Stat101

Chapter 3

Statistics for Describing, Exploring, and Comparing Data

إحصاءات عن وصفه، استكشاف، ومقارنة البيانات

Review and Preview مراجعة ومعاينة

Chapter 1

Distinguish between population and sample, parameter and statistic Good sampling methods: *simple random sample*, collect in appropriate ways

التمييز بين السكان وعينة، المعلمة والإحصائية

طرق أخذ العينات جيدة: عينة عشوائية بسيطة، وجمع بالطرق المناسبة

Chapter 2

Frequency distribution: summarizing data

Graphs designed to help understand data Center, variation, distribution, outliers, changing characteristics over time

توزيع تردد: تلخيص البيانات

الرسوم البيانية المصممة للمساعدة في فهم مركز البيانات، والتباين، التوزيع، القيم المتطرفة، وتغيير الخصائص مع مرور الوقت

Important Statistics الاحصائيات الهامة

Mean, median, standard deviation,

variance

متوسط، متوسط، الانحراف المعياري،
التباين

Understanding and Interpreting فهم وتفسير

these important statistics

هذه الإحصاءات الهامة

Preview معاينة

- **Descriptive Statistics الإحصاء الوصفي**

In this chapter we'll learn to summarize or describe the important characteristics

By Hoor

of a known set of data

في هذا الفصل سوف نتعلم لتلخيص أو وصف الخصائص المهمة للمجموعة معروفة من البيانات

❖ **Inferential Statistics** **الاحصائيات استنتاجي**

In later chapters we'll learn to use sample data to make inferences or generalizations about a population

في فصول لاحقة سوف تعلم كيفية استخدام البيانات النموذجية الوصول لاستنتاجات أو تعميمات حول السكان

Measures of Center تدابير من مركز

Key Concept المفهوم الرئيسي

Characteristics of center. Measures of center, including mean and median, as tools for analyzing data. Not only determine the value of each measure of center, but also interpret those values.

خصائص المركز. تدابير المركز، بما في ذلك المتوسط والوسيط، كأدوات لتحليل البيانات. لا فقط تحديد قيمة كل تدبير من المركز، ولكن أيضا تفسير هذه القيم.

Part 1 جزء 1
Basics Concepts of Measures of Center أساسيات مفاهيم تدابير من مركز

Measure of Center **مقياس لمركز**

❖**Measure of Center** **مقياس لمركز**

the value at the center or middle of a data set **قيمة في مركز أو وسط مجموعة من البيانات**

Arithmetic Mean **المتوسط الحسابي**

Arithmetic Mean (Mean) **الحسابي (متوسط)**

the measure of center obtained by adding the values and dividing the total by the number of values

مقياس مركز الحصول بإضافة القيم وقسمة المجموع على عدد من القيم

What most people call an *average*. **ما أكثر الناس دعوة في المتوسط**

Notation الرموز

denotes the sum of a set of values. **يدل على مبلغ من مجموعة من القيم .**

***x* is the variable usually used to represent**

the individual data values. x **هو متغير عادة ما تستخدم لتمثيل قيم البيانات الفردية.**

n **represents the number of data values in a**

sample. n تمثل عدد قيم البيانات في عينة.

N **represents the number of data values in a**

population. يمثل N عدد من قيم البيانات في عدد السكان.

x **is pronounced 'x-bar' and denotes the mean of a set**

of sample values ***x*** = بار "ويدل على متوسط مجموعة من عينة قيم س-'X س هو واضح =

μ **is pronounced 'mu' and denotes the mean of all values in a population** μ غير

$$\bar{x} = \frac{\Sigma x}{n}$$

واضحة "مو" ويدل على متوسط كافة القيم في عدد السكان

$$\mu = \frac{\Sigma x}{N}$$

Mean

Advantages مزايا

Is relatively reliable, means of samples drawn from the same population don't vary as much as other measures of center Takes every data value into account

يمكن الاعتماد عليها نسبيا، وسائل العينات المسحوبة من نفس السكان لا تختلف كثيرا مثل غيرها من التدابير من مركز تحيط كل قيمة البيانات في الاعتبار

Disadvantage مساوئ

Is sensitive to every data value, one extreme value can affect it dramatically;

is not a *resistant* measure of center

غير حساسة للكل قيمة البيانات، يمكن أن القيمة القصوى واحدة تؤثر عليه بشكل كبير.

ليس مقياسا مقاومة للمركز

Median الوسيط

Median

the middle value when the original data values are arranged in order of increasing (or decreasing) magnitude

القيمة المتوسطة عندما يتم ترتيب قيم البيانات الأصلية من أجل زيادة (أو النقصان) حجم

~ ❖often denoted by *x* (pronounced 'x-tilde')

('تيلدا-X' غالبا ما يشار إليه ب خ (وضوحا

is not affected by an extreme value - is a resistant measure of the center

لا تتأثر قيمة المتطرفة - هو مقياس مقاومة للمركز

Finding the Median العثور على الوسيط

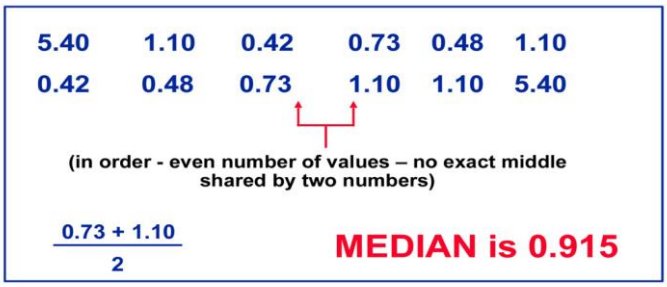
First *sort* the values (arrange them in order), the follow one of these

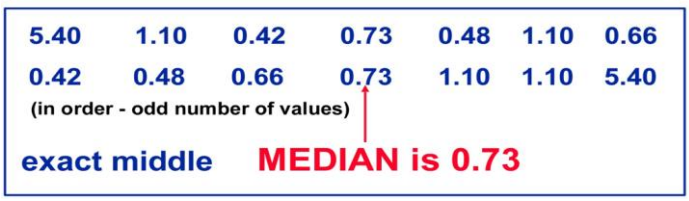
النوع الأول القيم (ترتيبها في الترتيب)، واتباع واحدة من هذه

1. **If the number of data values is odd, the median is the number located in the exact middle of the list.**
   إذا كان عدد قيم البيانات هو الغريب، والوسيط هو العدد الموجود في منتصف الدقيق للقائمة.

