5 واقل من 24.5 هو .... (13/25)

7- <u>مثال :</u> في إحدى المقابلات الشخصية لدراسة مدى نجاح تجربة معينة رصد الوقت المستغرق للرد على احد الأسئلة بالدقيقة لعينة مكونة من 25 طالب :

3.7	3.9	5.7	3.8	5.2
4.3	4.7	4.3	4.1	4.2
4.8	4.4	3.0	2.5	3.1
4.2	5.3	4.8	3.9	3.6
5.4	3.8	4.2	3.3	4.7

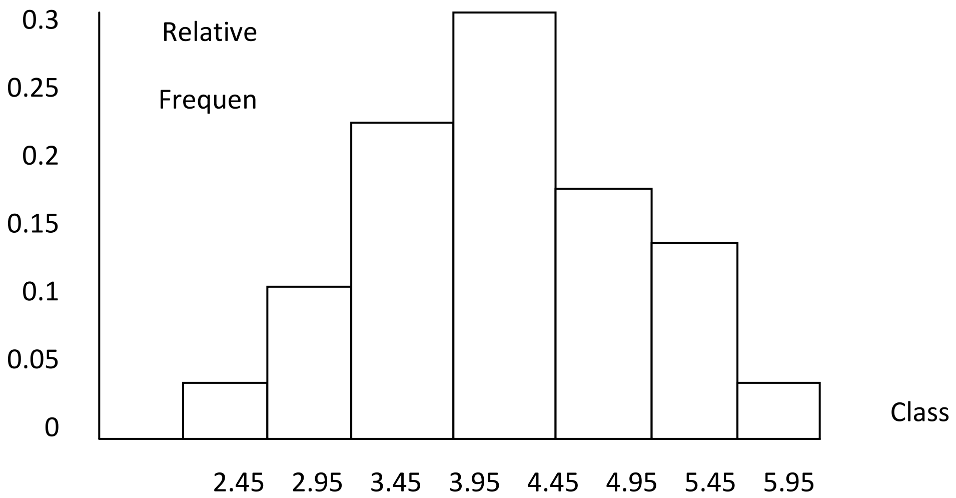
أ) ارسم المدرج التكراري النسبي لهذه العينة .

<u>ب) أكمل العبارات التالية مستخدما الفقرة (أ) :</u>

1- احتمال أن القيمة المختارة عشوائيا من العينة تكون اكبر من 4.45 هو .... (8/25)
2- احتمال أن القيمة المختارة عشوائيا من العينة تكون اقل من 3.45 هو .... (4/25)
3- احتمال وقوع القيمة المختارة عشوائيا من العينة في الفترة 34.5 إلى 4.45 هو .... (13/25)

الحل : أ) جدولة البيانات :

الفترة (الفصل)	حدود الفترة	التكرارات المطلقة		التكرار النسبي	التكرار المجمع	التكرار النسبي
		إشارة Tally	التكرار f_i			
1	2.45-2.95		1	1/25	1	1/25
2	2.95-3.45		3	3/25	4	4/25
3	3.45-3.95		6	6/25	10	10/25
4	3.95-4.45		7	7/25	17	17/25
5	4.45-4.95		4	4/25	21	21/25
6	4.95-5.45		3	3/25	24	24/25
7	5.45-5.95		1	1/25	25	25/25 = 1



2-10 الطريقة العددية :

Numerical Descriptive Measures

1- من أهم مميزات الطريقة البيانية لوصف القياسات هي عملية التمثيل المرئي للبيانات .
2- في كثير من الأوقات نحتاج إلى تقرير كتابي عن البيانات ، وفى هذه الحالة لا يمكن استخدام الطريقة البيانية للحصول على هذا النوع من التقارير .
3- أهم عيوب الطريقة البيانية لوصف القياسات يتمثل في أنها غير ملائمة لعملية صنع الاستدلال الإحصائي أو الاستقراء الإحصائي . نظرا لصعوبة إعطاء مقياس للجودة في حالة الاستدلال أو الاستقراء البياني .
4- مما سبق يتضح إننا نحتاج إلى الطريقة العددية لوصف القياسات وهى تتمثل في الحصول على مجموعة أو فئة من الأعداد تصف أو تصور التوزيع التكراري لهذه القياسات وتستخدم في نفس الوقت لصناعة الاستدلال أو الاستقراء خاصية معينة بالمجتمع .

2-11 القياسات العددية للنزعة المركزية :

Numerical Measures of Central Tendency

مقياس النزعة المركزية (مركز التوزيع التكراري) هو قيمة تتوسط البيانات. وتوجد طرق مختلفة لتعيين مركز مجموعة من القياسات وعلية توجد تعريفات مختلفة لمقياس النزعة المركزية تحتوى على سبيل المثال الوسط Mean ، الوسط Median ، المنوال Mode ، ونصف المدى Midrange

أ) الوسط الحسابي لمجموعة تحتوى على n من البيانات y1, y2, y3, ... , yn يعرف على أنة مجموع هذه البيانات مقسوما على عددها (n)

ب) الرمز $\bar{y}$ يستخدم للدلالة على الوسط الحسابي للعينة بينما μ (يقرأ ميوه) يستخدم ليشير إلى الوسط الحسابي للمجتمع .

ج) من أهم استخدامات الوسط الحسابي للعينة $\bar{y}$ في مجال صناعة الاستدلال الإحصائي هو استخدامه كتقدير للوسط الحسابي للمجتمع μ .

د) باستخدام رمز التجميع ، يمكن تعريف متوسط العينة بالصورة التالية :

$$\bar{y} = \frac{\sum_{i=1}^{n} (yi)}{n}$$

<u>هـ) مثال :</u>

اوجد الوسط الحسابي لمجموعة البيانات 13 ,11 ,10 ,7 ,5 ,2

<u>الحل :</u>

<u>أولا :</u> بما أن عدد القياسات n هو 6 نحسب

$$\sum_{i=1}^{6} y_i = 48 .$$

<u>ثانيا</u> : باستخدام رمز التجميع

$$\bar{y} = \frac{\sum_{i=1}^{6} y_i}{n} = \frac{48}{6}$$

<u>و) مثال</u> :

أوجد الوسط لمجموعة البيانات ,0.81 ,0.82 ,0.8 ,0.82 ,0.81 ,0.82 ,0.81 ,0.81 ,0.810.8

<u>الحل :</u>

أولا : بما أن عدد القياسات nهو 10 نحسب

$$\sum_{i=1}^{10} y_i = 8.11 .$$

ثانيا : باستخدام رمز التجميع

$$\bar{y} = \frac{\sum_{i=1}^{10} y_i}{n} = \frac{8.11}{10} = 0.811 .$$

يمكن أن تصبح العملية الأخيرة أسرع وأكثر سهولة إذا تم تجميع الأرقام المتشابهة معا ، أي أن

$$\bar{y} = \frac{1}{10}\left[(0.8 + 0.8) + \right.$$

$$(0.81+0.81+0.81+0.81+0.81)$$

$$\left.(0.82 + 0.82 + 0.82)\right]$$

$$= \frac{1}{10}\left(\underline{2}.0.8 + \underline{5}.0.81 + \underline{3}.0.82 = 0.811.\right)$$

ونستنتج من هذا المثال إنه يمكن حساب الوسط الحسابي للعينة باستخدام دالة التكرار النسبي $\tilde{f}(x)$ للعينة (بند 2-6 (10- ج)).

ز) بفرض وجود عينة حجمها n تحتوى على m من القيم العددية المختلفة

$$y_1, y_2, \dots\dots\dots\dots, y_m \qquad (m \le n)$$

بتكرارات نسبية مناظرة

$$\tilde{f}_1, \tilde{f}_2, \dots\dots\dots, \tilde{f}_m$$

وبتكرارات مناظرة

$$f_i = n\,\tilde{f}_i. \qquad (i = 1, 2, \dots\dots\dots, m)$$

فإن الوسط الحسابي للعينة يمكن حسابه باستخدام

$$\bar{y} = \frac{\sum_{i=1}^{m} y_{i.}\, n \tilde{f}_i}{n} = \frac{\sum_{i=1}^{m} y_{i.}\, \tilde{f}_i}{n}$$

ح) لإيجاد الوسط الحسابي لمجموعة من البيانات المبوبة (Mean Of The Grouped Data)

خلال جدول تكراري بطول فترة جزئية محددة ، فإن أسهل الطرق لإتمام هذه العملية هو تصميم جدول يلخص خطوات أو طريقة الحل .

