STATISTICS WORKSHEET-1

Q 1 Bernoulli random variables take (only) the values 1 and 0

   a) True

Q 2 Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?

   a) Central Limit Theorem

 Q 3 Which of the following is incorrect with respect to use of Poisson distribution?

b) Modeling contingency tables

Q 4 Point out the correct statement.

c) Sums of normally distributed random variables are again normally distributed even if the variables are dependent

Q 5 _____ random variables are used to model rates

Q 6   10. Usually replacing the standard error by its estimated value does change the CLT.

a) True

Q 7 Which of the following testing is concerned with making decisions using data?

b) Hypothesis

Q 8 4. Normalized data are centered at_and have units equal to standard deviations of the original data.

a) 0

Q 9 Which of the following statement is incorrect with respect to outliers

b) Outliers cannot conform to the regression relationship

Q 10 What do you understand by the term Normal Distribution?

The normal distribution, also known as the Gaussian distribution, is the most important probability distribution in statistics for independent, random variables. Most people recognize its familiar bell-shaped curve in statistical reports.

The normal distribution is a continuous probability distribution that is symmetrical around its mean, most of the observations cluster around the central peak, and the probabilities for values further away from the mean taper off equally in both directions. Extreme values in both tails of the distribution are similarly unlikely. While the normal distribution is symmetrical, not all symmetrical distributions are normal. For example, the Student's t, Cauchy, and logistic distributions are symmetric.

As with any probability distribution, the normal distribution describes how the values of a variable

are distributed. It is the most important probability distribution in statistics because it accurately describes the distribution of values for many natural phenomena. Characteristics that are the sum of many independent processes frequently follow normal distributions. For example, heights, blood pressure, measurement error, and IQ scores follow the normal distribution.

The mean is the central tendency of the normal distribution. It defines the location of the peak for the bell curve.

Q11 What is A/B testing?

A/B testing is one of the most popular controlled experiments used to optimize web marketing strategies. It allows decision makers to choose the best design for a website by looking at the analytics

results obtained with two possible alternatives A and B.

In this article we'll see how different statistical methods can be used to make A/B testing successful. I recommend you to also have a look at this notebook where you can play with the examples discussed in this article.

To understand what A/B testing is about, let's consider two alternative designs: A and B. Visitors of a website are randomly served with one of the two. Then, data about their activity is collected by web analytics. Given this data, one can apply statistical tests to determine whether one of the two designs has better efficacy.

Now, different kinds of metrics can be used to measure a website efficacy. With **discrete metrics**, also called **binomial metrics**, only the two values **0** and **1** are possible. The following are examples of popular discrete metrics.

- Click-through rate — if a user is shown an advertisement, do they click on it?

- Conversion rate — if a user is shown an advertisement, do they convert into customers?
- Bounce rate — if a user is visits a website, is the following visited page on the same website?

Q 12 4. What is linear regression in statistics?

Linear regression models the relationships between at least one explanatory variable and an outcome variable. This flexible analysis allows you to separate the effects of complicated research questions, allowing you to isolate each variable's role. Additionally, linear models can fit curvature and interaction effects.

Statisticians refer to the explanatory variables in linear regression as independent variables (IV) and the outcome as dependent variables (DV). When a linear model has one IV, the procedure is known as simple linear regression. When there are more than one IV, statisticians refer to it as multiple

regression. These models assume that the average value of the dependent variable depends on a linear function of the independent variables.

Linear regression has two primary purposes—understanding the relationships between variables and prediction.

1. The coefficients represent the estimated magnitude and direction (positive/negative) of the relationship between each independent variable and the dependent variable.
2. The equation allows you to predict the mean value of the dependent variable given the values of the independent variables that you specify.

Linear regression finds the constant and coefficient values for the IVs for a line that best fit your sample data. The graph below shows the best linear fit for the height and weight data points, revealing the mathematical relationship between them. Additionally, you can use the line's

equation to predict future values of the weight given a person's height.

Q 13 What are the various branches of statistics

Statistics is the main branch of [mathematics](). Used to perform different operations, i.e., Data collection, organization, analysis, and so on. In other words, statistics is a form of mathematical analysis that uses quantitative models to give a set of experimental data or studies of real life. Statistics examine the methodology for collecting, reviewing, analyzing, and making data conclusions. Some statistical measures include the following:

- **Mean:** It is an important concept in mathematics and statistics. The mean is an

average and the most common value in the collection of numbers.

- **Regression analysis:** It is a powerful statistical method. It allows us to examine the relationship between two or more variables of interest.

- **Skewness:** In statistics, skewness is a degree of asymmetry that is observed in a probability distribution. Distributions can display right (positive) skewness or left (negative) skewness to differing degrees. A normal distribution (bell curve) presents zero skewness.

- **Kurtosis:** It is a measure of the combined weight of a distribution's tails relative to the center of the distribution.

- **Variance:** It estimates the variability from the mean or average.

- **Analysis of variance:** The method of statistics that separates the variance data into several components used for additional tests.