

STATISTICS WORKSHEET-1

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
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
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
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
5. _____ random variables are used to model rates.
 - a) Empirical
 - b) Binomial
 - c) **Poisson**
 - d) All of the mentioned
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
 - a) True
 - 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
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
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

Q10 and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

Ans- Probability distribution which is symmetric about its mean. Shows data near mean occurs more frequently.

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

Ans. Missing data can be completely missing or partly in randomly

1) Delete

- a) Listwise- delete missing items and run only those has complete set of observation
- b) Pairwise- this will assume data is missing randomly not all cases with missing data and with data will be consider
- c) Discard variable itself in case data is more than 60% missing

2) Imputations

- a) Mean , median, mode – in case small number if observations are missing.
- b) Trend , seasonality basis extrapolation
- c) Linear interpolation
- d) Keep same as last observed
- e) K nearest neighbours
Select number of nearest neighbours and their distance metric. KNN identify most frequent value among neighbours and mean among the nearest neighbours

12. What is A/B testing?

Ans- It compares two versions of same variables to find out which performs better in controlled environment.

Eg-

100 customers

Newsletter 1 A and newsletter 2 B

Objective which newsletter brings higher traffic

H_0 = There is no difference in customer bringing A and B

H_a = conversion rate of B is higher than A

Control group = will receive newsletter A, 50 customers

Test group will receive newsletter B, 50 customers

Calculate daily conversion rate

For control grp = say 16%

For test grp = say 19%

Significance test

Choose H_0 or H_a is true.

13. Is mean imputation of missing data acceptable practice?

Ans. – Its not.

Because its doesn't consider relationship between variables, so results can be biased
Mean imputations can lead to underestimation of standard errors so the p values and going interpretation wrong.

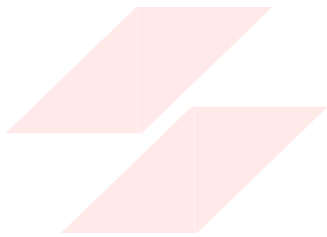
14. What is linear regression in statistics?

Ans- Linear model assumes linear relationships between multiple input variables x and single output variable y

Ex- $Y = ax + c$

15. What are the various branches of statistics?

Ans- descriptive and inferential statistics



FLIP ROBO