2. If the number of data values is even, the median is found by computing the mean of the two middle numbers.

وإذا كان عدد قيم البيانات حتى، وجدت على المتوسط عن طريق حساب متوسط اثنين من الأرقام المتوسطة





Mode الوضع

❖ **Mode the value that occurs with the greatest frequency**

وضع القيمة التي تحدث مع أكبر تردد

❖ **Data set can have one, more than one, or no mode**

مجموعة البيانات يمكن أن يكون واحد، أكثر من واحد، أو أي وضع

Bimodal two data values occur with the same greatest frequency

تحدث قيم البيانات اثنين ذات النسقين مع نفس أعظم تواتر

Multimodal more than two data values occur with the same greatest

تحدث المتعدد الوسائط أكثر من قيمتين البيانات مع نفسه أعظم

frequency No Mode no data value is repeated

تردد لا يتكرر الوضع لا قيمة البيانات

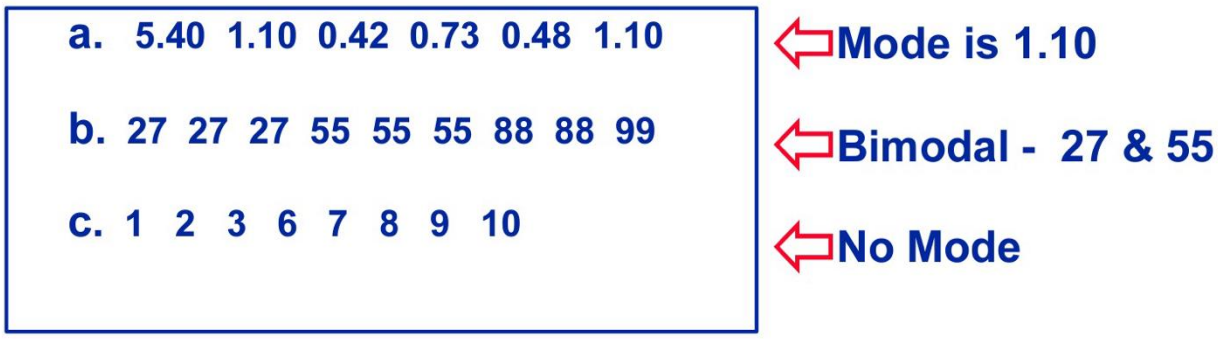
Mode is the only measure of central

الوضع هو المقياس الوحيد للوسط

tendency that can be used with nominal data

الميل التي يمكن استخدامها مع البيانات الاسمية

Mode - Examples وضع – أمثل



Definition فريف

Midrange المدى المتوسط

the value midway between the maximum and minimum values in the original data set

في منتصف الطريق بين قيمة القيم القصوى والدنيا في مجموعة البيانات الأصلية

Midrange = $\frac{maximum\ value+minimum\ value}{2}$

Midrange المدى المتوسط

Sensitive to extremes حساسة للتطرف

because it uses only the maximum and minimum values, so rarely used

لأنه يستخدم فقط القيم القصوى والدنيا، لذلك نادرا ما تستخدم

❖ **Redeeming Features** الميزات التعويض

(1) very easy to compute

(2) reinforces that there are several ways to define the center

(3) Avoids confusion with median

(1) من السهل جدا لحساب

(2) يعزز أن هناك عدة طرق لتحديد مركز

(3) يتجنب الخلط مع متوسط

Round-off Rule for Measures of Center جولة الاعادة القاعدة لتدابير من مركز

Carry one more decimal place than is present in the original set of values.

حمل واحد أكثر مكان عشري مما هو موجود في المجموعة الأصلية من القيم

Critical Thinking التفكير النقدي

Think about whether the results are reasonable.

التفكير في ما إذا كانت نتائج معقولة

Think about the method used to collect the sample data.

التفكير في الطريقة المستخدمة لجمع البيانات النموذجي

Part 2

Beyond the Basics of Measures of Center

ما وراء أساسيات التدابير من مركز

Mean from a Frequency Distribution يعني من توزيع التردد

Assume that all sample values in each class are equal to the class midpoint.

نفترض أن كل القيم النموذجية في كل فئة على قدم المساواة إلى منتصف الطبقة.