بفرض أن x تمثل مركز الفترة الجزئية (أو الفصل) Class Mark، f تمثل التكرار Frequency داخل هذه الفترة الجزئية (أو الفصل).

ط) مثال : أوجد الوسط الحسابي لمجموعة البيانات المبوبة في الجدول التالي :

حدود الفترة	التكـــــرار
171-175	4
166-170	8
161-165	14
156-160	22
151-155	27
146-150	19
141-145	17
136-140	11
131-135	3

الحل : أولا : نوجد الجدول التالي :

الفترة i	حدود الفترة	مركز الفترة X_i	التكرار f_i	$X_i . f_i$
1	171-175	(171+175)/2 = 173	4	692
2	166-170	(166+170)/2 = 168	8	1344
3	161-165	(161+165)/2 = 163	14	2282
4	156-160	(156+160)/2 = 158	22	3476
5	151-155	(151+155)/2 = 153	27	4131
6	146-150	(146+150)/2 = 148	19	2812
7	141-145	(141+145)/2 = 143	17	2431
8	136-140	(136+140)/2 = 138	11	1518
9	131-135	(131+135)/2 = 133	3	399
Total	…………	…………..	125	19085

ثانيا : نوجد قيمة الوسط الحسابي :

$$\bar{x} = \frac{\sum_{i=1}^{9} x_i . f_i}{n} = \frac{19085}{125} = 152.68 .$$

2- أ) الوسيط أو القيم الوسيطية لمجموعة من البيانات عددها n ، $y_1, y_2,..., y_n$ هي تلك القيمة y التي تتوسط القيم بعد ترتيبها تبعا لقيمها .

ب) إذا كان عدد البيانات في المجموعة زوجيا even فإنه توجد قيمات تتوسط هذه البيانات ويكون الوسيط في هذه الحالة هو الوسط الحسابي لهاتين القيمتين .

ج) مثال : اوجد الوسيط لكل مجموعة من القياسات التالية :

1) 5 , 3 , 2 , 7 , 4

2) 10 , 8, 13 , 14 , 9 , 8

الحل : 1) أولا : بترتيب القياسات تبعا لقيمها نجد أنها تأخذ الصورة :

2 , 3 , 4 , 5 , 7

ثانيا : نجد أن الوسيط أو القيمة المتوسطة هي 4 .

2) أولا : بترتيب القياسات تبعا لقيمها نجد أنها تأخذ الصورة :

8 , 8, , 9 , 10 , 13 , 14

ثانيا : وحيث أن عدد قياسات هذه المجموعة n هي 6 (زوجي) فإن الوسيط أو القيمة المتوسطة تعطى بالشكل : $\frac{9+10}{2} = 9.5$

د- مثال : أوجد الوسط الحسابي والوسيط لكل مجموعة من القياسات التالية :

. 5 , 7 , 8 , 10, , 10 , 11 , 13 , 14

الحل : أولا : لإيجاد $\bar{y}$ الوسط الحسابي للعينة ، نحسب . $\sum_{1=1}^{8} y_i = 78$ ثم نوجد

$$\bar{y} = \frac{\sum^{8} y_i}{i = 1} = \frac{78}{8} 9.75 \quad .$$

ثانيا: لإيجاد الوسيط أو القيمة المتوسطة نلاحظ أن القيم مرتبة ولا تحتاج إلى إعادة ترتيب وكذلك نلاحظ أن عددها زوجي وبالتالي نجد أن الوسيط لهذه المجموعة هو $10 = \frac{10+10}{2}$.

3- أ) المنوال Mode لمجموعة من البيانات هي القيمة صاحبة أكبر تكرار .

ب)إذا احتوت المجموعة على قيمتين لهما أكبر تكرار فإنه يقال أن مجموعة البيانات في هذه الحالة ثنائية المنوال Bimodal .

ج)أما إذا احتوت مجموعة البيانات على أكثر من قيمتين لهم أكبر تكرار فإنه يقال أن مجموعة البيانات في هذه الحالة متعددة المنوال Multimodal .

د) إذا لم تحتو مجموعة البيانات على أي تكرار فإن المجموعة في هذه الحالة لا تحتوى على منوال No Mode .

مثال:

أ) 5 , 3 , 5 , 4 , 1 ,5, 1 , 3 , 5 , 5 ,5

ب) 9 , 7 , 6 , 6 , 6, 5 , 4 , 3 , 2 ,2, 2 , 1

ج) 10 , 9 , 8 , 7 , 6, 3 , 2 , 1

الحل : أ) العدد 5 هو المنوال حيث أن له اكبر أكبر تكرار (5 مرات) .

ب)العددان 2 , 6 يمثلان المنوال حيث أن لكل منهما أكبر تكرار (3 مرات) .

ج) لا يوجد منوال لهذه المجموعة حيث أنة لا يوجد تكرار لأي من القياسات .

5- أ) نصف المدى midrange هو القيمة التي تتوسط أكبر وأصغر القياسات .

ب)نصف المدى = $\frac{\text{أكبر قيمة للقياسات + أصغر قيمة للقياسات}}{2}$

1-2 أفضل قياس للنزعة المركزية :

The Best Measure of central Tendency

تعتبر عملية اختيار أكثر الطرق ملاءمة لقياس النزعة المركزية عملية غير بسيطة ولا توجد طريقة معينة لتحديد أكثر الطرق ملاءمة لجميع أنواع البيانات .

طرق القياس المختلفة للنزعة المركزية لها الكثير من المميزات والعيوب والجدول التالي يوضح أهم مميزات وعيوب القياسات المختلفة للنزعة المركزية .

المعدل المتوسط	تعريف	مدى استخدامه	إيجاده	تأثره بالقيم المتطرقة	تأثره بجميع القيم	مميزاته وعيوبه
الوسط الحسابي	$\bar{x} = \frac{\sum x}{n}$	الأكثر استخداما	دائما	نعم	نعم	يعمل بكفاءة مع جميع الطرق الإحصائية
الوسيط	القيمة التي تتوسط القيم	غالبا	دائما	لا	لا	غالبا ما يستخدم في حالة وجود قيم متطرقة
المنوال	القيم الأكثر تكرارا	أحيانا	أحيانا لا يوجد وأحيانا أكثر من واحد	لا	لا	صالح للبيانات من المستوى الاسمي
نصف المدى	$\frac{\text{الأكبر + الأصغر}}{2}$	نادرا	دائما	نعم	لا	يتأثر بالقيم المتطرقة

13-2 مقاييس الاختلاف أو التشتت

Measures of Variability

1- في البند السابق تعرفنا على طرق إيجاد مقاييس النزعة المركزية والآن نتعرف على مقاييس الاختلاف أو التغير Variability أو التشتت Dispersion لمجموعة البيانات .

2- تأتى أهمية مقاييس الاختلاف لعدم كفاءة مقاييس النزعة المركزية وحدها لوصف مجموعة البيانات .

3- بفرض المجموعتين الآتيتين من البيانات :

$x_1 = 9$

$x_2 = 10$ $\qquad \bar{x} = 10$

$x_3 = 11$

$y_1 = 1$

$\bar{y}=10$ $\qquad$ $y_2 = 10$

$y_3 = 19$

ج) كل من المجموعتين له نفس الوسط الحسابي والذي يساوى 10 ، على الرغم من أن المجموعة الثانية تحتوى على اختلاف (بعد قيم المجموعة عن الوسط الحسابي) أكبر من المجموعة الأولى .

4- قد يكون الاختلاف أو التغير مطلوبا أحيانا في بعض البيانات فعلى سبيل المثال فإن الشركات المنتجة لقطع الغيار لا تحتاج إلى تغير أو اختلاف في المنتج الواحد والذي يشير إلى عدم جودة الإنتاج ، بينما يتغير الحال في أي اختبار تربوي لاختيار بعض العاملين في مجال التدريس ، حيث أن الاختلاف أو التغير في نتائج الاختبار تتيح المجال الأكبر لعملية الاختبار الصحيحة .

