Measure of Center

❖ **Measure of Center**

the value at the center or middle of a data set

1. Mean
2. Median
3. Mode
4. Midrange (rarely used)

Mean

❖ **Arithmetic Mean (Mean)**

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

What most people call an *average*.

Notation

Σ denotes the sum of a set of values.

x is the variable used to represent the individual data values.

n represents the number of data values in a sample.

N represents the number of data values in a population.

$\bar{x}$ is pronounced 'x-bar' and denotes the mean of a set of sample values

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

This is the sample mean

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

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

This is the population mean

Mean

- **Advantages**
  Is relatively reliable.
  Takes every data value into account

- **Disadvantage**
  Is sensitive to every data value, one extreme value can affect it dramatically; is not a *resistant* measure of center

Median

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

- often denoted by $\tilde{x}$ (pronounced 'x-tilde')

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

Finding the Median

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

1. If the number of data values is odd, the median is the value located in the exact middle of the list.

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

5.40	1.10	0.42	0.73	0.48	1.10
0.42	0.48	0.73	1.10	1.10	5.40

MEDIAN is 0.915

$\frac{0.73 + 1.10}{2}$ (even number of values – no exact middle shared by two numbers)

5.40	1.10	0.42	0.48	0.73	1.10	0.66
0.42	0.48	0.66	0.73	1.10	1.10	5.40

(odd number of values)

exact middle **MEDIAN is 0.73**

Mode

- **Mode**
  the value that occurs with the greatest frequency
- Data set can have one, more than one, or no mode

Bimodal	two data values occur with the same greatest frequency
Multimodal	more than two data values occur with the same greatest frequency
No Mode	no data value is repeated

9

Mode - Examples

a. 5.40 1.10 0.42 0.73 0.48 1.10	⇦ Mode is 1.10
b. 27 27 27 55 55 55 88 88 99	⇦ Bimodal - 27 & 55
c. 1 2 3 6 7 8 9 10	⇦ No Mode

10

Definition

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

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

11

Midrange

- Sensitive to extremes
  because it uses only the maximum and minimum values.
- Midrange is rarely used in practice

12

Round-off Rule for Measures of Center

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

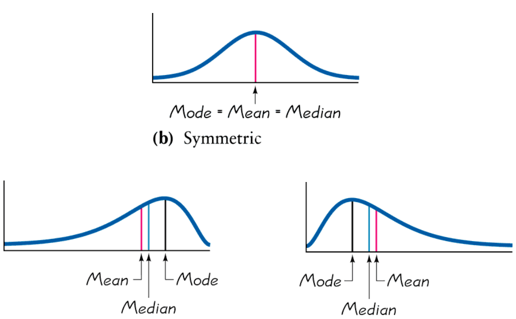
13

Skewed and Symmetric

- **Symmetric**
  distribution of data is symmetric if the left half of its histogram is roughly a mirror image of its right half
- **Skewed**
  distribution of data is skewed if it is not symmetric and extends more to one side than the other

14

Symmetry and skewness



(b) Symmetric

(a) Skewed to the Left (Negatively)

(c) Skewed to the Right (Positively)

15

Measures of Variation

spread, variability of data
width of a distribution

1. **Standard deviation**
2. **Variance**
3. **Range** (rarely used)

Standard deviation

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

Sample Standard Deviation Formula

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

Sample Standard Deviation (Shortcut Formula)

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

Population Standard Deviation

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

σ is pronounced 'sigma'

This formula only has a theoretical significance, it cannot be used in practice.

Variance

- The variance is a measure of variation equal to the square of the standard deviation.
- Sample variance: s^2 - Square of the sample standard deviation s
- Population variance: σ^2 - Square of the population standard deviation σ

Variance - Notation

s = *sample* standard deviation

s^2 = *sample* variance

σ = *population* standard deviation

σ^2 = *population* variance

Usual values in a data set are those that are **typical** and **not too extreme**.

Minimum usual value = (mean) – 2 × (standard deviation)

Maximum usual value = (mean) + 2 × (standard deviation)

Rule of Thumb

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.

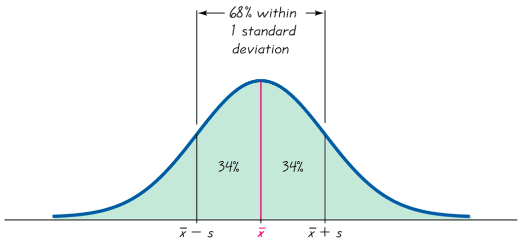
A value is unusual if it differs from the mean by more than two standard deviations.