Mean from a Frequency Distribution

يعني من توزيع التردد

use class midpoint of classes for variable x

x استخدام منتصف فئة من فئات للمتغير

$$\bar{x} = \frac{\Sigma(f \cdot x)}{\Sigma f}$$

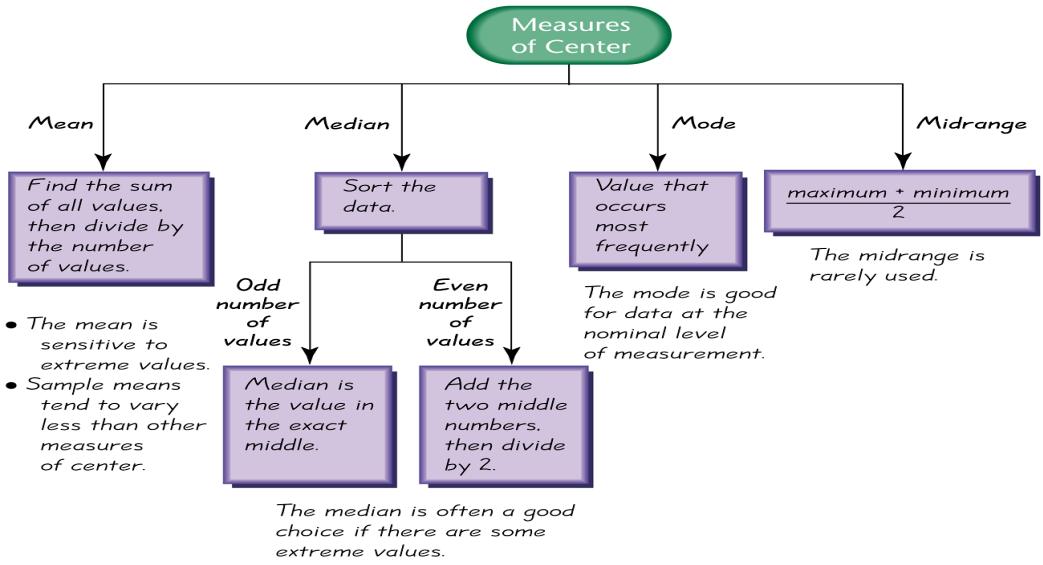
Weighted Mean المعنى الحقيقي

When data values are assigned different weights, we can compute a weighted mean.

عندما يتم تعيين قيم البيانات أوزان مختلفة، يمكننا أن تحسب على المتوسط الموزون.

$$\bar{x} = \frac{\Sigma(w \cdot x)}{\Sigma w}$$

Best Measure of Center أفضل مقياس للمركز



Skewed and Symmetric منحرفا ومتماثل

- **Symmetric متماثل**

distribution of data is symmetric if the left half of its histogram is roughly a mirror image of its right half

توزيع البيانات متماثل إذا كان النصف الأيسر من الرسم البياني لهو تقريبا صورة طبق الأصل من النصف الأيمن ل

- **Skewed منحرفا**

distribution of data is skewed if it is not symmetric and extends more to one side than the other

يميل توزيع البيانات إذا لم يكن متماثل ويمتد جانب أكثر من واحد من الآخر

Skewed Left or Right تميل إلى اليسار أو اليمين

- **Skewed to the left** تميل إلى اليسار

(also called negatively skewed) have a longer left tail, mean and median are to the left of the mode

(وتسمى أيضا منحرفة سلبا) لديها ذيل أطول اليسار، يعني وسيطة هي إلى اليسار من وضع

- **Skewed to the right تميل إلى اليمين**

(also called positively skewed) have a longer right tail, mean and median are to the right of the mode

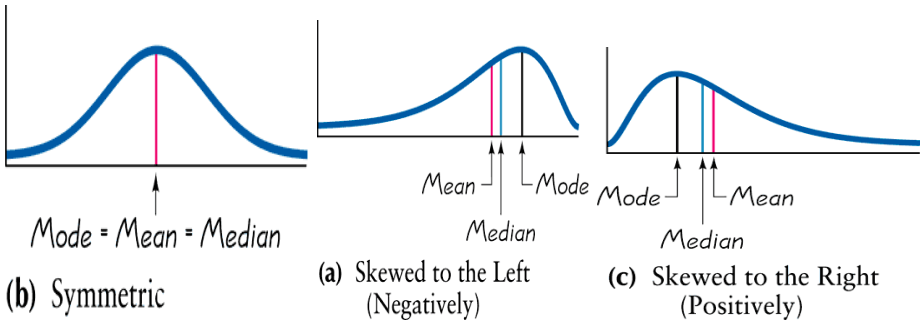
(وتسمى أيضا منحرفا بشكل إيجابي) لديها ذيل أطول الصحيح، يعني وسيطة هي على يمين وضع

Shape of the Distribution شكل التوزيع

The mean and median cannot always be used to identify the shape of the distribution.

المتوسط والوسيط لا يمكن دائما أن تستخدم لتحديد شكل التوزيع.

Skewness انحراف



Recap خلاصة

In this section we have discussed:في هذا القسم ناقشنا

- **Types of measures of center أنواع التدابير من مركز**

Mean تعني

Median الوسيط

Mode الوضع

- **Mean from a frequency distribution يعني من توزيع الترددات**
- **Weighted means وسائل مرجح**
- **Best measures of center أفضل مقاييس مركز**
- **Skewness انحراف**

Section 3-3 Measures of Variation القسم 3-3 تدابير التغيير

Key Concept المفهوم الرئيسي

Discuss characteristics of variation, in particular, measures of variation, such as standard deviation, for analyzing data.

مناقشة خصائص الاختلاف، ولا سيما التدابير من الاختلاف، مثل الانحراف المعياري، لتحليل البيانات.

Make understanding and interpreting the standard deviation a priority.

جعل فهم وتفسير الانحراف المعياري أولوية.

Part 1

Basics Concepts of Measures of Variation

أساسيات مفاهيم تدابير التغيير

Definition فريف

The range of a set of data values is the difference between the maximum data value and the minimum data value.

مجموعة من مجموعة من القيم البيانات هو الفرق بين قيمة البيانات القصوى والدنيا قيمة البيانات.

Range = (maximum value) – (minimum value)

(مجموعة = (القيمة القصوى) - (قيمة الحد الأدنى

It is very sensitive to extreme values; therefore not as useful as other measures of variation.

أنه حساس جدا إلى القيم المتطرفة. وليست مفيدة مثل غيرها من التدابير من الاختلاف.

Round-Off Rule for Measures of Variation

جولة أوف القاعدة لتدابير من الاختلاف

When rounding the value of a measure of variation, carry one more decimal place than is present in the original set of data.

