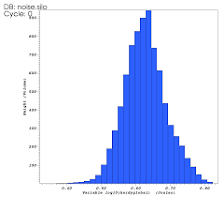
What is a histogram?

A histogram is a chart that plots the distribution of a numeric variable's values as a series of bars. Each bar typically covers a range of numeric values called a bin or class; a bar's height indicates the frequency of data points with a value within the corresponding bin.

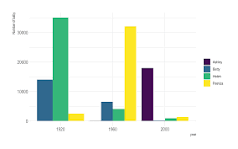


Bar Plot

# Bar Plot -- Mainly bar plot is used to show the relationship between the numeric and categorical values. In a bar chart, we

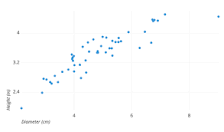
## have one axis representing a particular category of the columns and another axis representing the values or count of the

## specific category. Bar charts are plotted both vertically and horizontally and are plotted using the following line of code:



What is a scatter plot?

A scatter plot (aka scatter chart, scatter graph) uses dots to represent values for two different numeric variables. The position of each dot on the horizontal and vertical axis indicates values for an individual data point. Scatter plots are used to observe relationships between variables.

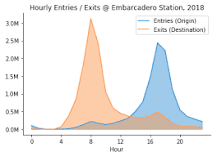


What is the definition of a pie chart?

A pie chart is a type of graph that represents the data in the circular graph. The slices of pie show the relative size of the data, and it is a type of pictorial representation of data. A pie chart requires a list of categorical variables and numerical variables.

What is the area chart?

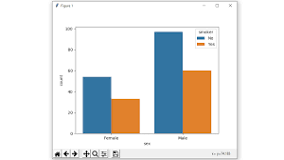
An area chart combines the line chart and bar chart to show how one or more groups' numeric values change over the progression of a second variable, typically that of time. An area chart is distinguished from a line chart by the addition of shading between lines and a baseline, like in a bar chart.



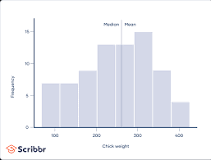
What is the count plot?

countplot() method is used to Show the counts of observations in each categorical bin using bars.

A counts plot is a variant of the strip plot with a better view of overlapping data points, used to visualise the distribution of many individual one-dimensional values.

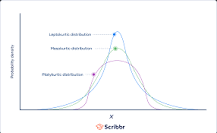


What is your definition of skewness?



Skewness is a measure of the asymmetry of a distribution. A distribution is asymmetrical when its left and right side are not mirror images. A distribution can have right (or positive), left (or negative), or zero skewness.

What do you mean by kurtosis?

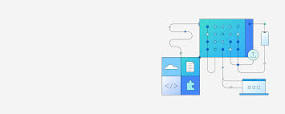


Kurtosis is a measure of the tailedness of a distribution. Tailedness is how often outliers occur. Excess kurtosis is the tailedness of a distribution relative to a normal distribution. Distributions with medium kurtosis (medium tails) are mesokurtic. Distributions with low kurtosis (thin tails) are platykurtic.

Formula

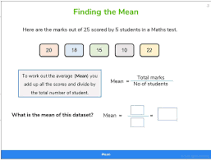
|  |  |  |
| --- | --- | --- |
|  | = | kurtosis |
|  | = | [fourth central moment](https://www.google.com/search?sca_esv=3c6329b6486d7f09&sca_upv=1&sxsrf=ADLYWIL6zzcNVuGOw5alRldrRM6nUDvtcg:1717911663625&q=Central+moment&stick=H4sIAAAAAAAAAOPgE-LQz9U3MLG0TFMCs0wNcs21tLKTrfRTU0qTE0sy8_P00_KLcktzEq2gtEJmbmJ6qkJiXnF5atEjRmNugZc_7glLaU1ac_IaowoXV3BGfrlrXklmSaWQGBcblMUjxcUFt4BnESufc2peSVFijkJufi6QBQDgHK5ViQAAAA&sa=X&ved=2ahUKEwiD09zT582GAxVzxTgGHZncEokQ24YFegQIHBAC) |
|  | = | [standard deviation](https://www.google.com/search?sca_esv=3c6329b6486d7f09&sca_upv=1&sxsrf=ADLYWIL6zzcNVuGOw5alRldrRM6nUDvtcg:1717911663625&q=Standard+deviation&stick=H4sIAAAAAAAAAOPgE-LQz9U3MLG0TFMCs8zKjYy0tLKTrfRTU0qTE0sy8_P00_KLcktzEq2gtEJmbmJ6qkJiXnF5atEjRmNugZc_7glLaU1ac_IaowoXV3BGfrlrXklmSaWQGBcblMUjxcUFt4BnEatQcEliXkpiUYpCSmpZJtgeAJOzqdqNAAAA&sa=X&ved=2ahUKEwiD09zT582GAxVzxTgGHZncEokQ24YFegQIHBAD) |

What is machine learning short answers?