Empirical (or 68-95-99.7) Rule

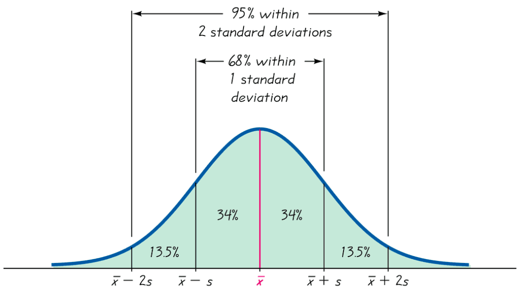
For data sets having a distribution that is approximately bell shaped, the following properties apply:

- **About 68% of all values fall within 1 standard deviation of the mean.**
- **About 95% of all values fall within 2 standard deviations of the mean.**
- **About 99.7% of all values fall within 3 standard deviations of the mean.**

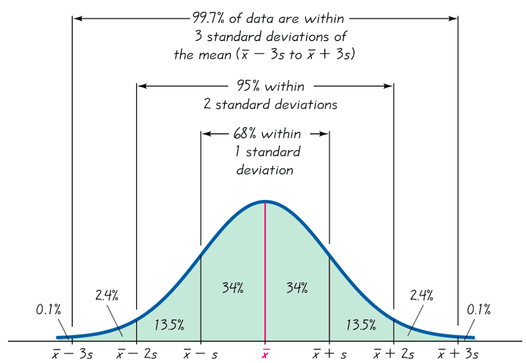
The Empirical Rule



The Empirical Rule



The Empirical Rule



Range (rarely used)

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

Range = (maximum value) – (minimum value)

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

29

Measures of Relative Standing

Z score

- *z* score (or standardized value)

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

31

Measure of Position: *z* score

Sample

$$z = \frac{x - \bar{x}}{s}$$

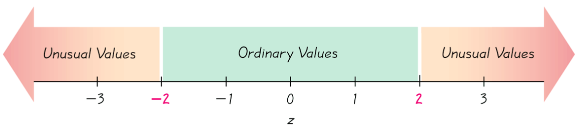
Population

$$z = \frac{x - \mu}{\sigma}$$

Round *z* scores to 2 decimal places

32

Interpreting Z scores



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

Ordinary values: $-2 \le z$ **score** ≤ 2

Unusual values: z **score** < -2 **or** z **score** > 2

33

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.

34

Finding the Percentile of a Data Value

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

Round it off to the nearest whole number

35

Converting from the kth Percentile to the Corresponding Data Value

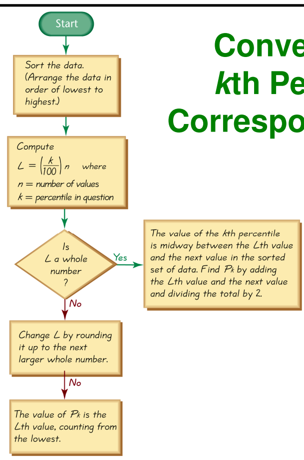
Notation

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

- n total number of values in the data set
- k percentile being used
- L locator that gives the position of a value
- P_k kth percentile

36

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



Quartiles

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.

- Q_1 (First Quartile) separates the bottom 25% of sorted values from the top 75%.
- Q_2 (Second Quartile) same as the median; separates the bottom 50% of sorted values from the top 50%.
- Q_3 (Third Quartile) separates the bottom 75% of sorted values from the top 25%.

Quartiles

$Q_1, \ Q_2, \ Q_3$

divide **ranked** scores into four equal parts

25% | 25% | 25% | 25%

(minimum) Q_1 Q_2 Q_3 (maximum)

(median)

Some Other Statistics

- Interquartile Range (or IQR): $Q_3 - Q_1$
- Semi-interquartile Range: $\dfrac{Q_3 - Q_1}{2}$
- Midquartile: $\dfrac{Q_3 + Q_1}{2}$
- 10 - 90 Percentile Range: $P_{90} - P_{10}$

5-Number Summary

- For a set of data, the **5-number summary** consists of the minimum value; the first quartile Q_1; the median (or second quartile Q_2); the third quartile, Q_3; and the maximum value.