عندما التقريب قيمة تدبير الاختلاف، وحمل واحد أكثر مكان عشري مما هو موجود في المجموعة الأصلية من البيانات.

Round only the final answer, not values in the middle of a calculation.

جولة فقط الجواب النهائي، وليس القيم في منتصف عملية حسابية.

Definition فريف

The standard deviation of a set of sample values, denoted by *s*, is a measure of variation of values about the mean.

الانحراف المعياري لمجموعة من القيم النموذجية، الرمز بواسطة الصورة، هو قياس التغير في القيم عن المتوسط

Sample Standard Deviation Formula نموذج الانحراف الفورمولا القياسية

$$s = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n - 1}}$$

Sample Standard Deviation (Shortcut Formula)

(نموذج الانحراف المعياري (الفورمولا اختصار

$$s = \sqrt{\frac{n\Sigma(x^2) - (\Sigma x)^2}{n(n-1)}}$$

Standard Deviation - Important Properties الانحراف المعياري - خصائص هامة

The standard deviation is a measure of variation of all values from the mean.

الانحراف المعياري هو مقياس لاختلاف كل القيم من الوسط.

- **The value of the standard deviation *s* is usually positive.**

قيمة الانحراف المعياري الصورة هي عادة إيجابية.

❖ **The value of the standard deviation *s* can increase dramatically with the inclusion of one or more outliers (data values far away from all others).**

قيمة الانحراف المعياري الصورة يمكن أن تزيد بشكل كبير مع إدراج واحد أو أكثر من القيم المتطرفة (قيم البيانات بعيدا عن كل الآخرين).

❖ **The units of the standard deviation *s* are the same as the units of the original data values.**

وحدات الانحراف المعياري الصورة هي نفس وحدات من قيم البيانات الأصلية

Comparing Variation in Different Samples مقارنة الاختلاف في عينات مختلفة

It's a good practice to compare two sample standard deviations only when the sample means are approximately the same.

انها ممارسة جيدة للمقارنة بين الانحرافات عينة القياسية فقط عندما يكون وسيلة العينة هي تقريبا نفس.

When comparing variation in samples with very different means, it is better to use the coefficient of variation, which is defined later in this section.

عند مقارنة الاختلاف في العينات مع وسائل مختلفة للغاية، فمن الأفضل استخدام معامل الاختلاف، والذي يعرف لاحقا في هذا القسم.

Population Standard Deviation الانحراف المعياري

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

This formula is similar to the previous formula, but instead, the population mean and population size are used.

هذه الصيغة هي مشابهة للصيغة السابقة، ولكن بدلا من ذلك، فإن عدد السكان يعني وتستخدم حجم السكان.

Variance التباين

The variance of a set of values is a measure of variation equal to the square of the standard deviation.

التباين في مجموعة من القيم هو مقياس لتباين يساوي مربع الانحراف المعياري.

❖**Sample variance: s^2 - Square of the sample standard deviation *s***

تباين العينة: S2 - ساحة للعينة الانحراف المعياري الصورة

❖**Population variance: σ^2 - Square of the population standard deviation σ**

تباين المجتمع ⏭ σ^2 ساحة الانحراف المعياري σ

Unbiased Estimator مقدر غير متحيز

The sample variance s^2 is an unbiased estimator of the population variance σ^2, which means values of s^2 tend to target the value of σ^2 instead of systematically tending to overestimate or underestimate σ^2.

بدلا من S2 تميل لاستهداف قيمة S2 تباين المجتمع، وهو ما يعني قيم S2عينة التباين هو مقدر غير متحيز لل S2و S2. يميل بشكل منهجي إلى المبالغة في تقدير أو نقلل

Variance - Notation التباين - الترقيم

***s* = *sample* standard deviation** الصورة = عينة الانحراف المعياري

s^2 = ***sample* variance** S2 = عينة التباين

σ = ***population* standard deviation** □ = الانحراف المعياري السكان

σ^2 = ***population* variance** □2 = تباين المجتمع

Part 2

Beyond the Basics of Measures of Variation ما وراء أساسيات تدابير التغيير

Range Rule of Thumb مجموعة القاعدة من الإبهام

is based on the principle that for many data sets, the vast majority (such as 95%) of sample values lie within two standard deviations of the mean.

ويستند على مبدأ أن للعديد من مجموعات البيانات، فإن الغالبية العظمى (مثل 95٪) من عينة القيم تقع ضمن اثنين الانحرافات المعيارية للمتوسط.

Range Rule of Thumb for Interpreting a Known Value of the Standard Deviation

مجموعة القاعدة من الإبهام لتفسير قيمة معروفة من الانحراف المعياري

Informally define *usual* values in a data set to be those that are typical and not too extreme. Find rough estimates of the minimum and maximum "usual" sample

values as follows:

تعريف رسمي القيم المعتادة في مجموعة البيانات إلى أن تكون تلك التي هي نموذجية وغير متطرفة جدا. البحث عن تقديرات تقريبية من الحد الأدنى والحد الأقصى "المعتادة" القيم العينة على النحو التالي:

Minimum "usual" value = (mean) – 2 ✎(standard deviation)

(قيمة الحد الأدنى "المعتادة" = (متوسط) - 2 (الانحراف المعياري

Maximum "usual" value = (mean) + 2 ✎(standard deviation)

(الحد الأقصى لقيمة "المعتادة" = (الوسط) + 2 (الانحراف المعياري

Range Rule of Thumb for Estimating a Value of the Standard Deviation *s*

مجموعة القاعدة من الإبهام لتقدير قيمة من الانحراف المعياري الصورة

To roughly estimate the standard deviation from a collection of known sample data use