Machine learning (ML) is a branch of artificial intelligence (AI) and computer science that focuses on the using data and algorithms to enable AI to imitate the way that humans learn, gradually improving its accuracy.

What is the mean in math definition?



A mean in math is the average of a data set, found by adding all numbers together and then dividing the sum of the numbers by the number of numbers.

Median: The middle number; found by ordering all data points and picking out the one in the middle

What is a mode in statistic?

Definition. The mode is the most common number that appears in your set of data. To find the mode count how often each number appears and the number that appears the most

times is the mode.

What is the Median?

A median is the middle number in a sorted list of numbers (either ascending or descending) used in statistical studies. To determine the median value in a sequence of numbers, the numbers must first be sorted or arranged in value order from lowest to highest or highest to lowest.

What is the Mean?

Mean is the average of the given numbers and is calculated by dividing the sum of given numbers by the total number of numbers. Mean = (Sum of all the observations/Total number of observations)

. What is Variance?

The term variance refers to a statistical measurement of the spread between numbers in a data set. More specifically, variance measures how far each number in the set is from the mean (average), and thus from every other number in the set.

What is Standard Deviation?

A standard deviation (or σ) is a measure of how dispersed the data is in relation to the mean. Low, or small, standard deviation indicates data are clustered tightly around the mean, and high, or large, standard deviation indicates data are more spread out.

What is a Confidence Interval?

What Is Confidence Interval? A confidence interval shows the probability that a parameter will fall between a pair of values around the mean. Confidence intervals show the degree of uncertainty or certainty in a sampling method. They are constructed using confidence levels of 95% or 99%

A confidence interval is a range of values that describes the uncertainty surrounding an estimate.

What is Hypothesis Testing?

Hypothesis testing uses sample data to evaluate a hypothesis about a population. A hypothesis test assesses how unusual the result is, whether it is reasonable chance variation or whether the result is too extreme to be considered chance variation.

What is the definition of covariance in statistics?

Covariance is a measure of the relationship between two random variables and to what extent, they change together. Or we can say, in other words, it defines the changes between the two variables, such that change in one variable is equal to change in another variable.

What is Correlation in statistics?

Correlation is a statistical measure that expresses the extent to which two variables are linearly related (meaning they change together at a constant rate). It's a common tool for describing simple relationships without making a statement about cause and effect.

What is the Coefficient of Variation in statistics?

The coefficient of variation (CV) is the ratio of the standard deviation to the mean. The higher the coefficient of variation, the greater the level of dispersion around the mean. It is generally expressed as a percentage.

How is the Mean Different from the Median?

The coefficient of variation (CV) is the ratio of the standard deviation to the mean. The higher the coefficient of variation, the greater the level of dispersion around the mean. It is generally expressed as a percentage.

What is the Purpose of a Histogram in statistics?

The histogram is a popular graphing tool. It is used to summarize discrete or continuous data that are measured on an interval scale. It is often used to illustrate the major features of the distribution of the data in a convenient form.

How Does a Scatter Plot Help in Understanding Data?

A scatter plot identifies a possible relationship between changes observed in two different sets of variables. It provides a visual and statistical means to test the strength of a relationship between two variables

Why is Standard Deviation Preferred Over Variance?

A scatter plot identifies a possible relationship between changes observed in two different sets of variables. It provides a visual and statistical means to test the strength of a relationship between two variables.

What is the Interquartile Range (IQR)?

Interquartile range is defined as the difference between the upper and lower quartile values in a set of data. It is commonly referred to as IQR and is used as a measure of spread and variability in a data set. This topic is often discussed in statistics with similar topics such as mean deviation and distribution.

What is a P-Value in Hypothesis Testing?

The P value is defined as the probability under the assumption of no effect or no difference (null hypothesis), of obtaining a result equal to or more extreme than what was actually observed. The P stands for probability and measures how likely it is that any observed difference between groups is due to chance.

What is a Null Hypothesis?

A null hypothesis is a type of statistical hypothesis that proposes that no statistical significance exists in a set of given observations. Hypothesis testing is used to assess the credibility of a hypothesis by using sample data

What is the Central Limit Theorem in statistics?

The central limit theorem (CLT) states that the distribution of sample means approximates a normal distribution as the sample size gets larger, regardless of the population's distribution

How is Correlation Different from Causation in statistics?

A correlation between variables, however, does not automatically mean that the change in one variable is the cause of the change in the values of the other variable. Causation indicates that one event is the result of the occurrence of the other event; i.e. there is a causal relationship between the two events.