

COR 142 C: Section 3.1

Dr. Chris Kao

Summaries for Symmetric Distributions

- Mean: Measure for center
- Standard Deviation: Measure for horizontal spread

Summaries for Symmetric Distributions

Recall: When working with numerical data, we need to describe the distribution using the three characteristics below.

- Shape: Symmetric, skewed, mode, etc
- Center: Typical value
- Spread: Horizontal variability

The Mean for Symmetric Distributions

Example 3.1. The dotplot below shows a fictitious distribution of test scores for a sample of students. The mean is for the dataset is 24.2

The Mean for Symmetric Distributions

Example 3.1. (continued)

a. How would you describe the shape of the distribution?

- The distribution is roughly symmetric.

b. What is a good measure of a “typical” value? Would you say the mean is a good representation of this value?

- Since the distribution is roughly symmetric, the mean is a good measure for the typical value.

Summary: Using the Mean to Describe Typical Values

- When the distribution of the data is (roughly) symmetric, the mean represents the typical value of the dataset.
- When the distribution of the data is skewed, the median is the more appropriate estimate of the typical value.

Computing the Mean

Suppose $\{x_1, x_2, x_3, \dots, x_n\}$ is a dataset. Then the mean, denoted by \bar{x} , of the dataset is defined by

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

Measuring the Spread: Symmetric Distributions

- Recall: The variability in a distribution can be measured by the horizontal spread.
- Why is this important? It is important to know if most of the data is near or far from the center.

Measuring the Spread: Symmetric Distributions

Example 3.2. The following histograms are records of daily high temperature in F in 2017 in St. Augustine, FL and Mt. Pleasant, MI.

Measuring the Spread: Symmetric Distributions

Example 3.2. (continued)

- a. How would you describe the shape of each distribution?
- The distribution for St. Augustine is left-skewed, whereas the distribution for Mt. Pleasant is more symmetric.

Measuring the Spread: Symmetric Distributions

Example 3.2. (continued)

b. Which distribution appears to have wider horizontal spread?

- Mt. Pleasant has a wider horizontal spread because its temperature distribution has a wider range of different temperature values.

Standard Deviation

- **Standard deviation:** It is a measure that can be thought of as the “typical distance” of the observations away from the mean.
- The standard deviation of a symmetric distribution is defined by

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

- For most distributions, a majority of the data is within **one standard deviation** of the mean.

Variance

The variance of a dataset is defined to be the square of its standard deviation: s^2 .

Note

For most applications in this class, it is preferred to the use standard deviation of datasets instead of the variance; the standard deviation has the same unit as the mean.