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 (cannot resist the extremes)

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 (resists the extremes) 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**

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

Definition

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

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

Midrange

- **Sensitive to extremes**
  because it uses only the maximum and minimum values.

- Midrange is rarely used in practice

Round-off Rule for Measures of Center

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

The mean of 2,4,4 is 3.3333.. Aprox to 3.**3**

The mean of 80.4 and 80.7 is 80.5**5**

Skewed and Symmetric

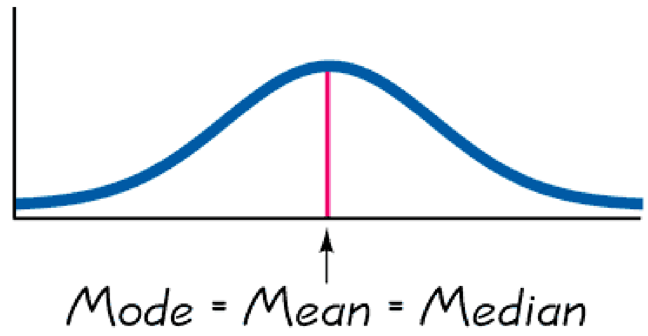
- **Symmetric**

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

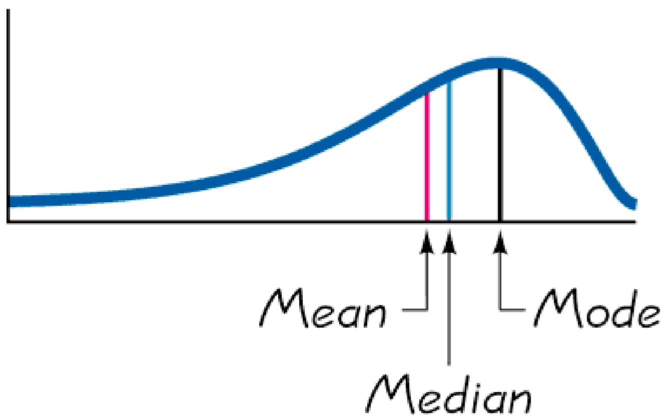
- **Skewed**

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

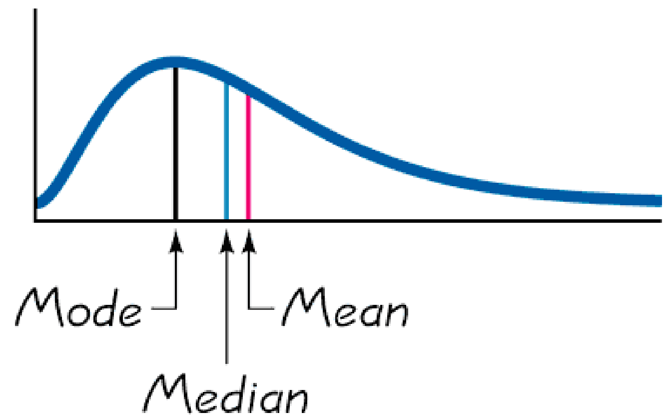
Symmetry and skewness



(b) Symmetric



(a) Skewed to the Left
(Negatively)



(c) Skewed to the Right
(Positively)

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.

