Statistics Worksheet-1

Q1) Bernoulli random variables take (only) the values 1 and 0.

Answer- a) True

Q2) 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?

Answer- a) Central Limit Theorem

Q3) Which of the following is incorrect with respect to use of Poisson distribution?

Answer- b) Modeling bounded count data

Q4) 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

Answer- d)All of the mentioned

Q5) _____ random variables are used to model rates.

Answer- c) Poisson

Q6) Usually replacing the standard error by its estimated value does change the CLT.

Answer- b) False

Q7) Which of the following testing is concerned with making decisions using data?

Answer- b) Hypothesis

Q8) Normalized data are centered at _____and have units equal to standard deviations of the original data.

Answer- a) 0

Statistics Worksheet-1

Q9) Which of the following statement is incorrect with respect to outliers?

Answer- c) Outliers cannot conform to the regression relationship

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.

The key properties of Normal Distribution are-

- A normal distribution is the proper term for a probability bell curve.
- In a normal distribution the mean is zero and the standard deviation is 1.It has zero skew and a kurtosis of 3.
- Normal distributions are symmetrical, but not all symmetrical distributions are normal.

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

Answer-

If data missing at random: deletion has no bias effect, but decreases the power of the analysis by decreasing the effective sample size.

- Recommended: Knn imputation, Gaussian mixture imputation

Q12) What is A/B testing?

Answer- An AB test is an example of statistical hypothesis testing, a process whereby a hypothesis is made about the relationship between two data sets and those data sets are then compared against each other to determine if there is a statistically significant relationship or not

Q13) Is mean imputation of missing data acceptable practice?

Answer-

- 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

Statistics Worksheet-1

Q14) What is linear regression in statistics?

Linear regression analysis is used to predict the value of a variable based on the value of another variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values.

$$Y = a + bX,$$

where ,

Y is the dependent variable

X is the independent variable

b is the slope of the line

a is the y-intercept.

15. What are the various branches of statistics?

Answer-There are three real branches of statistics:

1)Data Collection

2)Descriptive Statistics

3)Inferential Statistics

Data Collection- Data Collection is all about how the actual data is collected.

Descriptive Statistics- Descriptive Statistics deals with the presentation and collection of the data. It is a first part of statistical Analysis.

Inferential Statistics- It 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.