5-سبق أن تعرفنا في البند 2-6 على أسهل الطرق لقياس الاختلاف ، وهو المدى ، والذي يعرف على أنة الفرق بين أكبر وأصغر البيانات .

6- مثال :

أوجد المدى لكل من :

أ)المجموعة

74	34	73	23
17	26	29	28
49	52	8	88
45	32	96	37
62	23	62	81

المدى = 96- 8 = 88

ب)المجموعة

8.8	6.7	7.1	2.9
9.0	0.2	1.2	8.6
6.3	4.6	2.1	8.8

المدى = 9 – 0.2 =8.8

7-الشكلان الآتيان يوضحان توزيعين لمجموعتين من البيانات وبالمقارنة نجد أن المجموعتين رغم أن لهما نفس المدى إلا أن الاختلاف بين قياسات كل مجموعة لا يساوى الاختلاف بين قياسات المجموعة الأخرى ، مما يؤكد أن المدى لا يكفى وحدة كمقياس لعملية الاختلاف أو التغير .

8-للحصول على مقياس أكبر دقة من المدى لقياس الاختلاف فإننا نقوم بتوسيع أو بتعميم فكرة الوسيط كما يلي :

نفرض أن لدينا مجموعة تحتوى على n من البيانات ، $y_1, y_2, y_3, \ldots\ldots, y_n$ منظمة أو مرتبة تبعا لقيمها . فإنه يمكن تعريف المئين P على إنه تلك القيمة y بحيث أن p% من القياسات أقل منها و % (100-p) أكبر منها (تقع فوقها) .

9-يعتبر المئين أكثر دقة وأكثر حساسية من المدى كمقياس للتغير أو الاختلاف ولكن أكثر عيوبه أننا نحتاج إلى العديد من المئينات لوصف مجموعة البيانات كافيا .

10-المنحنى الخطى ويسمى (المضلع التكراري التراكمي) أو (Ogive) ويمكن الحصول علية برسم التكرار التراكمي الأقل من الحدود العليا للفترات الجزئية ضد الحدود العليا للفترات الجزئية ، ثم توصيل النقاط المتتابعة بخط مستقيم .عند استخدام التكرارات النسبية بدلا من التكرارات العادية فإن المنحنى يسمى بالضلع التكراري التراكمي النسبي (Relative Frequency Ogive) أو المضلع التكراري التراكمي المئوي (percentage Ogive)

والذي يمتاز بسهولة وسرعة الحصول على المعلومة من خلاله .

11-الجدول التالي يوضح درجات عينة مكونة من 40 طالب :

2.2	4.1	3.5	4.5	3.2	3.7	3.0	2.6
3.4	1.6	3.1	3.3	3.8	3.1	4.7	3.7
2.5	4.3	3.4	3.6	2.9	3.3	3.9	3.1
3.3	3.1	3.7	4.4	3.2	4.1	1.9	3.4

4.7 3.8 3.2 2.6 3.9 3.0 4.2 3.5

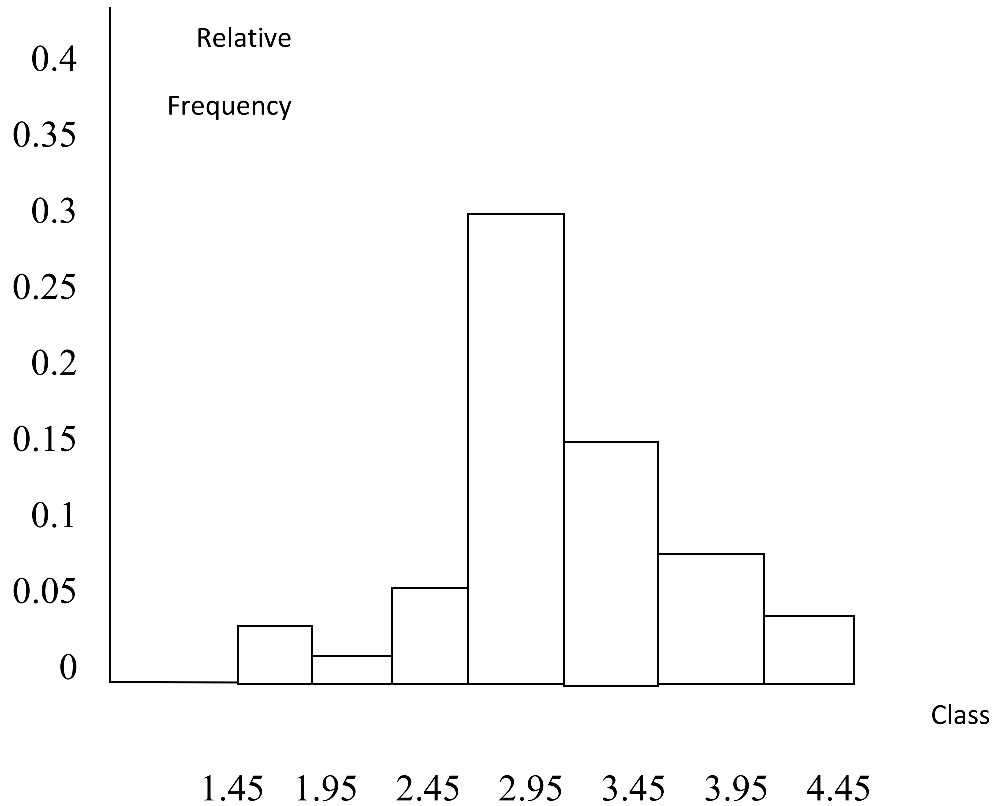
أ) ارسم المدرج التكراري النسبي لهذه البيانات .

ب) أوجد p_{80} (المئين 80) ، p_{85} (المئين 85) لتوزيع درجات الطلاب .

ج) ارسم المضلع التكراري التراكمي النسبي ثم اوجد باستخدامه كلا من المئين 25 والمئين 70

أ) جدولة البيانات :

الفترة الفصل (I)	حدود الفترة	التكرارات المطلقة f_i	التكرارات النسبية	التكرار التراكمي	التكرار التراكمي النسبي
1	1.45 – 1.95	2	2/40 = .05	2	2/40
2	1.95 – 2.45	1	1/4 = .30	3	3/40
3	2.45 – 2.95	4	4/40 = .1	7	7/40
4	2.95 – 3.45	15	15/40 = .38	22	22/40
5	3.45 – 3.95	10	10/40 = .25	32	32/40
6	3.95 – 4.45	5	5/40 = .13	37	37/40
7	4.45 – 4.95	3	3/40 = .08	40	40/40=1



ب)لإيجاد قيمة كل من p_{80} ، p_{85} نقوم بعمل الجدول التالي :

الفترات الجزئية	التكرار التراكمي	نسبة التكرار التراكمي
أقل من 1.45	0	(0/40) .100 = 0
اقل من 1.95	2	(2/40) .100 = 5.0
اقل من 2.45	3	(3/40) .100 = 7.5
اقل من 2.95	7	(7/40) .100 = 17.5
اقل من 3.45	22	(22/40) .100 = 55
اقل من 3.95	32	(32/40) .100 = 80
اقل من 4.45	37	(37/40) .100 = 92.5
اقل من 4.95	40	(40/40) .100 = 100

والآن لحساب قيمة p_{80} فإننا نبحث عن القيمة (الدرجة) y في العينة والتي تقع تحتها

$\frac{80}{100} x\ 40 = 32$ من القيم أو الدرجات في العينة . يوجد 32 قيمة تقع تحت الحد 3.95 وبالتالي فإن $80\% = (p_{80}) = 3.95$

ولإيجاد قيمة p_{85} فإننا نبحث عن القيمة (الدرجة) y في العينة والتي تقع تحتها $\frac{85}{100} x\ 40 = 34$ من القيم أو الدرجات في العينة . يوجد 32 قيمة تقع تحت الحد 3.95 . ولاستكمال هذا العدد من القيم فإننا نحتاج إلى 2 من 5 قيم موزعة بين الحدين 3.95 ، 4.45 وبالتالي نتحرك مسافة $\frac{2}{5} x\ 0.5 = 0.2$ بعد الحد 3.95 وعلية تكون قيمة

$$(85\%) = p_{85} = 3.95 + 0.2 = 4.15$$

ج) المضلع التكراري التراكمي النسبي :

12-أ) المثال السابق يوضح طريقة إيجاد المئين المناظر لقيمة معينة من مجموعة القيم .

