***Statistics Notes - 1***

Statistics: is the science of collecting, organizing and analyzing the data.

Data: Facts and pieces of information.

Descriptive of statistics: consists of organizing and summarizing data. Asking to describe data.

Inferential Statistics: consists of using data you have measured to form conclusion. Asking additional question about data.

***Sampling Method***

Population: the group we are interested in study.

Sample: a subset of population. Sample comes from population.

“**N**” – used to represent the size of population.

“**n**” – used to represent the size of sample.

***Sampling Methods***:

1. Simple Random Sampling: every members of population (N) has an equal chance of being selected for sample (n).
2. Stratified Sampling: the population (N) is split into non-overlapping group (“strata”), then sampling random sampling is done on each group to form a sample (n).

Ex: splitting a population of students into men and women, then sampling from each of the two group.

1. Systematic Sampling: every *n*th individual from population (N) is placed in the sample.

EX: if we add every 7th individual to walk out of a supermarket to your sample, you are performing systematic sampling.

1. Convenience Sampling: easily obtained individual from population (N) are placed in the sampling (n) this of sampling also called “voluntary response sampling”, because individual often select to be a part of the sampling.

***Types Of Variable***

Variable: is a property that can take on many values. A single variable is always spoken of in the singular. “Age” is a variable, but “Ages” is not variable.

Quantitative Variable : are measured numerically. With measurements of quantitative variable we can do thing like add, subtraction, multiple and divide and get meaningful result. In the previous example, “Age” was quantitative variable.

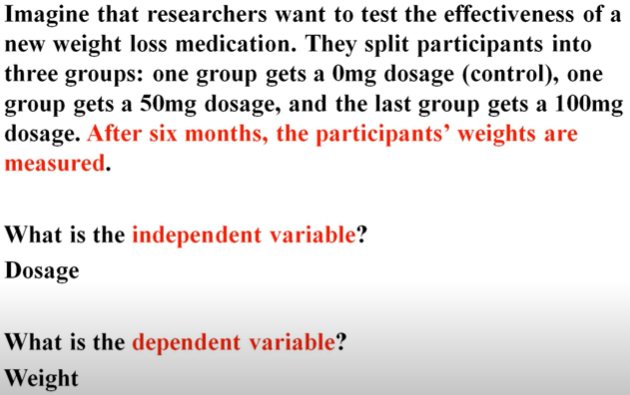
1. Discrete Variable: is a quantitative variable with finite number of value. Ex: if we rolled a six-sided die four times and measured how many times you rolled out an even number then out come will be {0, 1, 2, 3, 4}
2. Continuous Variable: is a quantitative variable with an infinite number of values. Ex: Temperature can take on an infinite number of values, such as 80 degree, or 80.01 degrees, or 80.005092359 degrees.

Qualitative / Categorical Variable: allow for classification based on some characteristics. In the previous example, “Gender” was a qualitative / categorical variable. Gender was categorized as either male or female.

