

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

a) **True** ✓

b) False

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** ✓

b) Central Mean Theorem

c) Centroid Limit Theorem

d) All of the mentioned

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

a) Modeling event/time data

b) **Modeling bounded count data** ✓

c) Modeling contingency tables

d) All of the mentioned

4. Point out the correct statement.

a) The exponent of a normally distributed random variables follows what is called the log-normal distribution

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

c) The square of a standard normal random variable follows what is called chi-squared distribution

d) **All of the mentioned** ✓

5. _____ random variables are used to model rates.

a) Empirical

b) Binomial

c) **Poisson** ✓

d) All of the mentioned

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

a) True

b) **False** ✓

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

a) Probability

b) **Hypothesis** ✓

c) Causal

d) None of the mentioned

8. 4. Normalized data are centered at _____ and have units equal to standard deviations of the original data.

- a) 0 ✓
- b) 5
- c) 1
- d) 10

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

- a) Outliers can have varying degrees of influence
- b) Outliers can be the result of spurious or real processes
- c) **Outliers cannot conform to the regression relationship** ✓
- d) None of the mentioned

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

- **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. 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 fits many natural phenomena. For example, heights, blood pressure, measurement error, and IQ scores follow the normal distribution. It is also known as the Gaussian distribution and the bell curve.

How do you handle missing data? What imputation techniques do you recommend?

➤ Following are the ways to handle missing data :

1. Deleting Rows with missing values
2. Impute missing values for continuous variable
3. Impute missing values for categorical variable
4. Other Imputation Methods
5. Using Algorithms that support missing values
6. Prediction of missing values
7. Imputation using Deep Learning Library — Datawig

➤ Following are the imputation technique we can use

1. Imputation Using (Most Frequent) or (Zero/Constant) Values
2. Imputation Using (Mean/Median) Values
3. Imputation Using (Most Frequent) or (Zero/Constant) Values
4. Imputation Using k-NN
5. Imputation Using Deep Learning
6. Imputation Using Multivariate Imputation by Chained Equation (MICE)

What is A/B testing?

- A/B testing, also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drive business metrics. Essentially, A/B testing eliminates all the guesswork out of website optimization and enables experience optimizers to make data-backed decisions. In A/B testing, A refers to 'control' or the original testing variable. Whereas B refers to 'variation' or a new version of the original testing variable. The version that moves your business metric(s) in the positive direction is known as the 'winner.' Implementing the changes of this winning variation on your tested page(s) / element(s) can help optimize your website and increase business ROI. The metrics for conversion are unique to each website. For instance, in the case of eCommerce, it may be the sale of the products. Meanwhile, for B2B, it may be the generation of qualified leads. A/B testing is one of the components of the overarching process of Conversion Rate Optimization (CRO), using which you can gather both qualitative and quantitative user insights. You can further use this collected data to understand user behaviour, engagement rate, pain points, and even satisfaction with website features, including new features, revamped page sections, etc. If you're not A/B testing your website, you're surely losing out on a lot of potential business revenue.

Is mean imputation of missing data acceptable practice?

It is not an acceptable practice as there are three problems with using mean-imputed variables in statistical analyses:

- Mean imputation reduces the variance of the imputed variables.
- Mean imputation shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence interval.
- Mean imputation does not preserve relationships between variables such as correlations.

What is linear regression in statistics?

- **Linear regression** is a regression model that estimates the relationship between one independent variable and one dependent variable using a straight line. Both variables should be quantitative. ... Linear regression most often uses mean-square error (MSE) to calculate the error of the model. It is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

What are the various branches of statistics?

- There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.
 1. Data collection is all about how the actual data is collected. For the most part, this needn't concern us too much in terms of the mathematics (we just work with what we are given), but there are significant issues to consider when actually collecting data.
 2. Descriptive statistics is the part of statistics that deals with presenting the data we have. This can take two basic forms – presenting aspects of the data either visually (via graphs, charts, etc.) or numerically (via averages and so on).
 3. Inferential statistics is the aspect that deals with making conclusions about the data. This is quite a wide area; essentially you are asking 'What is this data telling us, and what should we do?'