ب)هناك طرق عديدة ومختلفة للحصول على العملية العكسية .

(إيجاد القيمة المناظرة لمئين معين) .

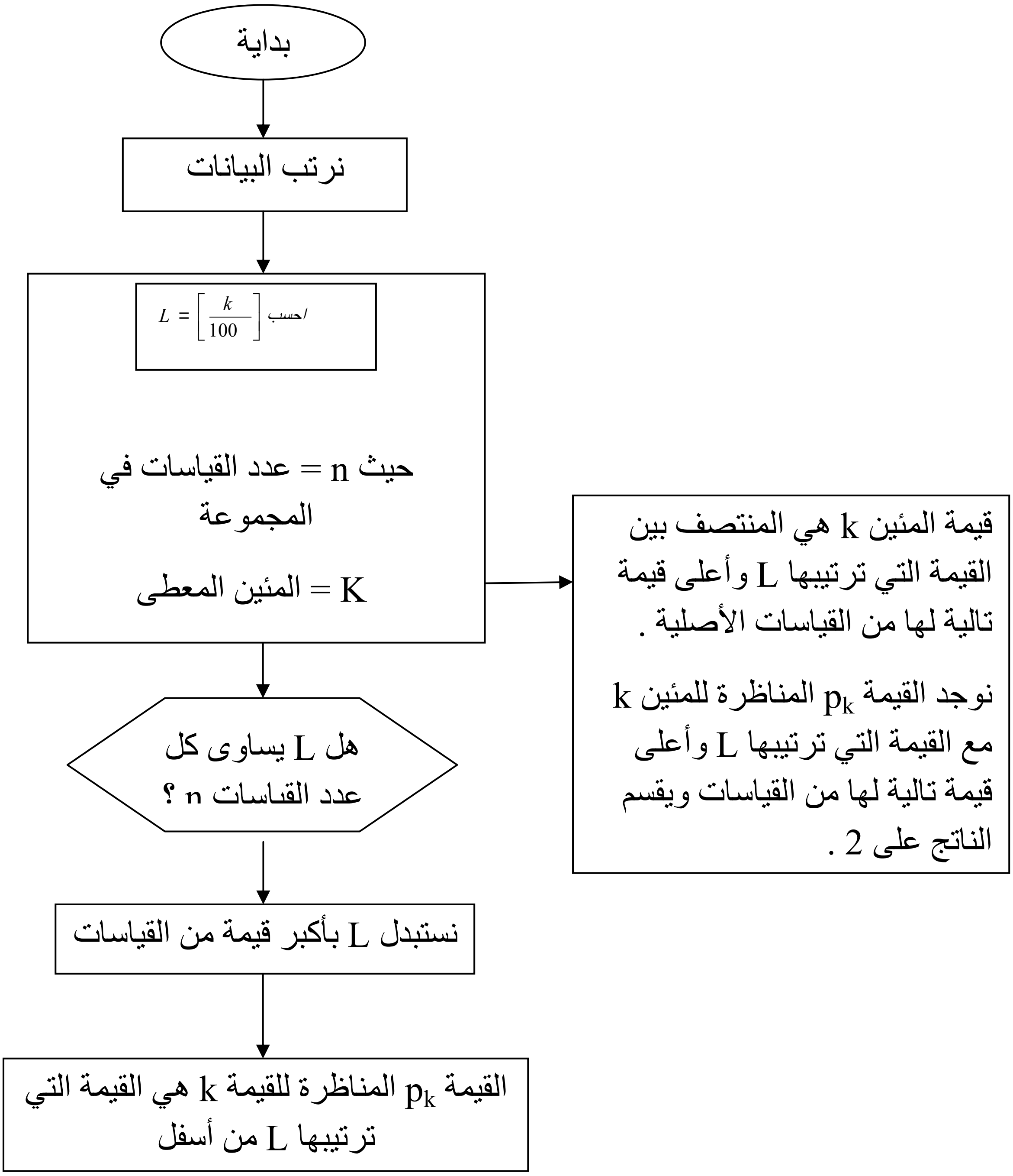
والطريقة التي تم اختيارها يمكن تلخيص خطواتها في خريطة التدفق شكل (1-2) والتي تستخدم فيها الرموز الآتية :

n عدد القياسات في المجموعة

k المئين المستخدم

L كاشف موضع القيم المراد إيجادها

P_k المئين k



إيجاد القيم المناظرة للمئين k

13- المئيان p_{25} ، p_{75} للعينة تسمى الربيع الأدنى (أو الأول) والربيع الأعلى (أو الثالث) للعينة ويرمز لهما بالرمز Q_3, Q_1 على التوالي . المئين p50 يسمى بالربيع الأوسط (أو الثاني) أو الوسط ويرمز له بالرمز Q_2 أي أن $p_{50} = Q_2$

14- أوجد قيم كل من Q_3, Q_1 للعينة الموجودة بمثال (11)

الحل :

لإيجاد Q_1 نبحث عن القيمة y في العينة والتي تقع تحتها

$\frac{25}{100} x\ 40 = 10$ من القيم .

نجد أنه يوجد 7 قيم تقع تحت الحد 2.95 ، ولاستكمال هذا العدد فإننا نحتاج إلى 3 من 15 قيمة موزعة بين الحدين 2.95 ، 3.45 .

وبالتالي فإننا نتحرك مسافة $\frac{3}{15} x\ 0.5 = 0.1$ بعد الحد 2.95 .

وعلية تكون Q1 = 2.95 + 0.1 = 3.05 ، والتي تتفق تماما مع النتيجة التي تم الحصول عليها من المدرج التكراري التراكمي .

وبالمثل نجد أن القيمة التي تقع تحتها $30 = \frac{75}{100} x\ 40$ من القيم . نجد أنه يوجد 22 قيمة تقع تحت الحد 3.45 ، ولاستكمال هذا العدد فإننا نحتاج إلى 8 من 10 قيمة موزعة بين الحدين 3.45 ، 3.95.

وبالتالي فإننا نتحرك مسافة $\frac{8}{10} x\ 0.5 = 0.4$ بعد الحد 3.45 وعلية تكون

. Q3 = 3.45 + 0.4 = 3.85

15- سنقوم الآن بعرض مقياس هام للتغير أو الاختلاف معتمدين على تشتت (Dispersion) أو انحراف (Deviation) القياسات عن الوسط الحسابي للعينة. بفرض أن $(y_i - \bar{y})$ تمثل الانحرافات رقم i عن الوسط الحسابي وبالطبع فإن الانحرافات الكبيرة تشير عادة إلى تغيرات واختلافات كبيرة .

16- يمكن تعيين هذه الانحرافات بأكثر من طريقة منها :

أ) مجموع n من الانحرافات عن الوسط الحسابي يساوى صفر . أي أن ،

$$\sum_{i=1}^{n} (y_i - \bar{y}) = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \bar{y}$$

$$= \sum_{i=1}^{n} y_i - n\,\bar{y}$$

$$= \sum_{i=1}^{n} y_i - n\,\frac{\sum_{i=1}^{n} y_i}{n} = 0$$

ب) لمنع حدوث المجموع الصفري ، يمكن استخدام متوسط القيم المطلقة للانحرافات والتي تسمى أحيانا بمتوسط الانحراف (Mean Deviation) وهذا المقياس يصعب حسابه ويصعب أيضا عند استخدامه إعطاء مقياس للجودة في عملية صنع الاستدلال أو الاستقراء .

مجموع مربعات الانحرافات $\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$ يعتبر مقياسا للانتشار Measure Of Spread.

كما يعرف (متوسط مربع الانحرافات في المجتمع بتباين المجتمع ويرمز له بالرمز σ^2 .