***Independent and Dependent Variables***

Independent variable: is any variable that is being manipulated

Dependent Variable: is any variable that is being measured



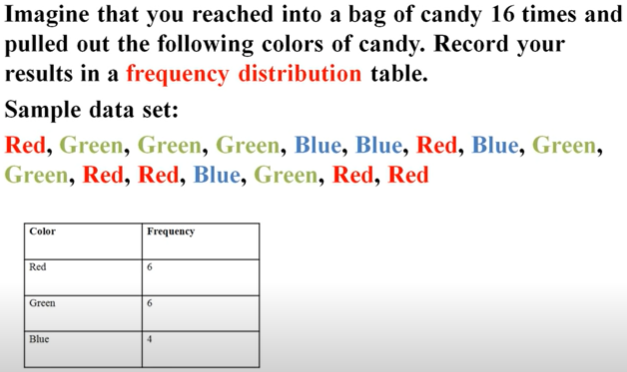
***Variable Measurement Scales***

***Four types of Measured Variable***

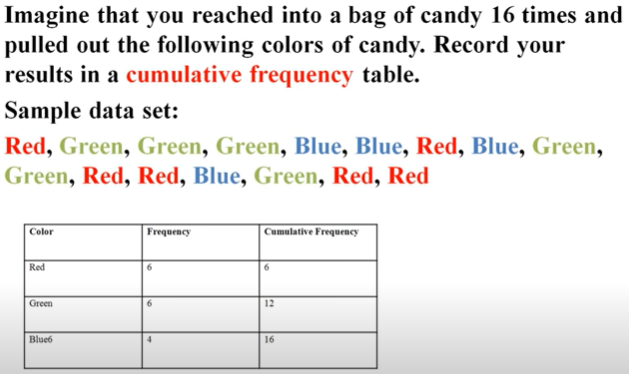
1. Nominal: is also known as qualitative /categorical data. Data that is split into categories. Ex: if we would like to collect data for variable “color” then we will get information such as “Red”, “green”, “blue”, and so on. This Qualitative information is known as Nominal data.
2. Ordinary: is data where order matters, but distance between value does not. Ex: three people in the race. One finished in 1st place, one in second place and one in 3rd place. This order can be placed in order, but we can’t necessarily measure the distance between values (may be 1st place finished four second ahead of 2nd place, and 2nd place finished nineteen second ahead of 3rd place).
3. Interval: data where order matters, and distances between values are equal and meaningful, and natural zero is not present. Ex: temperature is interval data. Then difference between 10 degree and 20 degree is 10 degree. The difference between 80 degree and 90 degree is 10 degree s. the scale at any given point is constant , and it is not possible to have measurement of 0 degree.
4. Ratio: data where order matters, and distances between values are equal and meaningful, and natural zero is present. Ex: weight is ratio data. The difference between 140 pounds and 155 degree is 15 pound. The difference between 280 pounds and 295 pounds is 15 pounds. The scale at any given point is constant, and it is possible to have a measurement of 0 pound.

***Frequency Distributions and Cumulative Frequency Table***

Frequency Distribution: lists each measured category and the number of occurrences for each category.

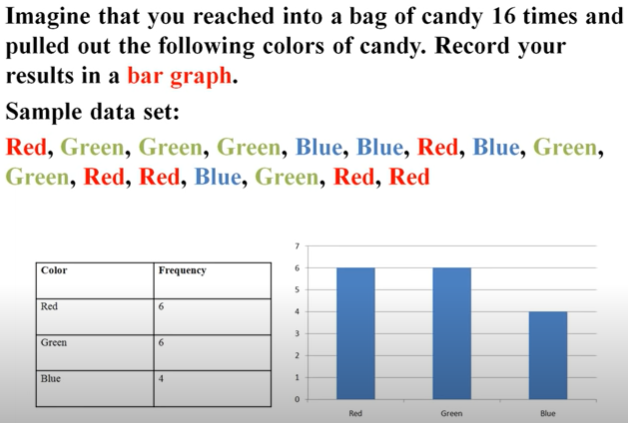


Cumulative Frequency Distribution: is the sum of the class and all classes below it in a frequency distribution. All that means is you’re adding up a value and all of the values that came before it

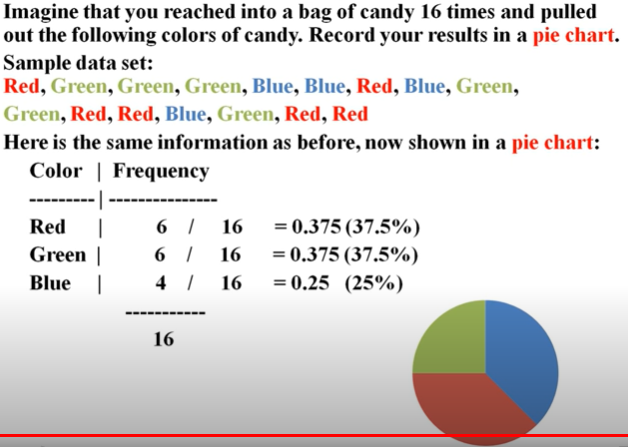


***Bar Graphs and Pie Charts***

Bar Graphs: list each measured category on the horizontal axis and the number of occurrences for each category on the vertical axis.

