

Statistics work set6

1. D All of the mentioned
2. a) Discrete
3. a) pdf
4. c) mean
5. b) standard deviation
6. b) standard deviation
7. c) 0 and 1
8. b) bootstrap
9. b) summarized

Question 10. What is the difference between a boxplot and histogram?

Answer

Box plot is graphical representation of numeric data value for the given data set .it is use for comparing different features in dataset.it is use to identify outliers and other statistical information such as median, range , min and max value.it is also use to find quartile ranges for any given variable in dataset.

Histogram is graphical representation of numeric data. It is mostly used for continous data. It is used to describe the distribution of data in dataset.

Question 11. How to select metrics?

Answer

To select the correct metrics for any model first we have to check whether given algorithm is regression type or classification type. When the outcome or variable which is needed to be predict is in continuous form that is it is giving any number than we choose the following metrics

- Mean Absolute Error (MAE),
- Mean Squared Error (MSE),
- Root Mean Squared Error (RMSE),
- R^2 (R-Squared)
- Adjusted R-squared

But if outcome or result to be predict is yes/no that is probability of happening or not happening thing than it is classification algorithm. In this case we use different metrics, some of them are mentioned below.

- Accuracy
- Confusion Matrix (not a metric but fundamental to others)
- Precision and Recall
- F1-score
- AU-ROC

Question 12 How do you assess the statistical significance of an insight?

Answer

For statistical significance we have to start with hypothesis value .we have to presume two thing that is null hypothesis and other one is alternate hypothesis.secondly we have to find out p-value and threshold of significance that is alpha value.after that we have to compare p-value and alpha value, if p-value is smaller than alpha value than null hypothesis is rejected and alternate hypothesis is considered and if p-value is greater than alpha value than null hypothesis is considered.

Question 13. Give examples of data that doesnot have a Gaussian distribution, nor log-normal

Answer

Exponential distribution and Any type of categorical data won't have a gaussian distribution or lognormal distribution.

Example :- time when next earthquake will occur.

2 Values of oil reserves among oil fields

Question 14. Give an example where the median is a better measure than the mean.

Answer

Income of any person is best example where we consider median is a better measure than mean

2 survey made for particular group of people and they have to give rating from 1 to 9. So in this case mean may be in

decimel but median will always give any value which will be exact rating value.

Question 15. What is the Likelihood?

Answer

The likelihood function measures the extent to which the data provide support for different values of the parameter. It indicates how likely it is that a particular population will produce a sample.