لتقدير ما يقرب من الانحراف المعياري من مجموعة من استخدام البيانات عينة معروفة

(مجموعة = (القيمة القصوى) - (قيمة الحد الأدنى

$$s \approx \frac{\text{range}}{4}$$

where

range = (maximum value) – (minimum value)

Properties of the Standard Deviation خصائص الانحراف المعياري

Measures the variation among data values

تدابير ¥ الاختلاف بين قيم البيانات

Values close together have a small standard deviation, but values with much more variation have a larger standard deviation

القيم قريبة من بعضها البعض لديهم انحراف معياري صغير، ولكن القيم بقدر أكبر من الاختلاف لديها انحراف معياري أكبر

Has the same units of measurement as the original data

لديه نفس وحدات القياس كما البيانات الأصلية

For many data sets, a value is *unusual* if it differs from the mean by more than two standard deviations

بالنسبة لكثير من مجموعات البيانات، قيمة غير عادية إذا كان يختلف عن المتوسط بأكثر من اثنين الانحرافات المعيارية

Compare standard deviations of two different data sets only if the they use the same scale and units, and they have means that are approximately the same

مقارنة الانحرافات المعيارية للمجموعتين بيانات مختلفة إلا إذا كان أنها تستخدم نفس حجم وحدة، ولديهم الوسائل التي هي تقريبا نفس

Empirical (or 68-95-99.7) Rule القاعدة التجريبية (أو 68-95-99.7)

For data sets having a distribution that is approximately bell shaped, the following properties apply:

لمجموعات البيانات التي لديها توزيع هذا هو تقريبا على شكل جرس، تنطبق الخصائص التالية:

- **About 68% of all values fall within 1 standard deviation of the mean.**

حوالي 68٪ من جميع القيم تقع ضمن نطاق 1 الانحراف المعياري للمتوسط.

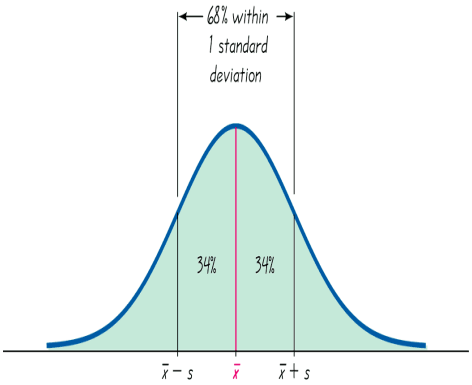
- **About 95% of all values fall within 2 standard deviations of the mean.**

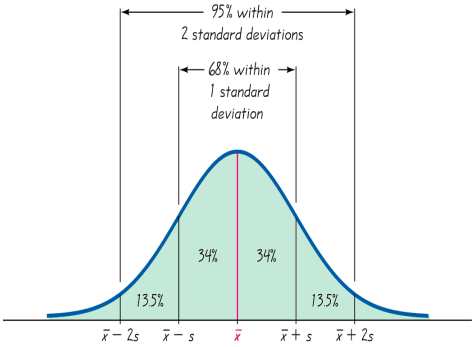
حوالي 95٪ من جميع القيم تندرج 2 الانحرافات المعيارية للمتوسط.

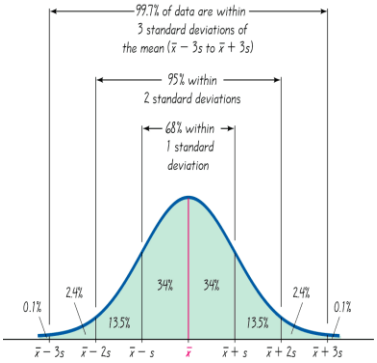
- **About 99.7% of all values fall within 3 standard deviations of the mean.**

عن 99.7٪ من إجمالي قيم تندرج ضمن 3 الانحرافات المعيارية للمتوسط.

The Empirical Rule القاعدة التجريبية







Chebyshev's Theorem نظرية

The proportion (or fraction) of any set of data lying within *K* standard deviations of the mean is always at least $1-1/K^2$, where *K* is any positive number greater than 1.

نسبة (أو جزء) من أي مجموعة من البيانات التي تقع ضمن الانحرافات مستوى كاف من متوسط دائما 1-1 ما لا يقل عن / هو أي رقم موجب أكبر من 1 K ، حيثK2.

- **For *K* = 2, at least 3/4 (or 75%) of all values lie within 2 standard deviations of the mean.**

لK = 24/3 ، على الأقل (أو 75٪) من جميع القيم تقع ضمن 2 الانحرافات المعيارية للمتوسط.

- **For *K* = 3, at least 8/9 (or 89%) of all**

values lie within 3 standard deviations of the mean.

على الأقل (أو 89٪) من جميع القيم تقع ضمن 3 معيار ، 39/8 = Kل

انحرافات المتوسط

Rationale for using *n* – 1 versus *n* الأساس المنطقي لاستخدام ن - 1 مقابل ن

There are only *n* – 1 independent values. With a given mean, only *n* – 1 values can be freely assigned any number before the last value is determined.

لا يوجد سوى ن - 1 القيم مستقلة. مع وسيلة معينة، إلا ن - 1 القيم يمكن تعيين بحرية أي عدد قبل أن يحدد القيمة الأخيرة.

Dividing by *n* – 1 yields better results than dividing by *n*. It causes s^2 to target $?^2$ whereas division by *n* causes s^2 to underestimate $?^2$.

□2 أن نقلل S2 بينما شعبه ن يسبب □2 لاستهداف S2 قسمة ن - 1 يعطي نتائج أفضل من قسمة ن. أنه يسبب.

Coefficient of Variation معامل الاختلاف

The coefficient of variation (or CV) for a set of nonnegative sample or population data, expressed as a percent, describes the standard deviation relative to the mean.