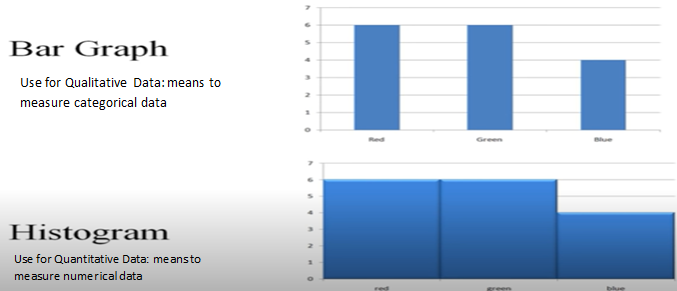


Pie Chart: is a circle divided into sectors, where each sector represents a category of data that is proportional to the total amount of data collected.

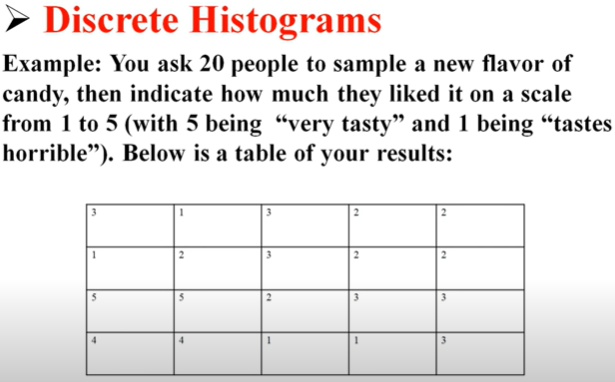
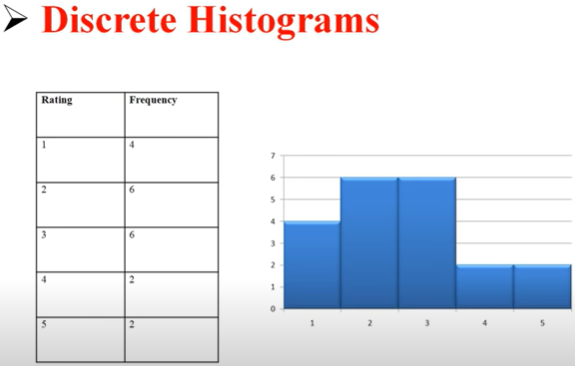


***Histograms (Discrete and Continuous) and Stem and Leaf Plots***

Histogram: is a bar graph that lists each measured category on the horizontal axis and the number of occurrences of each category on the vertical axis. The rectangles for each bar touch one another.



Discrete Histogram: are created when dealing with discrete values on the horizontal axis, while

