COR 142 C: Section 2.2

Dr. Chris Kao

Summarizing A Numerical Distribution

Three important features of a numerical distribution:

1. Shape
2. Center
3. Variability

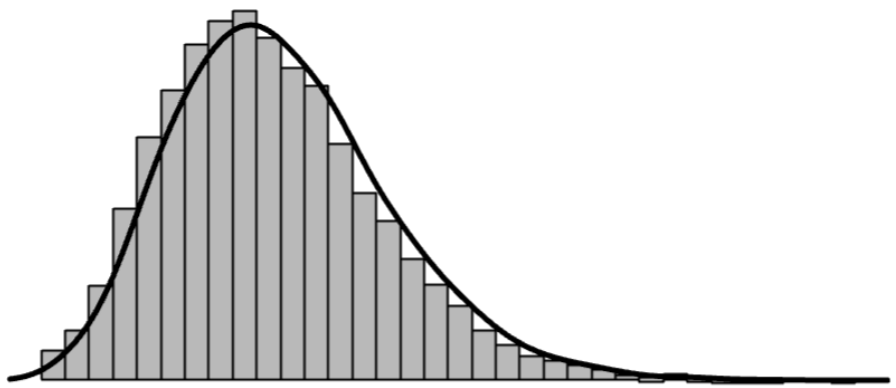
Shape

- Is the distribution symmetric or skewed?
- How many peaks (mounds) are there?
- Are unusually large or small values present?

Shape

Symmetric

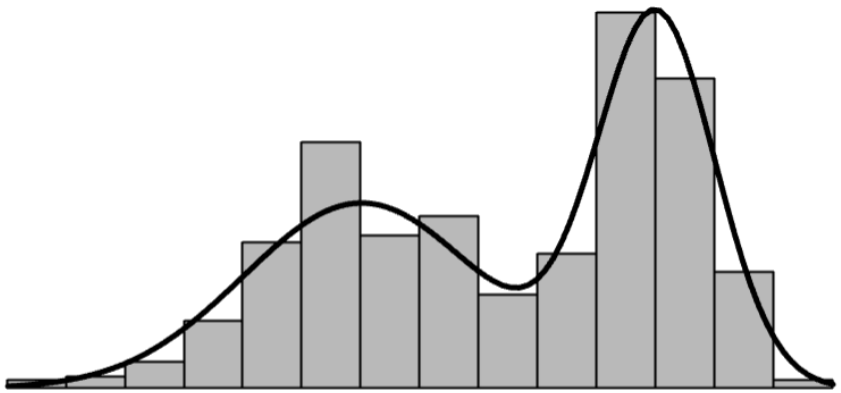
Skewed



Shape

Unimodal

Bimodal



Shape

Multimodal: More than two main peaks.

- Peaks can be different heights.
- Bimodal and multimodal data may indicate existence of different groups within the data.
  - In this case, the data should be separated into two (or more) groups and visualize each group separately.
- **Example:** Plotting men's and women's heights together.

Outliers

- Extremely large or small values.
- Data values that do not fit the pattern of the rest of the data.
- Not precisely defined (subject to opinions)
- When observing extremely large or small values:
  - Report these values
  - Know that they could be sources of error (typos, etc)
  - True outliers are unusually interesting data values.

Center

Center: The typical data value

Variability

Variability: Look at the horizontal spread in the histogram or dotplot.

> **Note**
>
> This is a somewhat subject concept to evaluate by “eyeballing” right now. We will be more rigorous about this definition later.