

2.6: Putting It Together- Summarizing Data Graphically and Numerically

Let's Summarize

In *Summarizing Data Graphically and Numerically*, we focused on describing the *distribution of a quantitative variable*.

- To analyze the distribution of a quantitative variable, we describe the *overall pattern of the data* (shape, center, spread) and any *deviations from the pattern* (outliers). We use three types of graphs to analyze the distribution of a quantitative variable: dotplots, histograms, and boxplots.
- We described the *shape* of a distribution as left-skewed, right-skewed, symmetric with a central peak (bell-shaped), or uniform. Not all distributions have a simple shape that fits into one of these categories.
- The *center* of a distribution is a typical value that represents the group. We have two different measurements for determining the center of a distribution: mean and median.
 - The *mean* is the average. We calculate the mean by adding the data values and dividing by the number of individual data points. The *mean* is the *fair share* measure. The mean is also called the *balancing point* of a distribution. If we measure the distance between each data point and the mean, the distances are balanced on each side of the mean.
 - The *median* is the physical center of the data when we make an ordered list. It has the same number of values above it as below it.
 - **General Guidelines for Choosing a Measure of Center**
 - *Always plot the data.* We need to use a graph to determine the shape of the distribution. By looking at the shape, we can determine which measure of center best describes the data.
 - Use the mean as a measure of center *only* for distributions that are reasonably symmetric with a central peak. When outliers are present, the mean is not a good choice.
 - Use the median as a measure of center for all other cases.
- The *spread* of a distribution is a description of how the data varies. We studied three ways to measure spread: *range* (max – min), the *interquartile range* (Q3 – Q1), and the *standard deviation*. When we use the median, Q1 to Q3 gives a typical range of values associated with the middle 50% of the data. When we use the mean, Mean \pm SD gives a typical range of values.
 - The interquartile range (IQR) measures the variability in the middle half of the data.
 - Standard deviation measures roughly the average distance of data from the mean.
- *Outliers* are data points that fall outside the overall pattern of the distribution. When using the median and IQR to measure center and spread, we use the $1.5 * \text{IQR}$ interval to identify outliers. Specifically, points outside the interval $Q1 - 1.5 * \text{IQR}$ to $Q3 + 1.5 * \text{IQR}$ are labeled as outliers.

Contributors and Attributions

CC licensed content, Shared previously

- Concepts in Statistics. **Provided by:** Open Learning Initiative. **Located at:** <http://oli.cmu.edu>. **License:** [CC BY: Attribution](#)

This page titled [2.6: Putting It Together- Summarizing Data Graphically and Numerically](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Lumen Learning](#).