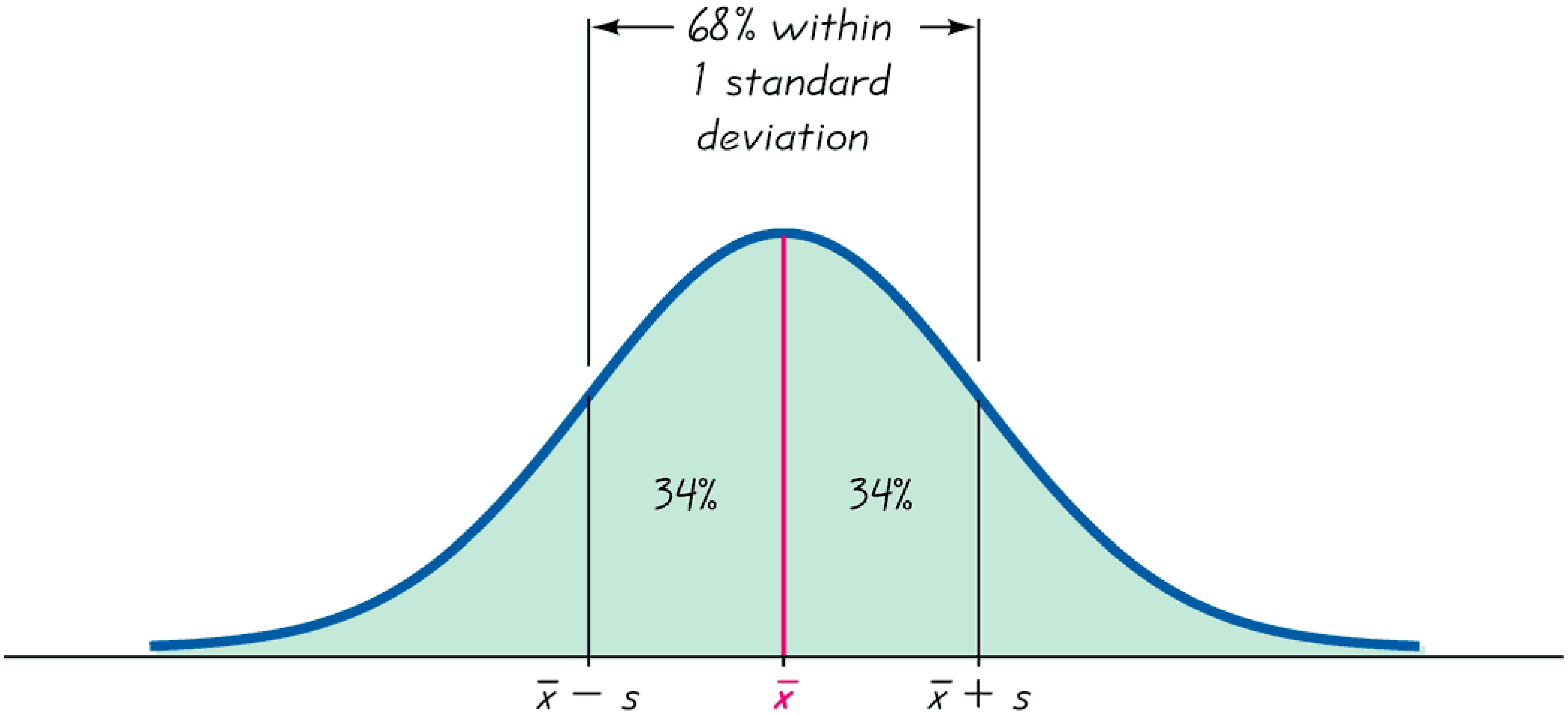
A value is unusual if it differs from the mean by more than two standard deviations.