معامل الاختلاف (أو السيرة الذاتية) لمجموعة من البيانات عينة أو السكان غير سلبي، كنسبة مئوية، ويصف الانحراف المعياري بالنسبة للوسيط.

Sample

$$CV = \frac{s}{\bar{x}} \cdot 100\%$$

Population

$$CV = \frac{\sigma}{\mu} \cdot 100\%$$

Recap خلاصة

In this section we have looked at:

في هذا القسم لقد بحثنا في:

- **Range** **نطاق**
- **Standard deviation of a sample and population**

الانحراف المعياري للعينة والسكان

- **Variance of a sample and population التباين في العينة والسكان**
- **Range rule of thumb** **حكم مجموعة من الإبهام**
- **Empirical distribution توزيع التجريبية**
- **Chebyshev's theorem نظريه**
- **Coefficient of variation (CV)** معامل الاختلاف (CV)

Section 3-4 Measures of Relative Standing and Boxplots

القسم 3-4 تدابير الدائمة النسبي وBoxplots

Key Concept المفهوم الرئيسي

This section introduces measures of relative standing, which are numbers showing the location of data values relative to the other values within a data set. They can be used to compare values from different data sets, or to compare values within the same data set. The most important concept is the *z* score. We will also discuss percentiles and quartiles, as well as a new statistical graph called the boxplot.

يقدم هذا القسم تدابير مكانة النسبي، وهي عبارة عن أرقام توضح موقع لقيم البيانات النسبية للقيم أخرى ضمن مجموعة البيانات. ويمكن استخدامها لمقارنة القيم من مجموعات البيانات المختلفة، أو لمقارنة القيم ضمن نفس مجموعة البيانات. مفهوم الأكثر أهمية هو النتيجة ض. سنناقش أيضا النسب المئوية والربعية، وكذلك رسم بياني إحصائي جديد يسمى boxplot.

Part 1

Basics of *z* Scores, Percentiles, Quartiles, and Boxplots

أساسيات عشرات ض، النسب المئوية Percentiles، الشريحة الربعية، وBoxplots

Z score Z النتيجة

- ***z* Score (or standardized value) عدد Z (أو قيمة معيارية)**

the number of standard deviations that a given value *x* is above or below the mean

عدد الانحرافات المعيارية أن قيمة س معين هو أعلى أو أدنى من المتوسط

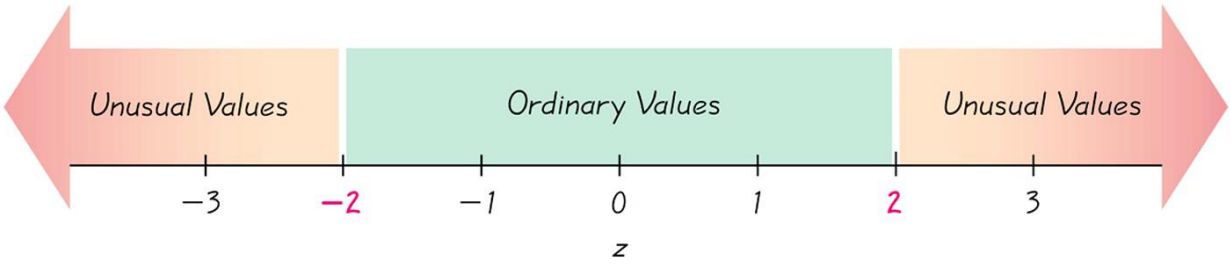
Measures of Position *z* Score تدابير الوظيفة ض نتيجة

Sample **Population**

$$z = \frac{x - \bar{x}}{s} \qquad z = \frac{x - \mu}{\sigma}$$

Round *z* scores to 2 decimal places عشرات ض وإيابا إلى 2 عشرية

Interpreting *Z* Scores العشرات Z تفسير



Whenever a value is less than the mean, its corresponding *z* score is negative

كلما كانت قيمة أقل من المتوسط، ض نتيجة المناظرة سلبية

Ordinary values: –2 ≤ *z* score ≤ 2

النتيجة ≥ 2 Z ≤ القيم العادية: -2

Unusual Values: *z* score < –2 or *z* score > 2

قيم غير عادية: ض النتيجة <-2 ض أو درجة> 2

Percentiles

are measures of location. There are 99 percentiles denoted $P_1, P_2, \ldots P_{99}$, which divide a set of data into 100 groups with about 1% of the values in each group.

، التي تقسم مجموعة من البيانات إلى 100 مجموعة مع حوالي P99 . . .، P2، P1 هي مقاييس الموقع. هناك 99 مئوي الرمز
1٪ من القيم الموجودة في كل مجموعة.

Finding the Percentile of a Data Value العثور على النسبة المئوية لقيمة البيانات

$$\text{Percentile of value } x = \frac{\text{number of values less than } x}{\text{total number of values}} \cdot 100$$

