Vocabulary: <u>Unit 1: Inferences and Conclusions from Data</u>

- **Center:**
- **Central Limit Theorem:**
- **Confidence Interval**
- **Empirical Rule**:
    - 68%
    - 95%
    - 99.7%
- **Margin of Error**:
- **Mean absolute deviation:**
- **Parameters:**
- **Population:**
- **Random:**
- **Sample:**
- **Sample Mean: Sample Proportion:**
- **Sampling Distribution:**
- **Sampling Variability**:
- **Shape:**
    - ***Symmetry***-
    - ***Number of Peaks***-
    - ***Direction of Skew***-.
    - ***Uniformity***-
- **Spread:**
- **Standard Deviation**:
- **Statistics**:
- **Variance:**

Vocabulary: Unit 1: Inferences and Conclusions from Data

- **Center:** Measures of center refer to the summary measures used to describe the most "typical" value in a set of data. The two most common measures of center are median and the mean.
- **Central Limit Theorem:** Choose a simple random sample of size n from any population with mean μ and standard deviation σ. When n is $\bar{x}$ ge (at least 30), the sampling distribution of the sample mean $\frac{\sigma}{\sqrt{n}}$ approximately normal with mean μ and standard deviation . Choose a simple random sample of size *n* from a large population with

  population parameter *p* having some characteristic of interest. Then the sampling distribution of the sample proportion $\hat{p}$ is approximately normal with mean *p* and standard deviation $\sqrt{\frac{p(1-p)}{n}}$. This approximation becomes more and more accurate as the sample size *n* increases, and it is generally considered valid if the population is much larger than the sample, i.e. $np \geq 10$ and $n(1-p) \geq 10$. The CLT allows us to use normal calculations to determine probabilities about sample proportions and sample means obtained from populations that are not normally distributed.
- **Confidence Interval** is an interval for a parameter, calculated from the data, usually in the form estimate ± margin of error. The confidence level gives the probability that the interval will capture the true parameter value in repeated samples.
- **Empirical Rule**: If a distribution is normal, then approximately
  - 68% of the data will be located within one standard deviation symmetric to the mean
  - 95% of the data will be located within two standard deviations symmetric to the mean
  - 99.7% of the data will be located within three standard deviations symmetric to the mean
- **Margin of Error**: The value in the confidence interval that says how accurate we believe our estimate of the parameter to be. The margin of error is comprised of the product of the z-score and the standard deviation (or standard error of the estimate). The margin of error can be decreased by increasing the sample size or decreasing the confidence level.
- **Mean absolute deviation:** A measure of variation in a set of numerical data, computed by adding the distances between each data value and the mean, then dividing by the number of data values. Example: For the data set {2, 3, 6, 7, 10, 12, 14, 15, 22, 120}, the mean absolute deviation is 20.
- **Parameters:** These are numerical values that describe the population. The population mean is symbolically represented by the parameter μ_X. The population standard deviation is symbolically represented by the parameter σ_X
- **Population:** The entire set of items from which data can be selected.
- **Random:** Events are random when individual outcomes are uncertain. However, there is a regular distribution of outcomes in a large number of repetitions.
- **Sample:** A subset, or portion, of the population.
- **Sample Mean:** A statistic measuring the average of the observations in the sample. It is written as $\bar{x}$. The mean of the population, a parameter, is written as μ.
- **Sample Proportion:** A statistic indicating the proportion of successes in a particular sample. It is written as $\hat{p}$. The population proportion, a parameter, is written as *p*.
- **Sampling Distribution:** A statistics is the distribution of values taken by the statistic in all possible samples of the same size from the same population.
- **Sampling Variability**: The fact that the value of a statistic varies in repeated random sampling.
- **Shape:** The shape of a distribution is described by symmetry, number of peaks, direction of skew, or uniformity.
  - ***Symmetry***- A symmetric distribution can be divided at the center so that each half is a mirror image of the other.
  - ***Number of Peaks***- Distributions can have few or many peaks. Distributions with one clear peak are called unimodal and distributions with two clear peaks are called bimodal. Unimodal distributions are sometimes called bell-shaped.
  - ***Direction of Skew***- Some distributions have many more observations on one side of graph than the other. Distributions with a tail on the right toward the higher values are said to be skewed right; and distributions with a tail on the left toward the lower values are said to be skewed left.
  - ***Uniformity***- When observations in a set of data are equally spread across the range of the distribution, the distribution is called uniform distribution. A uniform distribution has no clear peaks.
- **Spread:** The spread of a distribution refers to the variability of the data. If the data cluster around a single central value, the spread is smaller. The further the observations fall from the center, the greater the spread or variability of the set. (range, interquartile range, Mean Absolute Deviation, and Standard Deviation measure the spread of data)
- **Standard Deviation**: The square root of the variance. $\sigma = \sqrt{\frac{1}{n}\sum(x_i - x)^2}$
- **Statistics**: These are numerical values that describe the sample. The sample mean is symbolically represented by the statistic $\bar{x}$. The sample standard deviation is symbolically represented by the statistic s_x
- **Variance:** The average of the squares of the deviations of the observations from their mean. $\sigma^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$