Answer1: A. True

Answer2: A. Central Limit Theorem

Answer3: B. Modeling bounded count data

Answer4: D. All of the mentioned

Answer5: C. Poisson

Answer6: B. False

Answer7: B. Hypothesis

Answer8: A. 0

Answer9. C. Outliers cannot conform to the regression relationship

Answer10:

Normal Distribution - Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve. The normal distribution is the most common type of distribution assumed in technical stock market analysis and in other types of statistical analyses. The standard normal distribution has two parameters: the mean and the standard deviation. For a normal distribution, 68% of the observations are within +/- one standard deviation of the mean, 95% are within +/- two standard deviations, and 99.7% are within +- three standard deviations.

The normal distribution model is motivated by the Central Limit Theorem. This theory states that averages calculated from independent, identically distributed random variables have approximately normal distributions, regardless of the type of distribution from which the variables are sampled. Normal distribution is sometimes confused with symmetrical distribution. Symmetrical distribution is one where a dividing line produces two mirror images, but the actual data could be two humps or a series of hills in addition to the bell curve that indicates a normal distribution.

Answer11:

There are many ways to approach missing data. The most common which I believe is to ignore it. But making no choice means that your statistical software is choosing for you. Most of the time, your software is choosing list wise deletion. List wise deletion may or may not be a bad choice, depending on why and how much data are missing. Another common approach among those who are paying attention is imputation. Imputation simply means replacing the missing values with an estimate, then analyzing the full data set as if the imputed values were actual observed values.

The following are the common methods:

a. Mean Imputation

b. Substitution

c. Hot deck imputation

d. Cold deck imputation

e. Regression imputation

Answer: 12

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 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 behavior, 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.

Answer: 13

True, imputing the mean preserves the mean of the observed data.  So if the data are missing completely at random, the estimate of the mean remains unbiased. That’s a good thing. Plus, by imputing the mean, you are able to keep your sample size up to the full sample size. That’s good too. If all you are doing is estimating means (which is rarely the point of research studies), and if the data are missing completely at random, mean imputation will not bias your parameter estimate. In other words, yes, you get the same mean from mean-imputed data that you would have gotten without the imputations. And yes, there are circumstances where that mean is unbiased. Even so, the standard error of that mean will be too small. Because the imputations are themselves estimates, there is some error associated with them.  But your statistical software doesn’t know that.  It treats it as real data. Ultimately, because your standard errors are too low, so are your p-values.  Now you’re making Type I errors without realizing it.

Answer: 14

Linear regression analysis 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. This form of analysis estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values. There are simple linear regression calculators that use a “least squares” method to discover the best-fit line for a set of paired data. You then estimate the value of X (dependent variable) from Y (independent variable).

You can perform the linear regression method in a variety of programs and environments, including:

* R linear regression
* MATLAB linear regression
* Sklearn linear regression
* Linear regression Python
* Excel linear regression

Answer15:

The two main branches of statistics are descriptive statistics and Inferential statistics. Both of these are employed in scientific analysis of data and both are equally important for the student of statistics.

a. Descriptive Statistics - deals with the presentation and collection of data. This is usually the first part of a statistical analysis. It is usually not as simple as it sounds, and the statistician needs to be aware of designing experiments, choosing the right focus group and avoid biases that are so easy to creep into the experiment. Different areas of study require different kinds of analysis using descriptive statistics. For example, a physicist studying turbulence in the laboratory needs the average quantities that vary over small intervals of time. The nature of this problem requires that physical quantities be averaged from a host of data collected through the experiment.

b. Inferential Statistics - as the name suggests, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics. In the end, it is the inferences that make studies important and this aspect is dealt with in inferential statistics. Most predictions of the future and generalizations about a population by studying a smaller sample come under the purview of inferential statistics. Most social sciences experiments deal with studying a small sample population that helps determine how the population in general behaves. By designing the right experiment, the researcher is able to draw conclusions relevant to his study. While drawing conclusions, one needs to be very careful so as not to draw the wrong or biased conclusions. Even though this appears like a science, there are ways in which one can manipulate studies and results through various means. For example, data dredging is increasingly becoming a problem as computers hold loads of information and it is easy, either intentionally or unintentionally, to use the wrong inferential methods. Both descriptive and inferential statistics go hand in hand and one cannot exist without the other. Good scientific methodology needs to be followed in both these steps of statistical analysis and both these branches of statistics are equally important for a researcher.