11-10-2021

STATISTICS-WORKSHEET 1

Q1. A

Q2. A

Q3. C

Q4. D

Q5. C

Q6. B

Q7. B

Q8. A

Q9. C

Q10. What do you understand by the term 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.

Q11. How do you handle missing data? What imputation techniques do you recommend?

Missing data is a huge problem for data analysis because it distorts findings. It's difficult to be fully confident in the insights when you know that some entries are missing values. Hence, why they must be addressed. According to data scientists, there are three types of missing data. These are Missing Completely at Random (MCAR) – when data is completely missing at random

across the dataset with no discernable pattern. There is also Missing At Random (MAR) – when data is not missing randomly, but only within sub-samples of data. Finally, there is Not Missing at Random (NMAR), when there is a noticeable trend in the way data is missing.

When dealing with missing data, data scientists can use two primary methods to solve the error: imputation or the removal of data.The imputation method develops reasonable guesses for missing data. It's most useful when the percentage of missing data is low. If the portion of missing data is too high, the results lack natural variation that could result in an effective model.The other option is to remove data. When dealing with data that is missing at random, related data can be deleted to reduce bias.

Q12. What is A/B testing?

A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

Q13. Is mean imputation of missing data acceptable practice?

Mean imputation is so simple and yet so dangerous.

- Bad practice in general
- If just estimating means: mean imputation preserves the mean of the observed data
- Leads to an underestimate of the standard deviation
- Distorts relationships between variables by "pulling" estimates of the correlation toward zero source

Q14. What is linear regression in statistics?

Linear regression quantifies the relationship between one or more predictor variable(s) and one outcome variable. Linear regression is commonly used for predictive analysis and modeling. Linear regression is also known as multiple regression, multivariate regression, ordinary least squares (OLS), and regression.

Q15. What are the various branches of statistics?

Statistics is concerned with developing and studying different methods for collecting, analyzing and presenting the empirical data.

Statistics is all about the interpretation of data. To begin, I shall first give a brief, informal overview of the major types of statistics, and discuss the different types of data that we may encounter

Statistics is the branch of mathematics that deals with data. Data is a collection of values. For most of what we do, it will be numerical data, such as the inflation rate, the number of bees in a colony, or the marks in a class test , but it can also take other forms, such as the political party a voter intends to vote for, the football team they support, and soon.

There are three real branches of statistics

- Data Collection
- Descriptive Statistics
- Inferential Statistics