ASSIGNMENT – 39
STATISTICS

---

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

Ans- a) True

---

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) Central Limit Theorem

---

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) Modeling bounded count data

---

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) All of the mentioned

---

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

Ans- c) Poisson

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

Ans- b) False

---
7. 1. 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) Hypothesis(null)

---
8. 4. 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- a) 0

---
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- c) Outliers cannot conform to the regression relationship
---
---
10. What do you understand by the term Normal Distribution?

Ans- 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 always symmetrical, but not all symmetrical distributions are normal.

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

Ans- Types of Missing Data--

1.Missing Completely At Random (MCAR)
2.Missing At Random (MAR)
3.Not Missing At Random (NMAR)

1. Mean or Median Imputation: A common technique is to use the mean or median of the non-missing observations. This can be useful in cases where the number of missing observations is low. However, for large number of missing values, using mean or median can result in loss of variation in data and it is better to use imputations.

2. Multivariate Imputation by Chained Equations (MICE): It imputes data on a variable-by-variable basis by specifying an imputation model per variable. MICE uses predictive mean matching for continuous variables, logistic regressions for binary variables, bayesian polytomous regressions for factor variables, and proportional odds model for ordered variables to impute missing data.

3. Random Forest: Random forest is a non-parametric imputation method applicable to various variable types that works well with both data missing at random and not missing at random. Random forest uses multiple decision trees to estimate missing values and outputs out of bag imputation error estimates.

---

12. What is A/B testing?

Ans- A/B testing is a basic randomized control experiment. It is used to compare the 2 versions of a variable to find out which performs better in a controlled environmen.A/B testing is one of the most prominent and widely used statistical tools. Basically the products are divided into two parts – A and B. Here A will remain unchanged while you make significant changes in B. Now, on the basis of the response from customer groups who used A and B respectively, you try to decide which is performing better.

---

13. Is mean imputation of missing data acceptable practice?

Ans- Yes, 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.

---

14. What is linear regression in statistics?

Ans- Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered as an explanatory variable, and the other is considered as a dependent variable.

---

15. What are the various branches of statistics?

Ans-Statistics is divided into two main branches: (1) Descriptive Statistics (2) Inferential Statistics

(1) Descriptive Statistics:-

Descriptive statistics deals with the collection of data and presents in various forms, such as tables, graphs and diagrams and finding averages and other measures which would describe the data.

(2) Inferential Statistics:-

Inferential statistics deals with techniques used for the analysis of data, making estimates and drawing conclusions from limited information obtained through sampling and testing the reliability of the estimates.