

Statistics

Statistics- the branch of mathematics that deals with the collection, organization, analysis, and interpretation of numerical data.

Dot Plot- a statistical chart consisting of data points plotted on a fairly simple scale.

Histogram- a diagram consisting of rectangles whose area is proportional to the frequency of a variable and whose width is equal to the class interval.

Box Plot- a graphic way to display the median, quartiles, maximum and minimum values of a data set on a number line.

Interquartile Range- the range of values of a frequency distribution between the first and third quartiles.

Standard Deviation- a measure of how spread out numbers are.

Outlier- A value that "lies outside" (is much smaller or larger than) most of the other values in a set of data. $y < Q1 - 1.5 \times IQR$ or $y > Q3 + 1.5 \times IQR$

Frequency- the rate at which something occurs or is repeated over a particular period of time or in a given sample.

Frequency Table- a table that shows the total for each category or group of data.

Relative Frequency- how often something happens divided by all outcomes.

Joint Frequency- It is called joint frequency because you are joining one variable from the row and one variable from the column.

Joint Relative Frequency- the ratio of the number of observations of a joint frequency to the total number of observations in a frequency table.

Marginal Frequency- the total in a row or column in a two way table.

Marginal Relative Frequency- the ratio of the sum of the joint relative frequency in a row or column and the total number of data values.

Conditional Frequency- an entry in a relative frequency table. It is the ratio of a joint frequency to the total number of observations.

Conditional Relative Frequency- the ratio of a joint frequency to a marginal frequency in its row or column.

Residuals- the distance from the predicted location of a point on a line to an actual data point.

Correlation- mathematical relationship between two variables.

Causation- the capacity of one variable to influence another.

Correlation Coefficient- The quantity r , called the linear correlation coefficient, measures the strength and the direction of a linear relationship between two variables. The value of r is such that $-1 < r < 1$.

Spread- the name given in statistics to describe how the data lies. It is measured in a variety of ways such as the range, the interquartile range and the standard deviation.

Shape- a way to classify data sets into 3 categories that describe the shape of the data distribution: symmetric, left skewed, right skewed.

Mean- The mean is the average of the numbers: a calculated "central" value of a set of numbers.

Median- The middle number in a sorted list of numbers.

Mode- The number which appears most often in a set of numbers.

Range- The difference between the lowest and highest values.

Gap- refer to areas of a graphic display where there are no observations.

Cluster- refers to a type of sampling method.

Two-Way Frequency Table- involves listing all of the values for two categorical variables. All of the values for one of the variables is listed in a vertical column. The values for the other variable are listed along a horizontal row.

Bivariate Data- a study that examines the relationship between two variables

Univariate Data- a study that looks at only one variable

Categorical Data- represent characteristics such as a person's gender, marital status, hometown, or the types of movies they like. Categorical data can take on numerical values (such as "1" indicating male and "2" indicating female), but those numbers don't have mathematical meaning.

First Quartile- (in a frequency distribution) the smallest quartile; the twenty-fifth percentile; the value of the variable below which one quarter of the elements are located.

Third Quartile- (in a frequency distribution) the largest quartile; the 75th percentile; the value of the variable below which three quarters of the elements are located.

Quantitative Data- data expressing a certain quantity, amount or range.

Scatter Plot- is a useful summary of a set of bivariate data (two variables), usually drawn before working out a linear correlation coefficient or fitting a regression line. It gives a good visual picture of the relationship between the two variables, and aids the interpretation of the correlation coefficient or regression model.

Line of Best Fit (*Trend line or Linear regression*)- A line on a graph showing the general direction that a group of points seem to be heading.

Interpolation- a value inside the set of data points.

Extrapolation- a value outside the set of data points

Bias-unfair

Unbiased- fair