16- متوسط مربع الانحرافات لعينة حجمها n يستعمل كتقدير لتباين المجتمع σ^2 ، ويسمى S^2 بتباين العينة (Sample Variance) .

$$s^2 = \frac{1}{n-1}\sum_{i-1}^{n}\left(y_i - \bar{y}\right)^2$$

17- بفرض أن لدينا مجموعة تحتوى على n من القياسات $y_1, y_2, \ldots\ldots, y_n$ فإن تباين العينة Sample Variance ويعطى بالمعادلة

$$s^2 = \frac{1}{n-1}\sum_{i-1}^{n}\left(y_i - \bar{y}\right)^2$$

<u>18- مثال :</u>

احسب الوسط الحسابي والتباين للعينة 4, 2, 3, 5, 6

الحل : (أ) $\bar{y} = \frac{\sum_{i=1}^{5} y_i}{5} = \frac{20}{5} = 4$

(ب) لحساب تباين العينة توجد الجدول التالي :

y_i	$y_i - \bar{y}$	$\left(y_i - \bar{y}\right)^2$
4	(4 – 4)= 0	0
2	(2 – 4)= -2	4
3	(3 – 4)= -1	1
5	(5 – 4)= 1	1
6	(6 – 4)= 2	4
$\sum y_i = 20$	$\sum \left(y_i - \bar{y}\right) = 0$	$\sum \left(y_i - \bar{y}\right)^2 = 10$

$$s^2 = \frac{1}{n-1}\sum_{i-1}^{n}\left(y_i - \overline{y}\right)^2 = \frac{1}{5-1}\sum_{i-1}^{5}\left(y_i - \overline{y}\right)^2 = \frac{10}{4} = 2.5$$

14-2 طريقة مختصرة لحساب التباين:

A Short Method for Calculating the Variance

1-لإيجاد قيمة المقدار $s^2 = \frac{1}{n-1}\sum_{i-1}^{n}\left(y_i - \overline{y}\right)^2$ فإننا نحتاج إلى حساب المقدار $\sum_{i-1}^{n}\left(y_i - \overline{y}\right)^2$

ولجعل حساب المقدار الأخير أكثر سهولة نتبع الأتي :

$$\sum_{i-1}^{n}\left(y_i - \overline{y}\right)^2 = \sum_{i-1}^{n}\left(y_i^2 - 2y_i\overline{y} + \overline{y}^2\right)^2$$

$$= \sum_{i=1}^{n} y_i^2 - 2\overline{y}\sum_{i=1}^{n} y_i + n\overline{y}^2$$

$$= \sum_{i=1}^{n} y_i^2 - 2n\overline{y}^2 + n\overline{y}^2$$

$$= \sum_{i=1}^{n} y_i^2 - n\overline{y}^2$$

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

2-هذه العملية الحسابية تحتاج إلى :

(أ) المجموع الحسابي العادي للقياسات $\sum_{i=1}^{n} y_i$

(ب) مجموع مربعات القياسات $\sum_{i=1}^{n} y_i^2$

4- يراعى التفرقة بين كل من $\sum_{i=1}^{n}\mathbf{y}_i^2$ و $\left(\sum_{i=1}^{n}\mathbf{y}_i\right)^2$ المستخدمة في الفقرة (1).

(أ) لحساب $\sum_{i=1}^{n} y_i^2$ فإننا أولا نحسب مربعات القياسات ثم نوجد المجموع الجبري لهذه المربعات .

(ب) لحساب $\left(\sum_{i=1}^{n} y_i\right)^2$ فإننا أولا نوجد المجموع الجبري للقياسات ثم نوجد مربع هذا المجموع .

-5 مثال :

احسب S^2 للعينة في مثال 2 – 13 (18) .

الحل : (أ) نكون الجدول

yi	yi^2
4	16
2	4
3	9
5	25
6	36
$\sum y_i = 20$	$\sum y_i^2 = 90$

(ب)نحسب أن

$$\sum_{i-1}^{n} (y_i - \bar{y})^2 = \sum_{i-1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{5} y_i\right)}{5}$$

$$= 90 - \frac{(20)}{5} = 90 - 80 = 10$$

(ج) إذن

$$s^2 = \frac{1}{5-1} \sum_{i=1}^{5} (y_i - \bar{y})^2 = \frac{10}{4} 25$$

5-مثال :

احسب الوسط الحسابي والتباين للعينة 3 ,2 ,5 ,7 ,6 ,5

الحل:

(أ) نكون الجدول

y_i	y_i^2
5	25
6	36
7	49
5	25
2	4
3	9
$\sum y_i^2 = 28$	$\sum y_i^2 = 148$

إذن $\bar{y} = \frac{\sum_{i=1}^{6} y_i}{6} = \frac{28}{6} = 4.67$

(ب) لحساب التباين :

$$\sum_{i-1}^{6} (y_i - \bar{y})^2 = \sum_{i-1}^{6} y_i^2 - \frac{\left(\sum_{i=1}^{6} y_i\right)^2}{6}$$

نحسب أولا

$$=148 - \frac{(28)^2}{6} = 148 - 130.67 = 17.33$$

وعلية $\mathbf{S^2 = \frac{1}{6-1}\sum_{i=1}^{6}(y_i - \bar{y})^2 = \frac{17.33}{5} = 3.47}$

6- الانحراف المعياري للعينة Sample Standard Deviation التي حجمها n ، يرمز له بالرمز S ، يعرف على إنه الجذر التربيعى الموجب لتباين

العينة s^2 أي أن $s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}$

7- **<u>مثال :</u>**

الانحراف المعياري للعينة الموجودة في مثال (4) هو $s = \sqrt{s^2} = \sqrt{2.5} = 1.58$

8- إذا كانت هناك عينة حجمها n تحتوى على m من القيم العددية

المختلفة $(m \le n)$ $y_1 , y_2 , ..., y_m$

بتكرارات نسبية مناظرة $\tilde{f}_1 , \tilde{f}_2 , \ldots\ldots\ldots , \tilde{f}_m$

وبتكرارات مناظرة $(i = 1,2 \ldots\ldots, m)$ ، $f_i = n\,\tilde{f}_i$

فإن $s^2 = \frac{1}{n-1}\sum_{i=1}^{m}(y_i - \bar{y})^2 \cdot f_i$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{m} y_i^2 \cdot f_i - \frac{1}{n}\left(\sum_{i=1}^{m} y_i \cdot f_i\right)^2\right]$$

9- **مثال** : أوجد s^2, s للبيانات الممثلة بالجدول التكراري التالي :

مركز الفترة	التكرار
171 – 175	4
166 – 170	8
161 – 165	14
156 – 160	22
151 – 155	27
146 – 150	19
141 – 145	17
136 – 140	11
131 - 135	3

الحل : أولا نوجد الجدول الأتي :

الفترة i	حدود الفترة	مركز الفترة x_i	x_i^2	التكرار f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
1	171 – 175	173	29929	4	692	119716
2	166 – 170	168	28224	8	1344	225792
3	161 – 165	163	26569	14	2282	371966
4	156 – 160	158	24964	22	3476	549208
5	151 – 155	153	23409	27	4131	632043
6	146 – 150	148	21904	19	2812	416176
7	141 – 145	143	20449	17	2431	347633
8	136 – 140	138	19044	11	1518	209484
9	131 - 135	133	17689	3	399	53067
Total	----------	--------	--------	125	19085	2925085

ثانيا : $$s^2 = \frac{1}{125-1}\left[\sum_{i=1}^{9} x_i^2 . f_i - \frac{1}{125}\left(\sum_{i=1}^{9} x_i . f_i\right)^2\right]$$

$$= \frac{1}{124}\left[(2925085) - \frac{1}{125}(19085)^2\right] = 90.22$$

ثالثا : $$s = \sqrt{s^2} = \sqrt{90.22} = 9.50$$

15-2 أهمية الانحراف المعياري :

The Signinficance of the Standard Deviation

1-**نظرية تشيبى تشيف :** Tchebysheff's Theorem

بفرض أن مجموعة من القياسات حجمها n ، $y_1 , y_2 , ..., y_n$. لأي عدد $k \geq 1$ ، فأنة يوجد على الأقل $\left(1-\frac{1}{k^2}\right)$ من هذه القياسات تقع داخل kانحراف معياري عن الوسط الحسابي $\left(\bar{y} \pm KS\right)$للمجموعة .

تظهر أهمية هذه النظرية لإمكانية استخدامها لأي جزء من (أو كل) القياسات.

سنرمز للمجتمع باستخدام الوسط الحسابي للمجتمع μ . والانحراف المعياري للمجتمع σ.
أو سنرمز للعينة المأخوذة من هذا المجتمع باستخدام $\bar{y}$ و S ، الوسط الحسابي والانحراف المعياري للعينة على الترتيب .

