

2.9: Chapter 2 Review

2.1 Display Data

A **stem-and-leaf plot** is a way to plot data and look at the distribution. In a stem-and-leaf plot, all data values within a class are visible. The advantage in a stem-and-leaf plot is that all values are listed, unlike a histogram, which gives classes of data values. A **line graph** is often used to represent a set of data values in which a quantity varies with time. These graphs are useful for finding trends. That is, finding a general pattern in data sets including temperature, sales, employment, company profit or cost over a period of time. A **bar graph** is a chart that uses either horizontal or vertical bars to show comparisons among categories. One axis of the chart shows the specific categories being compared, and the other axis represents a discrete value. Some bar graphs present bars clustered in groups of more than one (grouped bar graphs), and others show the bars divided into subparts to show cumulative effect (stacked bar graphs). Bar graphs are especially useful when categorical data is being used.

A **histogram** is a graphic version of a frequency distribution. The graph consists of bars of equal width drawn adjacent to each other. The horizontal scale represents classes of quantitative data values and the vertical scale represents frequencies. The heights of the bars correspond to frequency values. Histograms are typically used for large, continuous, quantitative data sets. A frequency polygon can also be used when graphing large data sets with data points that repeat. The data usually goes on y-axis with the frequency being graphed on the x-axis. Time series graphs can be helpful when looking at large amounts of data for one variable over a period of time.

2.2 Measures of the Location of the Data

The values that divide a rank-ordered set of data into 100 equal parts are called percentiles. Percentiles are used to compare and interpret data. For example, an observation at the 50th percentile would be greater than 50 percent of the other observations in the set. Quartiles divide data into quarters. The first quartile (Q_1) is the 25th percentile, the second quartile (Q_2 or median) is 50th percentile, and the third quartile (Q_3) is the 75th percentile. The interquartile range, or IQR , is the range of the middle 50 percent of the data values. The IQR is found by subtracting Q_1 from Q_3 , and can help determine outliers by using the following two expressions.

- $Q_3 + IQR(1.5)$
- $Q_1 - IQR(1.5)$

2.3 Measures of the Center of the Data

The **mean** and the **median** can be calculated to help you find the "center" of a data set. The mean is the most commonly used estimate of central tendency, but the median is the best measurement when a data set contains several outliers or extreme values. The **mode** will tell you the most frequently occurring datum (or data) in your data set.

The mean, median, and mode are extremely helpful when you need to analyze your data, but if your data set consists of ranges which lack specific values, the mean may seem impossible to calculate. However, the mean can be approximated if you add the lower boundary with the upper boundary and divide by two to find the midpoint of each interval. Multiply each midpoint by the number of values found in the corresponding range. Divide the sum of these values by the total number of data values in the set.

2.6 Skewness and the Mean, Median, and Mode

Looking at the distribution of data can reveal a lot about the relationship between the mean, the median, and the mode. In a **right (or positively) skewed** distribution, the mean will be higher than the median. In a **left (or negatively) skewed** distribution, the mean will be lower than the median. In a **normally shaped distribution**, the mean and median will be (approximately) equal.

2.7 Measures of the Spread of the Data

The standard deviation can be used to represent the spread of data, and it is a measure of how far, on average, all of the individual scores are from the mean.

There are different equations to use if we are calculating the standard deviation of a sample or of a population.

- $s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}}$ or $s = \sqrt{\frac{\Sigma f_i(x_i - \bar{x})^2}{n-1}}$ is the formula for calculating the standard deviation of a sample. To calculate the standard deviation of a population, we would use the population mean, μ , and the formula is $\sigma = \sqrt{\frac{\Sigma(x_i - \mu)^2}{N}}$ or $\sigma = \sqrt{\frac{\Sigma f_i(x_i - \mu)^2}{N}}$

This page titled [2.9: Chapter 2 Review](#) is shared under a [CC BY](#) license and was authored, remixed, and/or curated by .