Chapter 3 Numerically Summarizing Data
Section 3.1 Measures of Central Tendency

We will discuss

- **Mean**
- **Median**
- **Mode**
- **How do mean and median together tell us about the shape of a distribution.**

→ **Next page**

(Arithmetic) Mean

Correct the following in the book:

Page 107, in the "Definition" near the bottom of the page, change the symbol for the population arithmetic mean from "m" to "μ"
(μ, pronounced "mew", is the counterpart of "m" in Greek alphabet.)

Recall:

Parameters: Descriptive measures of a population.
(Usually represented by Greek letters)

Statistic: Descriptive measures of a sample.
(Usually represented by Roman letters.)

Population Arithmetic Mean: μ (This is a parameter!)

$$\mu = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{\sum x_i}{N}$$

where $N = Population\ Size$

("$\sum$" (pronounced "Sigma")is the Greek counter part of "S". In math it means "Sum".)

Sample Arithmetic Mean: $\overline{x}$ (This is a statistic!!)

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum x_i}{n}$$

where $n = Sample\ Size$

Median (Notation: M)

Arrange all data values in a row, from the smallest to the largest. Note that some of the values may appear more than once.

- Odd number of data points:

  Example: 11 data points

  10, 11, 12, 14, 14, 17, 18, 20, 20, 20, 23

  Median: M = the one at the center
  = the 6th value
  = 17

- Even number of data points:

  Example: 12 data point

  62, 68, 68, 74, 77, 82, 84, 88, 90, 94

  Median: $M = \frac{77+82}{2} = 79.5$

  (The midpoint between the top of the bottom group and the bottom of the top group.)

Mode

The value that appears the most frequently

Example:

10, 11, 12, 14, 14, 17, 18, 20, 20, 20, 23

The mode is 20.

Example:

20, 21, 22, 24, 24, 27, 28, 29, 29, 30, 31

There are two modes: 24 and 29 (They appear equally frequently, and more frequently than any other value.) We say that the data are "bimodal".

Example:

20, 21, 22, 24, 24, 27, 28, 29, 29, 30, 31, 31, 35, 36

There are three modes: 24, 29, 31.
When there are three or more modes, we say the data are "multimodal".

Example:

10, 11, 13, 14, 15, 17, 18, 20, 22, 25, 30, 31, 32, 35

No mode! (All values that do appear appear equally frequently.)

How do mean and median together tell us about the shape of a distribution.

Note that the median is not sensitive to extreme values, whereas the mean is!!

When extreme values are present, it is median becomes a better measure of center.

- **Skewed left: Mean smaller than median.**
- **Symmetric: Mean equal to the median.**
- **Skewed right: Mean larger than median.**

(See animation.)

Section 3.2 Measures of Dispersion

We will discuss

- **Range**
- **Variance**
- **Standard Deviation**

and also

- **Empirical Rule**
- **Chebyshev's Inequality**

Range (Notation: R)

$$R = \text{largest value} - \text{smallest value}$$

Example:

10, 11, 12, 14, 14, 17, 18, 20, 20, 20, 23

$$R = 23 - 10 = 13$$

σ is the lower case of $\sum$ in the Greek alphabet. It corresponds to the lower case s in the English alphabet.

Variance

Population Variance: σ^2 ("sigma squared")

$$\sigma^2 = \frac{(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_N-\mu)^2}{N} = \frac{\sum(x_i-\mu)^2}{N}$$

Sample Variance: s^2

$$s^2 = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n-1} = \frac{\sum(x_i-\bar{x})^2}{n-1}$$

Note that the two formulas are different—

The denominator in the expression for s^2 ***is*** $n-1$***,*** ***not*** n ***!!!!!!!!!!!!***

Unfortunately we are not ready to understand this peculiar $n-1$ at this stage. Wait until we learn the sampling distribution and "unbiased estimators" in a later chapter.

Equivalent Forms More Suitable for Hand Computation:

***** When doing computation, do <u>not</u> round any intermediate values!!!!!! (Otherwise your final answer will be off.)**

Population Variance: σ^2 ("sigma squared")

$$\sigma^2 = \frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{N}}{N}$$

Sample Variance: s^2

$$s^2 = \frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}$$

Standard Deviation

Population Standard Deviation: σ ("sigma")

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} = \sqrt{\frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{N}}{N}}$$

Example: If the population variance was calculated to be $178.8\,\text{min}^2$, then the population standard deviation is

$\sigma = \sqrt{178.8\,\text{min}^2} \approx 13.37161172\,\text{min} \approx 13.4\,\text{min}$

Sample Standard Deviation: s

$$s = \sqrt{s^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}}$$

Example: If the population variance was calculated to be $198.6666667\,\text{min}^2$, then the population standard deviation is

$s = \sqrt{198.6666667\,\text{min}^2} \approx 14.09491634\,\text{min} \approx 14.1\,\text{min}$

***** When doing computation, do <u>not</u> round any intermediate values!!!!!! (Otherwise your final answer will be off.)**

Interpretation of the Standard Deviation

When comparing the distribution of the same variable in two different populations, remember:

> **Larger standard deviation means more dispersion (i.e. more spread)**

The Empirical Rule ("68-95-99.7 percent rule"): Only works well for nearly bell-shaped distributions. [in Sec. 3.2]

Chebyshev's Inequality: Rigorous mathematical statement, but providing only very conservative statements. [in Sec. 3.2]

Section 3.3: Measures of Central Tendency and Dispersion from Grouped Data