2-الجدول التالي يوضح العلاقة بين قيم k المختلفة والفترات $\bar{y} \pm KS$:

k	الفترة $\bar{y} \pm KS$	الفترة تحتوى على الأقل $\left(1\frac{1}{k^2}\right)$
1	$\bar{y} \pm s$	$\left(1-\frac{1}{1}\right) = 0$
2	$\bar{y} \pm 25$	$\left(1-\frac{1}{2^2}\right) = \frac{3}{4}$
3	$\bar{y} \pm 3s$	$\left(1-\frac{1}{3^2}\right) = \frac{8}{9}$
10	$\bar{y} \pm 10s$	$\left(1-\frac{1}{10^2}\right) = \frac{99}{100}$

3-**مثال** :

الوسط الحسابي والتباين لمجموعة من القياسات عددها 20 هما25،35 على التوالي . استخدام نظرية تشيبي تشيف لوصف توزيع هذه المجموعة من القياسات .

الحل: بما أن $\bar{y} = 35$ ، $s^2 = 25$ فان $s = 5$

إذن :

(أ) على الأقل $\frac{3}{4}$ من القياسات تقع داخل الفترة $35 \pm 2\ (5)$ أو 25 إلى 45

(ب) على الأقل $\frac{8}{9}$ من القياسات تقع داخل الفترة $35 \pm 3\ (5)$ أو 20 إلى 50

(ج) على الأقل $\frac{15}{16}$ من القياسات تقع داخل الفترة $35 \pm 4\ (5)$ أو 15 إلى 55

4- مثال :

الوسط الحسابي والتباين لعينة من القياسات عددها 50 هما 42 ، 36 على التوالي . استخدام نظرية تشيبي تشيف لوصف توزيع هذه العينة من القياسات .

الحل: بما أن $\bar{y} = 42$ ، $s^2 = 36$ فان $s = 6$

إذن :

(أ)على الأقل $\frac{3}{4}$ من القياسات تقع داخل الفترة $42 \pm 2\ (6)$ أو 30 إلى 54 .

(ب) على الأقل $\frac{8}{9}$ من القياسات تقع داخل الفترة $42 \pm 3\ (6)$ أو 24 إلى 60 .

(ج) على الأقل $\frac{15}{16}$ من القياسات تقع داخل الفترة $42 \pm 4\ (6)$ أو 18 إلى 66 .

5- القاعدة التجريبية : Empirical Rule

إذا كان لدينا مجموعة من القياسات تأخذ تقريبا شكل التوزيع الطبيعي (شكل الناقوس – Approximately Bell – Shaped) أو شكل الكومة – Mound – Shaped) الموضح بالشكل التالي فان :

أ) الفترة $\mu \pm \sigma$ تحتوى تقريبا على 68% من القياسات .

ب)الفترة $\mu \pm 2\sigma$ تحتوى تقريبا على 95% من القياسات .

ج) الفترة $\mu \pm 3\sigma$ تحتوى تقريبا على 99.7% من القياسات .

وتسمى القاعدة التجريبية في بعض الأحيان بالقاعدة 68-95-99 .

6- مثال: إذا كان الوسط الحسابي والتباين لعينة حجمها 100 من القياسات

8.7 ، 0.36 على التوالي . بفرض أن توزيع هذه القياسات يأخذ شكل الكومة (Mound – Shaped) . صف هذه العينة مستخدما القاعدة التجريبية .

<u>الحل :</u>

نوجد جدول الفترات التالي :

k	$\bar{y} \pm ks$	$\bar{y}-ks$ to $\bar{y}+ks$		
1	$\bar{y} \pm s$	7.2	to	8.8
2	$\bar{y} \pm 2s$	6.6	to	8.8
3	$\bar{y} \pm 3s$	6.0	to	8.8

ومن الجدول يتضح :

أ) % 68 من القياسات تقع داخل الفترة من 7.2 إلى 8.4 .
ب)% 95 من القياسات تقع داخل الفترة من 6.6 إلى 9.0 .

ج)% 68 من القياسات تقع داخل الفترة من 6.0 إلى 9.6 .

7- عندما يكون حجم العينة n صغيرا ، فان توزيع هذه العينة لا يأخذ شكل الكومة (Mound – Shaped) وعلية فان القاعدة التجريبية (Empirical Rule) لا يمكن استخدامها لوصف هذا النوع من العينات .

وحيث أن نظرية تشيبي تشيف يمكن تطبيقها لأي جزء من (أو كل) العينة ، فإننا

نستخدمها دون النظر إلى حجم العينة .

<u>16-2 التأكد من حساب s :</u> A Check on the Calculation of S

1- عينة القياسات التي تأخذ شكل الكومة (Mound – Shaped) أو تقريبا شكل التوزيع الطبيعي (Normal) ، يمكن استخدام مدى العينة (Sample Range) للتأكد من صحة قيمة s (الانحراف المعياري) للعينة المحسوبة.

2- تبعا لنظرية تشيبى والقاعدة التجريبية فأنة على الأقل $\frac{3}{4}$ أو 95% من القياسات تقع

داخل الفترة $\bar{y} \pm 2S$. وعلية فان مدى العينة ، R ، يساوى تقريبا4S ، أي أنة $s \cong \frac{R}{4}$.

3- ويمكن إيضاح ذلك كما يلي :

المجتمعات	العينات
$range \approx 4\sigma$ or $\sigma \approx \frac{range}{4}$	$range \approx 4S$ or $S \approx \frac{range}{4}$

5-**مثال :** تأكد من صحة قيمة S المحسوبة لمجموعة البيانات (أ) وعددها 20 في البند 2-13 (6) .

الحل: من مجموعة البيانات المعطاة نجد أن المدى 88 =R=96-8 وبالتالي فإن $S \cong \frac{R}{4} = \frac{88}{4} = 22$.وبمقارنة الناتج 22 بقيمة S المحسوبة 25.46نلاحظ الفرق بينهما .

6-مثال : بالرجوع إلى المجموعة (ب) من البيانات في البند 2-13(6) والتي تحتوى على 12 قياسا نجد أن المدى R=9-0.2=8.8 وبالتالي فإن $S \cong \frac{R}{4} = \frac{8.8}{4} = 2.2$ نلاحظ أن الفرق هنا بين القيمة المحسوبة لـS وهى 3.21 والقيمة التقريبية لها هي 2.2 أكبر من مثيلة في المثال السابق .

7-من المثالين السابقين يمكن استنتاج أن عملية التقريب تعتمد على حجم العينة n والجدول التالي يوضح العلاقة بين n والقيمة التي يتم تقسيم المدى عليها .

n	اقسم R على
5-10	2.5
10-25	3
25-100	4
100-	5

8-**مثال :** استخدم المدى في إيجاد القيمة التقريبية للتأكد من القيمة 3.21 المحسوبة للانحراف المعياري للعينة الممثلة بالمجموعة الثانية في مثال 2-13 (6)

الحل : بما أن R=8.8 وحجم العينة n=12 فإن

$$S \cong \frac{R}{3} = \frac{8.8}{3} = 2.93.$$

والتي تبدو أكثر قربا وأكثر دقة بالمقارنة بالقيمة المحسوبة 3.21 .

9-**مثال** : استخدام المدى في إيجاد القيمة التقريبية للتأكد من القيمة المحسوبة للانحراف المعياري للعينة الممثلة بالمجموعة الثانية في مثال 2-13(18).

الحل : العينة تحتوى على 5 قياسات هي 4,2,3,5,6 ومدى هذه المجموعة هو R=6-2=4 وبالتالي فإن

$$S \cong \frac{R}{2.5} = \frac{4}{2.5} = 1.6$$

وهى قريبة جدا من القيمة المحسوبة للانحراف المعياري S=1.58

17-2 تأثير عملية التكويد فى حساب الوسط $\bar{y}$ والتباين S^2:

The Effect of Coding on $\bar{y}$ and S^2

1- لتبسيط العمليات الحسابية عند حساب قيمة كل من الوسط الحسابي والتباين لعينة من القياسات يمكن استخدام عملية تكويد القياسات والتي تتمثل في طرح (أو جمع) مقدار ثابت من (أو على) كل قياس وكذلك ضرب (أو قسمة) كل قياس في (أو على) مقدار ثابت أو الاثنين معا .