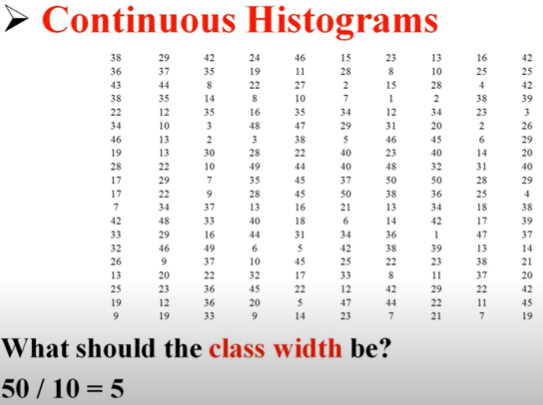
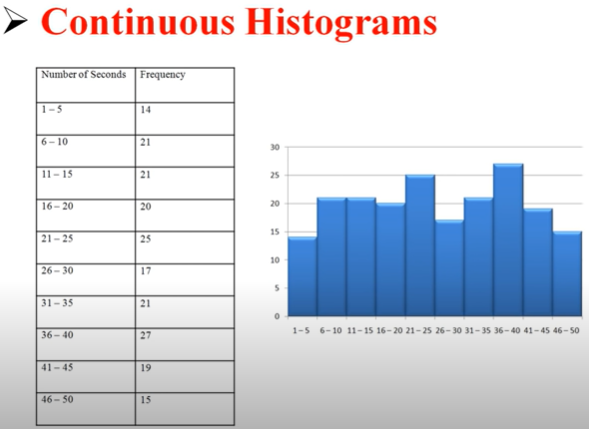
 

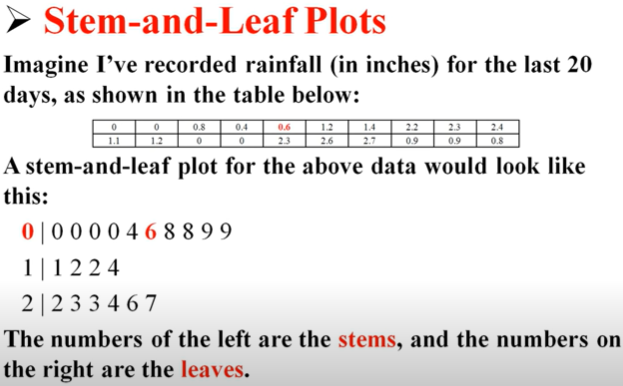
Continuous histograms: are created when dealing with continuous values on the horizontal axis.

There are no distinct categories, so to organize data on the horizontal axis we will need to create classes.

Class: is an interval of many values. For Ex, 1-10 is a class that covers all numbers from 1 to 10.

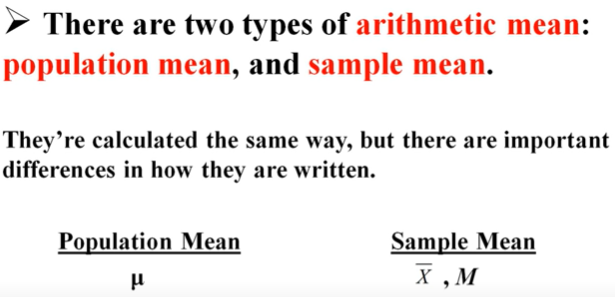
1. Lower Class Limit: smallest value within each class.
2. Upper Class Limit: largest value within each class.
3. Class Width: difference between consecutive lower class limit.

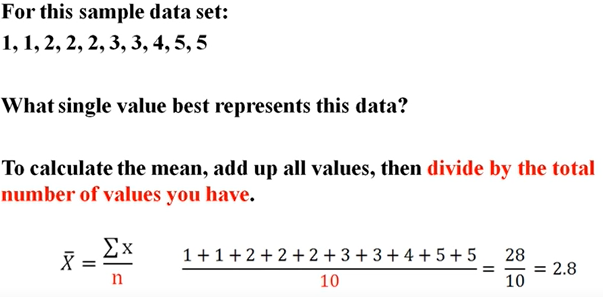
 



***Arithmetic Mean for Samples and Populations***

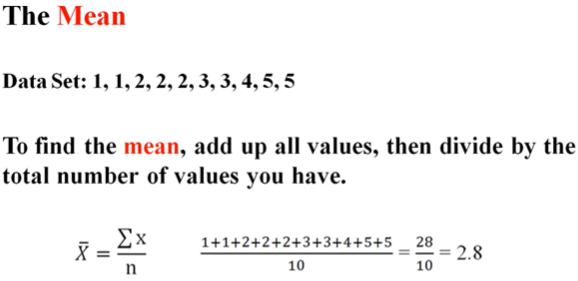
Arithmetic Mean: is a single value meant to “sum up” a data set.