Converting from the *k*th Percentile to the Corresponding Data Value

إلى بيانات القيمة المقابلة KTH تحويل من المئين

Notation

$$L = \frac{k}{100} \cdot n$$

n **total number of values in the data set**

ن عدد من القيم في مجموعة البيانات

k **percentile being used**

المئين ك المستخدمة

L **locator that gives the position of a value**

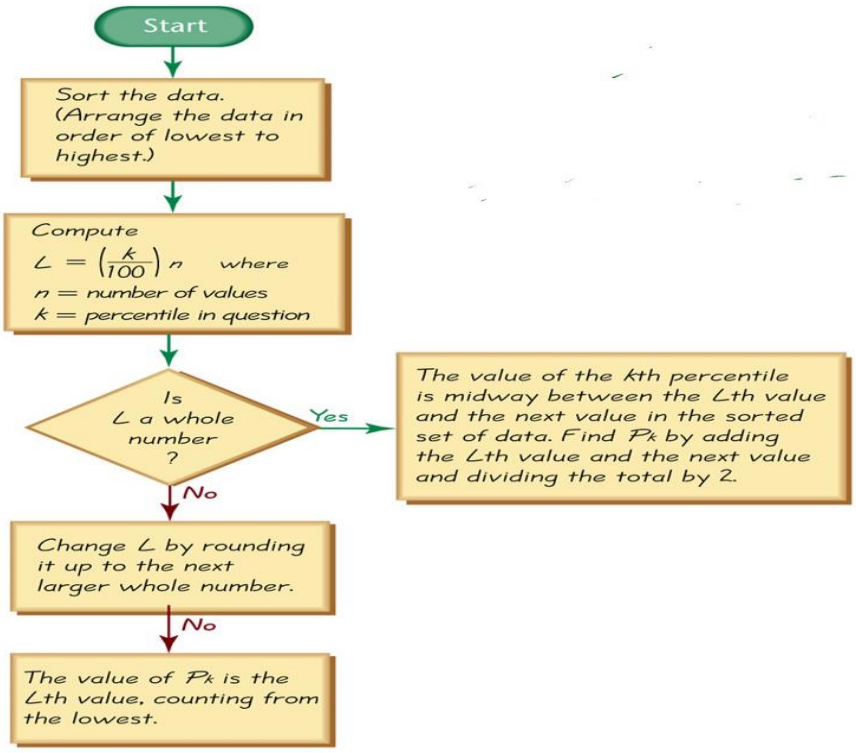
لتحديد المواقع التي تعطي الموقف من قيمة L

P_k ***k*th percentile**

المئوية KTH كيه

Converting from the *k*th Percentile to the Corresponding Data Value

إلى بيانات القيمة المقابلة KTH تحويل من المئين



Quartiles الربعية

Are measures of location, denoted Q_1, Q_2, and Q_3, which divide a set of data into four groups with about 25% of the values in each group.

هي مقاييس موقع، Q1الرمز ، Q2، Q3 من القيم 25٪ حوالي مجموعات مع أربع مجموعات إلى البيانات تقسم مجموعة من والتي ، الموجودة في كل مجموعة.

❖ **Q_1 (First Quartile) separates the bottom 25% of sorted values from the top 75%.**

Q1 (75٪ أعلى من مرتبة القيم من 25٪ أسفل بين يفصل الربعية) الأولى.

❖ **Q_2 (Second Quartile) same as the median; separates the bottom 50% of sorted values from the top 50%.**

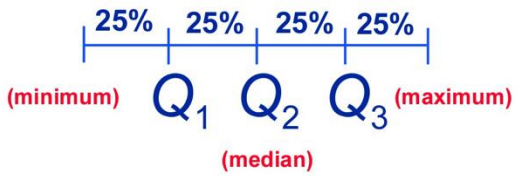
Q2 (50٪ أعلى من مرتبة القيم من 50٪ أسفل بين يفصل المتوسط. نفس الثانية) الربعية الشريحة.

❖ **Q_3 (Third Quartile) separates the bottom 75% of sorted values from the top 25%.**

الشريحة الربعية الثالثة) يفصل بين أسفل 75٪ من القيم مرتبة من أعلى 25٪) Q3.

$$Q_1,\ Q_2,\ Q_3$$

divide ranked scores into four equal parts



Some Other Statistics

❖**Interquartile Range (or IQR):** $Q_3 - Q_1$

الشرائح الربعية المدى (أو الربعية):

❖**Semi-interquartile Range: مدى نصف الشرائح الربعية:**

❖ **Midquartile:**

❖ **10 - 90 Percentile Range:** $P_{90} - P_{10}$

5-Number Summary عدد ملخص

❖**For a set of data, the 5-number summary consists of the minimum value; the first quartile**

لمجموعة من البيانات، وملخص 5 عدد يتكون من قيمة الحد الأدنى. في الربع الأول

Q_1 **; the median (or second quartile وسط (أو الربع الثاني**

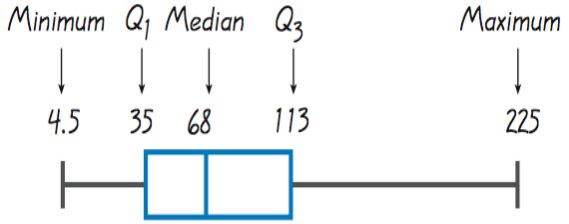
Q_2**); the third quartile, في الربع الثالث**

Q_3 **; and the maximum value. والقيمة القصوى**

Boxplot مربع مؤامرة

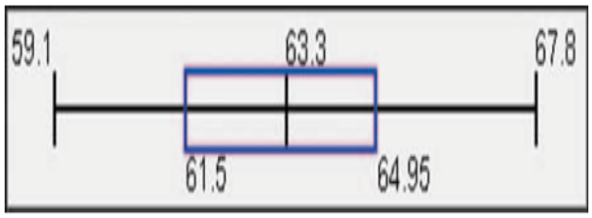
❖ **A boxplot (or box-and-whisker- diagram) is a graph of a data set that consists of a line extending from the minimum value to the maximum value, and a box with lines drawn at the first quartile,** Q_1 **; the median; and the third quartile,** Q_3**.**

الرسم البياني) هو الرسم البياني لمجموعة البيانات التي تتكون من الخط يمتد من قيمة الحد -whiskerأو مربع و) boxplotو
Q3. المتوسط؛ والربع الثالث .Q1الأدنى إلى الحد الأقصى المسموح به، ومربع مع الخطوط المرسومة في الربع الأول،



Boxplot of Movie Budget Amounts Boxplot من ميزانية الفيلم المبالغ

Boxplots - Normal Distribution Boxplots - توزيع عادى



Normal Distribution: Heights from a Simple Random Sample of Women

التوزيع الطبيعي:

مرتفعات من عينة عشوائية بسيطة من النساء

Boxplots - Skewed Distribution Boxplots - توزيع مشوه

Skewed Distribution: Salaries (in thousands of dollars) of NCAA Football Coaches

توزيع الانحراف:

الرواتب (بآلاف الدولارات) من الرابطة الوطنية لرياضة كرة القدم المدربين

Part 2

Outliers and Modified Boxplots التعديل Boxplotsالقيم المتطرفة و

Outliers القيم المتطرفة

❖ **An outlier is a value that lies very far away from the vast majority of the other values in a data set.**

عزلاء هو القيمة التي تقع بعيدا جدا من أن الغالبية العظمى من القيم الأخرى في مجموعة البيانات.