2- النظريات الآتية توضح علاقة كل من الوسط الحسابي $\bar{y}$ والتباين S^2 بعد إجراء عملية التكويد بقيمها الأصلية .

3- نظرية (1) : بفرض مجموعة تحتوى على n من القياسات $y_1, y_2, \ldots\ldots\ldots\ldots, y_n$ ، بفرض أن $x_i=y_i-c$, ، $i=1,\ldots,n$ فإنه:

أ) $\bar{x} = \bar{y} - c$ أو $\bar{y} = \bar{x} + c$

ب) $S_x^2 = S^2$.

البرهان : أ)

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

$$= \frac{\sum_{i=1}^{n} y_i}{n} - \frac{\sum_{i=1}^{n} c}{n} = \bar{y} - c$$

$$\bar{y} = \bar{x} + c .$$

ب)بالمثل

$$(n-1)S_x^2 = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \sum_{i=1}^{n}\left(y_i - c - [\bar{y} - c]\right)^2$$

$$= \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = (n-1)S^2$$

$$\therefore S_x^2 = S^2$$

4- **مثال** : أوجد الوسط الحسابي والتباين لعينة تحتوى على 5 قياسات هي .13,12,14.12.10

الحل:

بفرض أن $x_i=y_i-10$. البيانات المكودة هي 3,2,4,2,0 وعليه

$$\sum_{i=1}^{5} x_i = 11 \;,\; \sum_{i=1}^{5} x_1^2 = 33 \;.$$

ونجد أن
$$\bar{x} = \frac{\sum_{i=1}^{5} x_i}{n} = \frac{11}{5} = 2.2$$

و
$$S_x^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right] = \frac{1}{4}\left[33 - \frac{1}{5}(121)\right] = 2.2$$

وباستخدام نظرية (1) نجد أن ،

$$\bar{y} = \bar{x} + 10 = 2.2 + 10 = 12.2$$
$$S^2 = S_x^2 = 2.2$$

5- نظرية (2) : بفرض مجموعة تحتوى على n من القياسات $y_1, y_2, \ldots, y_n$ ، بفرض أن $x_i=ky_i$ ، $i=1,\ldots,n$ فإنه :

(أ) $$\bar{y} = \frac{1}{k}\bar{x} \quad or \quad \bar{x} = k\bar{y}$$

(ب) $$S_x^2 = k^2 S^2 \quad or \quad S^2 = \frac{1}{k^2} S_x^2 .$$

البرهان : (أ)

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{n} k\, y_i}{n}$$

$$= k \frac{\sum_{i=1}^{n} y_i}{n} = k\,\bar{y}$$

$$\therefore\ \bar{y} = \frac{1}{k}\bar{x}$$

ب-بفرض أن $(n-1)S_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$

$$\sum_{i=1}^{n}(ky_i - k\bar{y})^2$$

$$= k^2\sum_{i=1}^{n}(y_i - \bar{y})^2 = (n-1)k^2S^2$$

$$(n-1)S_x^2 = (n-1)k^2S^2$$

أو $S_x^2 = k^2S^2$ إذن $S^2 = \frac{1}{k^2}S_x^2$.

6- **<u>مثال</u>**: أوجد الوسط الحسابي $\bar{y}$ والتباين S^2 للعينة 0.03,0.04,0.01,0.05, 0.01, 0.07, باستخدام التكويدة $x_i=100\,y_i,\ i=1,\ldots,6$.

<u>الحل</u>: القياسات المكودة هي 5,1,7,3,4,1

وعلية فإن $\sum_{i=1}^{6} x_i = 21 \ and\ \sum_{i=1}^{6} x_i^2 = 101$.

وبالتالي $\bar{x} = \frac{21}{6} 3.5\ and\ S_x^2 = \frac{1}{5}\left[101 - \frac{(21)^2}{6}\right] = 5.5.$

وباستخدام نظرية (2) نجد أن ،

$$\bar{y} = \frac{\bar{x}}{100} = \frac{3.5}{100} = .035\ and\ S^2 = \frac{S_x^2}{(100)^2} = \frac{5.5}{10000} = .00055\ .$$

7-**<u>نظرية (3) :</u>** بفرض مجموعة تحتوى على n من القياسات $y_1, y_2, \ldots\ldots, y_n$ ، بفرض أن $x_i=k(y_i-c)$ ، $i=1, \ldots, n$ فإنه :

(أ) $\bar{y} = \frac{1}{k}\bar{x} + c\ , or\ , \quad \bar{x} = k\left(\bar{y} - c\right)$

(ب) $S_x^2 = k^2 S^2 \ or \ S^2 = \frac{1}{k^2} S_x^2 .$

البرهان: "تمرين"

8-مثال : أوجد الوسط الحسابي $\bar{y}$ والتباين S^2 لمجموعة القياسات 100.0, 100.3, 100.8, 100.4, 100.6, 100.3 باستخدام التكويدة

$.x_i = 10(yi-100),\ i=1,\ldots,6$

الحل : القياسات المكودة هي 0,3,8,4,6,3 وعلية فإن

$$\sum_{i=1}^{6} x_i = 24 \quad and \quad \sum_{i=1}^{6} x_i^2 = 134 \quad .$$

وبالتالي $\bar{x} = \frac{24}{6} 4 \ and \quad s_x^2 = \frac{1}{5}\left[134 - \frac{\left(24\right)^2}{6}\right] = 7.6 .$

وباستخدام نظرية (3) نجد أن ،

$$y = \frac{1}{10}\left(4\right) + 100 = 100.4 \ and \quad S^2 = \frac{1}{\left(10\right)^2} S_x^2 = \frac{7.6}{100} = .076 .$$

9-يجب مراعاة عدم استخدام عملية التكويد إلا عندما تكون أكثر سهولة وأكثر سرعة ، كحالة عدم وجود آلات حاسبة أو حاسبات آلية ، ويعتبر استخدامها عندئذ مضيعة للوقت .

10-تعتبر عملية التكويد أكثر تأثيرا عندما تحتوى القياسات على أعداد كسرية عشرية ، وكذلك الأعداد الكبيرة التي تنتهي بالأصفار ، مثال ذلك ، 500 ,200 ,17 ,0.0035 ,0.0021.

تمارين

في التمارين من 1 إلى 4 عين كلا من : طول الفترة – مركز الفترة – حدود الفترة للجداول التكرارية المعطاة .

-1

الدرجة	التكرار
87-80	16
95-88	37
103-96	50
111-104	29
119-112	14

-2

الزمن (دقيقة)	التكرار
7.5-0.0	16
15.1-7.6	18
22.7-15.2	17
30.3-22.8	15
37.9-30.4	19

-3

الوزن (كجم)	التكرار
12.1-16.2	16
26.1-12.2	15
31.1-26.2	12
36.1-31.2	8
41.1-36.2	3

-4

المبيعات (جنية)	التكرار
21-0	2
43-22	5
65-44	8
87-66	12
109-88	14
131-110	20

- في التمارين من 5 إلى 8 كون الجدول التكراري للجداول التكرارية المعطاة أمام كل تمرين :

5-الجدول التكراري تمرين (1) .

6-الجدول التكراري تمرين (2) .

7- الجدول التكراري تمرين (3) .

8- الجدول التكراري تمرين (4) .

- في التمارين من 9 إلى 12 كون الجدول التكراري للجداول التكرارية المعطاة أمام كل تمرين :

9-الجدول التكراري تمرين(1)   10-الجدول التكراري تمرين(2)

11-الجدول التكراري تمرين(3)   12-الجدول التكراري تمرين(4)

• في التمارين من 13 إلى 16 ارسم المدرج التكراري للجداول التكرارية الآتية :

-13

الدرجة	التكرار
43-42	14
45-44	11
47-46	8
49-48	6
51-50	4
53-52	3
55-54	1
57-56	2
59-58	0
61-60	1

-14

العمر	التكرار
9-0	37
19-10	35
29-20	39
39-30	46
49-40	33
59-50	22
69-60	20
79-70	17
89-80	9
99-90	3

-15

الزمن (دقيقة)	التكرار
5.9-0	10
11.9-6	8
17.9-12	9
23.9-18	9
29.9-24	4

-16

الأسئلة	التكرار
25-16	22
35-26	10
45-36	6
55-46	2
65-56	4
75-66	5
85-76	1

17- الجداول التكرارية الآتية توضح تكرار الأرقام العشرية لمائة رقم لكل من π و $\frac{22}{7}$.