Empirical (or 68-95-99.7) Rule

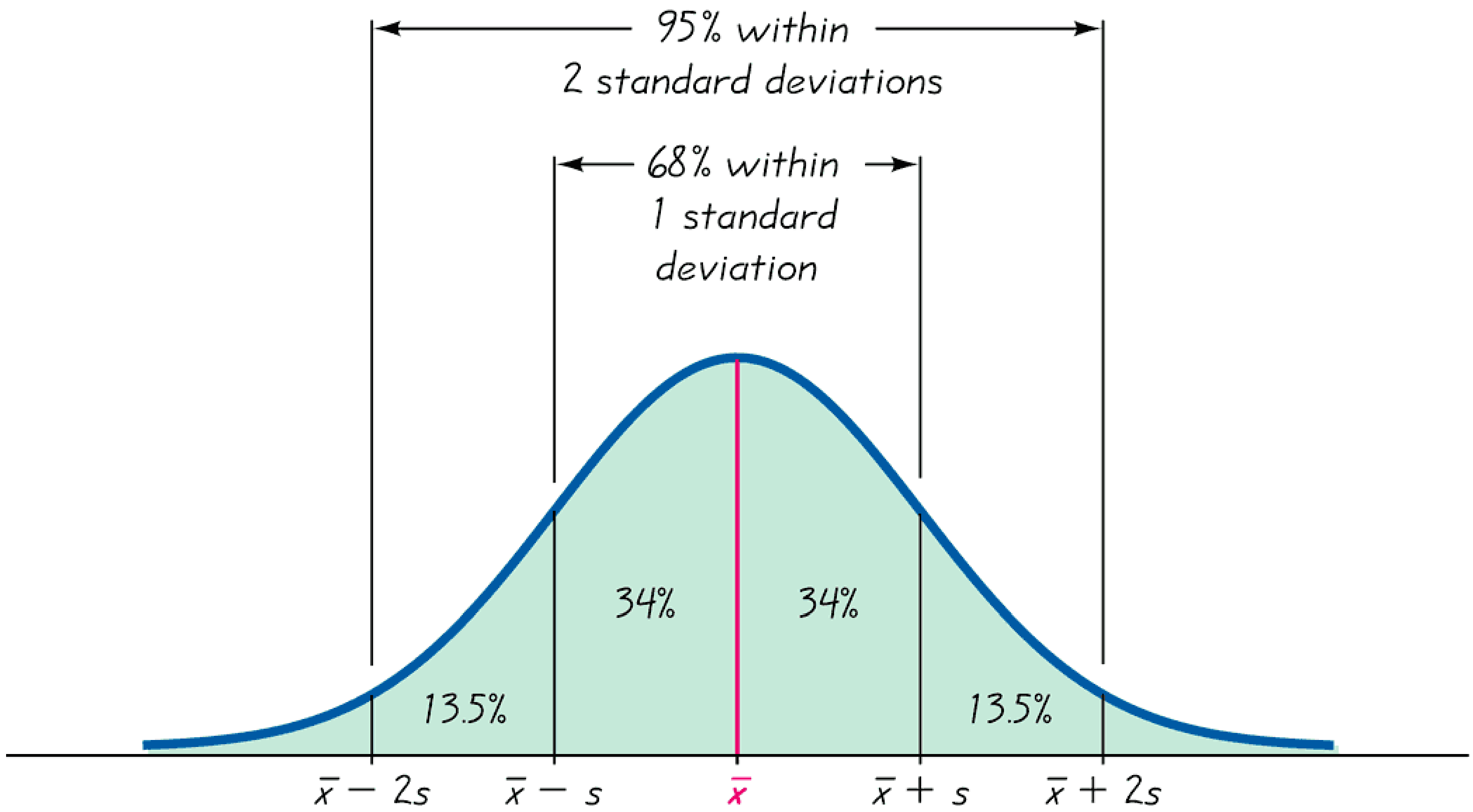
For data sets having a distribution that is approximately bell shaped, the following properties apply:

- 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.

