STATISTICS – WORKSHEET 1

(13-09-2021 to 20-09-2021)

Solutions

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1. Bernoulli random variables take (only) the values 1 and 0.
a) True
b) False

Ans. A

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

Ans. A

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

Ans. B

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

Ans. D

STATISTICS – WORKSHEET 1

(13-09-2021 to 20-09-2021)

Solutions

5. _____ random variables are used to model rates.
a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned

Ans. A

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
a) True
b) False

Ans. B

7. Which of the following testing is concerned with making decisions using data?
a) Probability
b) Hypothesis
c) Causal

d) None of the mentioned

Ans. B

8. Normalized data are centered at_____and have units equal to standard deviations of the original data.
a) 0
b) 5
c) 1
d) 10

Ans. C

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

Ans. D

FLIP ROBO

STATISTICS – WORKSHEET 1

(13-09-2021 to 20-09-2021)

Solutions

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

Normal Distribution: It states that it is symmetric about the mean, this is also called as Gaussian distribution. In graph form it will be in the shape of bell where the mean is 0 and standard deviation is 1 with 0 skewness (This range is an ideal one, which is not possible in reality.)

- All symmetrical distributions are not normal distribution.
- For normal distribution – 68% of data lies within +/- 1 std
  95% of data lies within +/- 2 std
  99.7% of data lies within +/- 3 std

  All the data above 3 std are considered as outliers.

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

There are different ways to handle missing data.

1) Dropping the rows with missing data – This can be done when the no. of missing data is very negligible, i.e. 1 or 2 rows of a column in a large dataset, and If the column is the target variable it is best to drop those rows.
2) Filling with 0 – Based on the column and their behaviour, some columns will be left blank intentionally where there will be no value for them. So those kind of data can be filled with 0.
3) Filling with Mean – Some columns play major role in determining the target variable, they might be missed due to unavailability of data or any other reasons, such kind of rows can be filled with mean or median. Median in case of object data type and mean for numerical data type.
4) Leave as it is.

It is recommended to use these imputation techniques based on the situation

12. What is A/B testing?

A/B testing is also known as bucket testing or split-run testing, which is used to research the user experience. It consists of two variables A and B includes application of hypothesis testing. Generally it is comparison of two versions of a same variable, testing a subject response to the same variable to determine, which is the more effective version.

13. Is mean imputation of missing data acceptable practice?

It is not acceptable in all the cases. It is the last preferred option to replace by mean, that will not preserve the relationship between variables also leads to underestimate of standard error.

14. What is linear regression in statistics?

Linear regression in the statistics is the linear approach for modelling the relationship between the one or more explanatory variables and the response.

15. What are the various branches of statistics?

There are two main branches of statistics, they are:

1) Descriptive Statistics
2) Inferential Statistics