Descriptive Statistics

Sreeja Bolla

May 30, 2025

1 Measures of Central Tendency

Arithmetic Mean: Also called the average or mean, the arithmetic mean is calculated by summing all the values in a dataset and dividing by the number of values. For n values $\{x_1, x_2, ..., x_n\}$, this can be written $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Mode: The mode is the most frequently occurring value in the dataset.

Median: This is the middle value in a dataset. For a set with n values, if n is odd, the median can be found by ordering the set and taking the $\frac{n+1}{2}th$ element. If n is even, the median can be found by taking the average of the $\frac{n}{2}th$ and $\frac{n}{2}+1th$ element.

Example: A group of ten athletes measured their verticals by jumping and recording the result in inches. The final measurements were $\{24, 24, 18, 23, 29, 27, 21, 26, 25, 24\}$. Find the mean, median, and mode of their verticals.

Solution: The sum of all the heights is 24 + 24 + 18 + 23 + 29 + 27 + 21 + 26 + 25 + 24 = 241. Since there are ten values, the mean is $\frac{241}{10} = 24.1$ inches. To find the median, we must order the elements to get $\{18, 21, 23, 24, 24, 24, 25, 26, 27, 29\}$. The median is the average of 24 and 24, which means the median vertical is 24 inches. The mode is also 24 inches, since this value repeats three times, more than any other height.

2 Measures of Spread

Range: The range is the length of values that the dataset covers, which is calculated by taking the difference between the largest and smallest values in the set.

Variance: This measures the spread of the data around its mean, calculated by summing the squared distance between each value and the mean and dividing by the number of data values. For a set with n values $\{x_1, x_2, ..., x_n\}$, $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$. The variance is also the standard deviation squared.

Sample Variance: For a sample taken from the data, this measures the spread of the sample data around its mean, calculated by summing the squared distance between each sample value and the sample mean and dividing by the number of sample data values. For a for a sample of size n $\{x_1, x_2, ..., x_n\}$, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$. The variance is also the standard deviation squared.

Standard Deviation: The standard deviation measures by how much the data deviates from the mean, calculated by taking the square root of the variance formula. Thus, for a set of n values $\{x_1, x_2, ..., x_n\}$, $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$.

Sample Standard Deviation: The sample standard deviation measures by how much the sample data deviates from its mean, calculated by taking the square root of the sample variance formula. Thus, for a sample of size n $\{x_1, x_2, ..., x_n\}$, $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$.

Interquartile Range: The IQR of a set of data is the range of the middle half of the data. This can be calculated by taking the difference between the median for the second half of the data and the median for the first half.

Standard Error of the Mean: For the set of all means for samples taken from a data set, the standard error, SE, is the spread of the sample means from the arithmetic mean for the entire dataset. This is calculated by SE = $\frac{\sigma}{\sqrt{n}}$

Example: Every month, every Girl Scout group in the state reports their cookie sales to be analyzed. There are twelve groups in MA and they reported the following numbers, {10, 12, 13, 15, 25, 30, 40, 44, 48, 49, 56, 57}. Find the range, variance, standard deviation, and IQR for this

set.

Solution: The largest value is 57 and the smallest is 10, so the range is 57 - 10 = 47. The sum of all the values in the set is 10 + 12 + 13 + 15 + 25 + 30 + 40 + 44 + 48 + 49 + 56 + 57 = 499, so the mean is $\frac{400}{12}$ = 33.25 cookies. The sum of the squared difference between each value and the mean is 540.5625 + 451.5625 + 410.0625 + 333.0625 + 68.0625 + 10.5625 + 45.5625 + 115.5625 + 217.5625 + 248.0625 + 516.5625 + 564.0625 = 3521.25. Thus, the variance is $\frac{3521.25}{12}$ = 293.4375 cookies. The standard deviation is the square root of the variance, so $\sigma = \sqrt{293.4375}$, which is about 17.13 cookies. The median of the first half of values is the average of 13 and 15, or 14. The median of the second half is the average of 48 and 49, or 48.5. Thus, the IQR = 48.5 - 14 = 34.5.