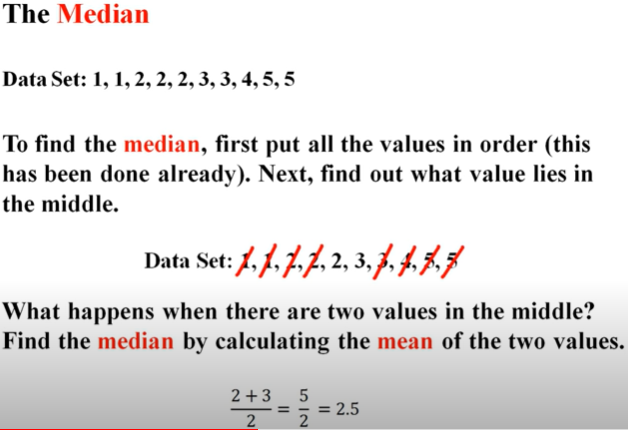
 

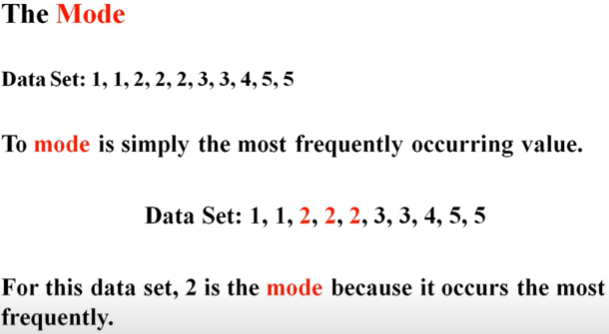


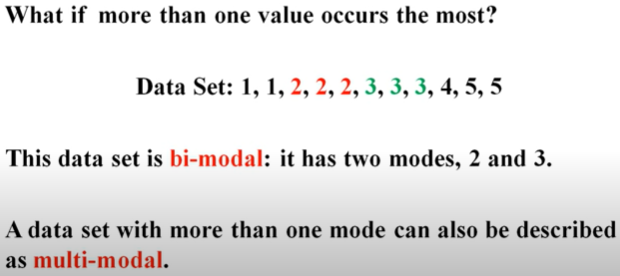
***Central Tendency: Mean, Median, and Mode***

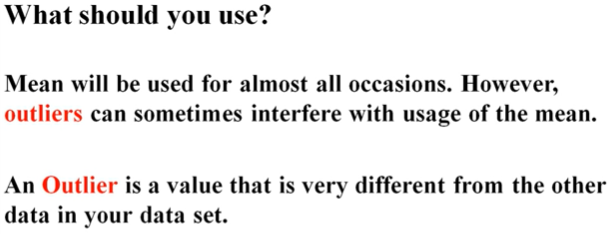
Central Tendency refers to the measure used to determine the “center” of a distribution of data. It is used to find a single score that is most representative of an entire data set.

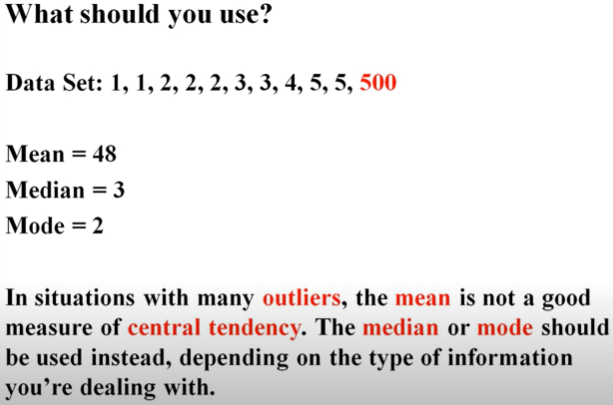












***Measured of Dispersion: Variance and Standard Deviation of Population***