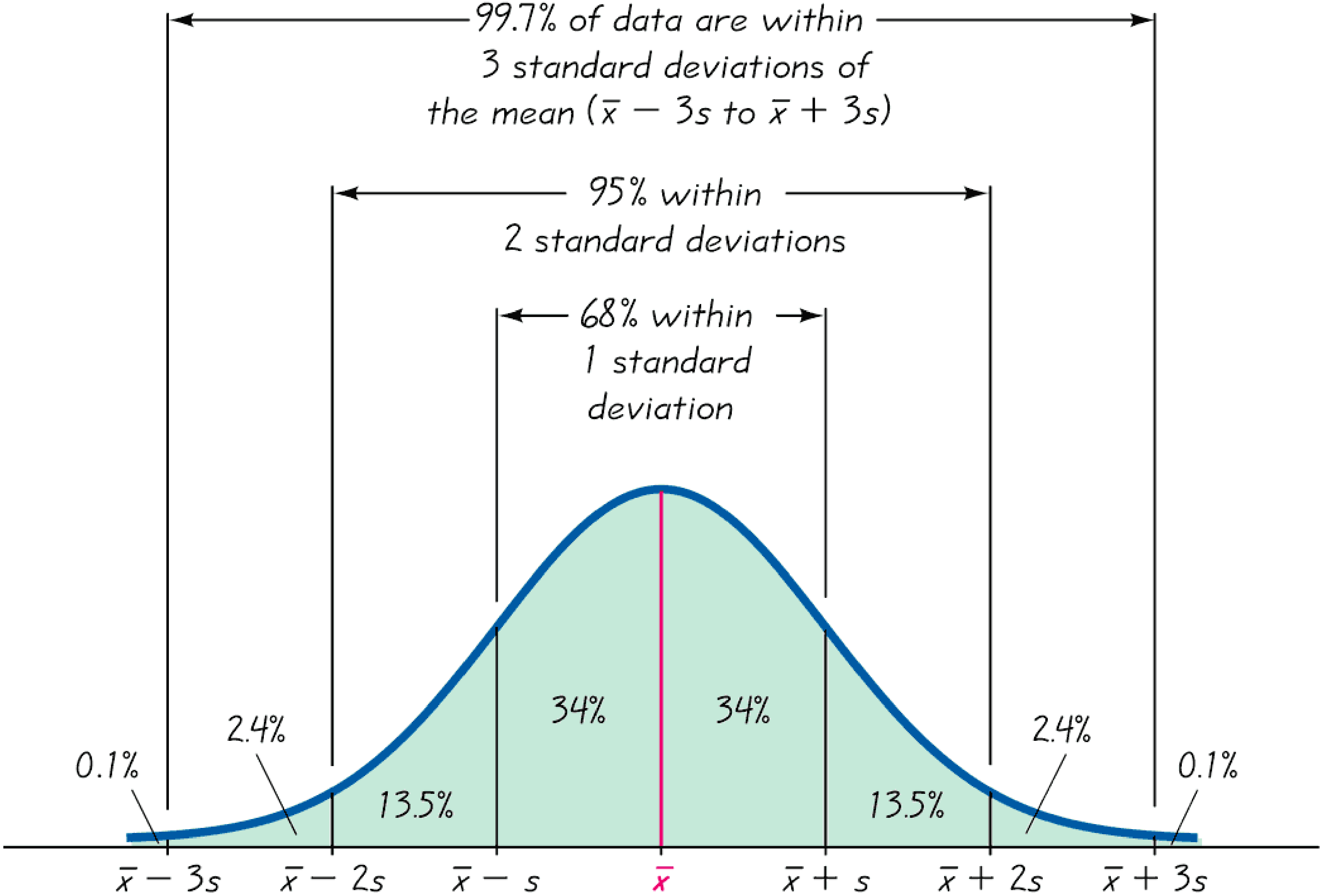
The Empirical Rule



The Empirical Rule



The Empirical Rule



Chebyshev’s Theorem

- The above empirical rule applies only for bell-shaped distributions
- The Chebyshev’s Theorem implies
  - 75% of all values line within 2 standard deviation of the mean
  - 89% of all values line within 3 standard deviation of the mean

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.

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

Measure of Position: *z* score

Sample

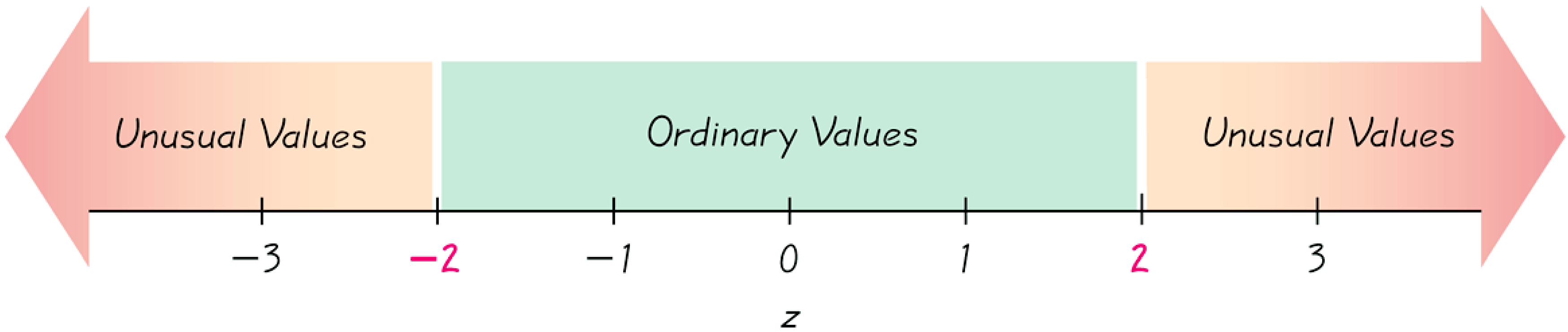
$$z = \frac{x - \bar{x}}{s}$$

Population

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

Round *z* scores to <u>2 decimal places</u>

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

Percentiles

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