Important Principles مبادئ هامة

❖ **An outlier can have a dramatic effect on the mean.**

عزلاء يمكن أن يكون لها تأثير كبير على متوسط

❖ **An outlier can have a dramatic effect on the standard deviation.**

عزلاء يمكن أن يكون لها تأثير كبير على الانحراف المعياري.

❖ **An outlier can have a dramatic effect on the scale of the histogram so that the true nature of the distribution is totally obscured.**

عزلاء يمكن أن يكون لها تأثير كبير على حجم الرسم البياني بحيث الطبيعة الحقيقية للتوزيع تحجب تماما

Outliers for Modified Boxplots القيم المتطرفة لBoxplots التعديل

For purposes of constructing *modified boxplots*, we can consider outliers to be data values meeting specific criteria.

تعديل، يمكن أن نعتبر القيم المتطرفة أن تكون قيم البيانات تلبية معايير محددة boxplots لأغراض بناء.

In modified boxplots, a data value is an outlier if it is . . .

above Q_3 by an amount greater than 1.5 ⌲IQR

or below Q_1 by an amount greater than 1.5 ⌲IQR

تعديل، قيمة البيانات هي عزلاء إذا كان boxplots في. . .

بمبلغ أكبر من 1.5 الربعية Q3 فوق

أو

بمبلغ أكبر من 1.5 الربعي Q1 أدناه

Modified Boxplots Boxplots تعديل

Boxplots described earlier are called skeletal (or regular) boxplots.

Boxplots هو موضح سابقا boxplots (أو العادية الهيكل العظمي).

Some statistical packages provide modified boxplots which represent outliers as

special points.

التي تمثل القيم المتطرفة كنقاط خاصة تعديلها boxplots بعض الحزم الإحصائية توفر.

Modified Boxplot Construction تعديل البناء Boxplot

A modified boxplot is constructed with these specifications:

تعديل لهذه المواصفات boxplot يتم إنشاء:

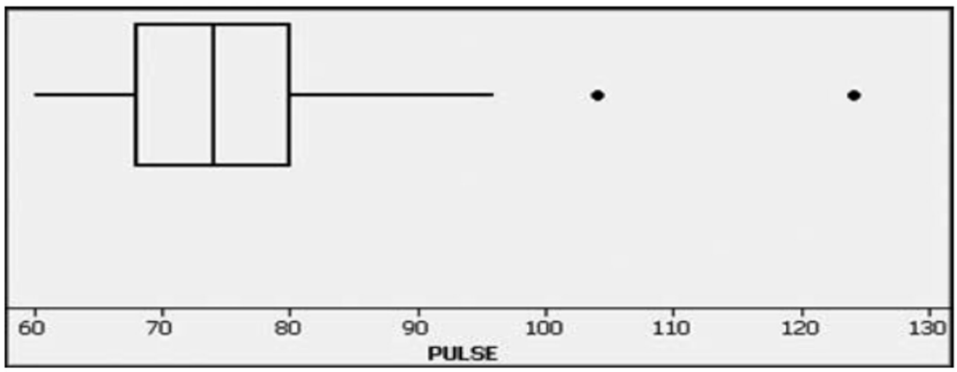
- **A special symbol (such as an asterisk) is used to identify outliers.**

يتم استخدام رمز خاص (مثل النجمة) لتحديد القيم المتطرفة.

- **The solid horizontal line extends only as far as the minimum data value that is not an outlier and the maximum data value that is not an outlier.**

الخط الأفقي الصلبة يمتد فقط بقدر ما قيمة البيانات الدنيا التي ليس عزلاء والحد الأقصى لقيمة البيانات التي ليست عزلاء

Modified Boxplots - Example Boxplots المعدلة - مثال



Pulse rates of females listed in Data Set 1 in Appendix B.

(معدلات النبض من الإناث المدرجة في بيانات مجموعة 1 في الملحق (ب

Recap خلاصة

In this section we have discussed: في هذا القسم ناقشناها

- ***z* Scores ض عشرات**
- ***z* Scores and unusual values ض عشرات والقيم غير عادية**
- **Percentiles النسب المئوية**

- **Quartiles** الربعية
- **Converting a percentile to corresponding data values**

تحويل المئوية لقيم البيانات المناظرة

- **Other statistics** إحصاءات أخرى
- **5-number summary** ملخص 5 عدد
- **Boxplots and modified boxplots** Boxplots وboxplots تعديل
- **Effects of outliers** آثار القيم المتطرف

Putting It All Together ضع كل شيء معا

Always consider certain key factors: تنظر دائما بعض عوامل رئيسية هي :

- **Context of the data** سياق البيانات
- **Source of the data** مصدر البيانات
- **Sampling Method** طريقة أخذ العينات
- **Measures of Center** تدابير من مركز
- **Measures of Variation** تدابير التغيير
- **Distribution** توزيع
- **Outliers** القيم المتطرفة
- **Changing patterns over time** تغيير أنماط على مر الزمن
- **Conclusions** الاستنتاجات
- **Practical Implications** نواتج عملية