أ) ارسم المدرج التكراري للجدولين مع ملاحظة الفرق

ب) العددان π و $\frac{22}{7}$. كلاهما حقيقيان . وضح الفرق بينهما من خلال الرسم .

• في التمارين من 18 إلى 23 أوجد كلا من :

أ) الوسط الحسابي ب) الوسيط ج) المنوال د) منتصف المدى .

(18) –

66 67 77 65 68 71 77 77 73 74

(19) –

42 100 77 54 93 85 67 77 62 58

(20) –

0.27 0.17 0.17 0.16 0.13 0.24 0.240.24

0.18 0.17 0.21 0.16 0.12 0.160.14

(21) –

3.7 2.9 3.4 0 1.5 1.8 2.3 2.4 1 1 2

4.4 2 4.5 0 1.7 4.4 3.3 2.4 2.1 2.1

(22) –

38 27 14 18 34 16 42 28 24 40 20 23 31

21 30 25 17 28 33 25 23 19 51 18 29 17

37

(23) –

61 57 57 58 57 61 54 68 51 49 64 50 48

56 46 54 49 51 47 55 55 54 42 51 56 57

51 60 62 43 55 56 61 52 69 64 46 65 52

55 51 54

- في التمارين من 24 إلى 27 أوجد الوسط الحسابي لكل من الجداول التكرارية الآتية .

24- الجدول التكراري تمرين (13).

25- الجدول التكراري تمرين (14).

26- الجدول التكراري تمرين (15).

27- الجدول التكراري تمرين (16).

28- (1) أوجد :

أ) الوسط الحسابي ب) الوسيط ج) المنوال د) منتصف المدى .

للقياسات التالية : 236.000 – 206.000 – 179.000 – 236.000 – 108.000

(2) أعد حساب المطلوب في الفقرة (1) بعد إضافة 20.000 لكل قياس .

(3) عموما ، ما هو تأثير عملية جميع مقدار ثابت k على كل قياس ؟ وما هو تأثير عملية طرح مقدار ثابت k من كل قياس ؟

(4) أعد حساب المطلوب في الفقرة (1) بعد قسمة كل قياس على 1000.

(5) عموما ، ما هو تأثير عملية قسمة (أو ضرب) مقدار ثابت k على (أو في) كل قياس ؟

- في التمارين من 24 إلى 34 أوجد المدى – التباين – الانحراف المعياري لكل من الجداول التكرارية الآتية .

29- الجدول التكرارية تمرين (18).

30- الجدول التكراري تمرين (19).

31- الجدول التكراري تمرين (20).

32- الجداول التكراري تمرين (21).

33- الجدول التكراري تمرين (22).

34- الجداول التكراري تمرين (23).

- في التمارين من 35 إلى 38 أوجد :

(أ) المدى (ب) التباين (ج) الانحراف المعياري

لكل من الجداول التكرارية الآتية .

35- الجداول التكراري تمرين (13).

36- الجداول التكراري تمرين (14).

37- الجداول التكراري تمرين (15).

38- الجداول التكراري تمرين (16).

39- (أ) أوجد :

(أ) المدى (ب) التباين (ج) الانحراف المعياري

للقياسات التالية :

66	67	77	65	68	71	77	77	73	74

(2) أعد حساب المطلوب في الفقرة (1) بعد إضافة 15 لكل قياس . ثم قارن النتائج بالفقرة (1) موضحا تأثير عملية جمع مقدار ثابت k على كل قياس ؟

(3) أعد حساب المطلوب في الفقرة (1) بعد طرح 5 من كل قياس . ثم قارن النتائج بالفقرة (1) موضحا تأثير عملية طرح مقدار ثابت k على كل قياس ؟

(4) أعد حساب المطلوب في الفقرة (1) بعد ضرب 10 من كل قياس . ثم قارن النتائج بالفقرة (1) موضحا تأثير عملية ضرب مقدار ثابت k في كل قياس ؟

(5) أعد حساب المطلوب في الفقرة (1) بعد أخذ نصف قيمة كل قياس . ثم قارن النتائج بالفقرة (1) موضحا تأثير عملية قسمة كل مقدار ثابت k ؟

40- الوسط الحسابي والانحراف المعياري لعينة من القياسات هما 56 ، 4 على التوالي . فإذا كانت هذه البيانات تأخذ شكل الجرس (شكل الكومة) . استخدام القاعدة التجريبية للإجابة على ما يلي :

(أ) ما هي نسبة القياسات الواقعة بين القيمة 52 ، 60 ؟

(ب) ما هي نسبة القياسات التي تقع على بعد 8 من الوسط ؟

(ج) ما هما القيمتان اللتان يحصران بينهما حوالي 99.7 % من القياسات ؟

41- القياسات الآتية توضح درجات 20 من الطلاب تم اختيارهم عشوائيا من مجتمع الدراسة لتجربة ما :

18.7	18.6	16.3	18.1
22	20.3	19.9	15.2
18.2	21.2	17.7	19.7
20.2	19.4	19.7	20.9
19.6	17.9	17.2	19.8

(أ) ما هو مدى هذه القياسات ؟

(ب) ارسم المدرج التكراري النسبي للقياسات المعطاة مستخدما فترات جزئية بطول يساوى 1. (يمكن أن تبدأ بالقيمة 15.15)

(ج) بدراسة المدرج التكراري في الفقرة (ب) :

(1) ما هو احتمال وقوع قياس تم اختياره عشوائيا من هذه القياسات داخل الفترة الجزئية 17.15 – 21.15 ؟

(2) ما هو احتمال أن القياس الذي تم اختياره من العينة يكون أكبر من 19.15 ؟

(د) أعد ترتيب القياسات تبعا لقيمها بداية من 15.2 .

(هـ) ما هو الوسيط لمجموعة القياسات ؟

(و) المئين .... هو أي عدد يقع بين 16.3 ، 17.2 ؟

(ز) المئين .... هو أي عدد يقع بين 19.19 ، 20.2 ؟

(ح) احسب كلا من $\bar{y}$ و S^2 و S لمجموعة القياسات .

(ط) هل مجموعة القياسات المعطاة تتفق مع نظرية تشيبى تشيف ؟ أكد إجابتك بحساب نسبة القياسات الواقعة في الفترات $\bar{y} \pm ks$ حيث k=1,2,3 .

(ى) هل القاعدة التجريبية تصف بدقة هذه القياسات ؟

42- الوسط الحسابي والتباين لدرجات عينة من الطلاب حجمها 100 هما 76 و 65.71 على التوالي .

(أ) استخدام نظرية تشيبى تشيف لوصف هذه العينة .

(ب) بفرض أن هذه الدرجات تتبع التوزيع الطبيعي تقريبا ، استخدم القاعدة التجريبية لوصف هذه الدرجات .

43- استخدم المدى لحساب القيمة التقريبية للانحراف المعياري S للعينة الموجودة في تمرين (41) وللتأكد من القيمة المحسوبة لتلك العينة .

44- أوجد القيمة التقريبية للانحراف المعياري S باستخدام المدى ثم احسب قيمة كل من $\bar{y}$ و S و S لمجموعة البيانات :

(أ) 5 ، 5 ، 6 ، 4 ، 5 .

(ب) 1 ، 2 ، 1 ، 2 ، 3 ، 2 ، 0 ، 4 ، 1 ، 3 ، 0.

قارن القيم التقريبية للانحراف المعياري S بالقيم المحسوبة .

45- **<u>أكمل :</u>**

1- الهدف من علم الإحصاء هو .... حول المجتمع باستخدام البيانات الموجودة

في ....

2-مجموعات الأعداد يمكن أن توصف .... أو باستخدام القياسات الوصفية ....

3-الطريقة البيانية توضح .... أو التوزيع التكراري المدى يصف كيفية توزيع القياسات على محور القياس .

4-صناعة الاستدلال هي بروتوكول ذو .... وهما .... و ....

5-من أهم القياسات الوصفية العددية .... و ....

6-الوسط الحسابي يحدد .... توزيع القياسات .

7-الانحراف المعياري يوضح .... لمجموعة القياسات .