**FLIP ROBO**

# 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
   Ans. *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

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

**FLIP ROBO**

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

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

*The Normal Distribution is a continuously probability distribution that is symmetrical around its mean most of the observations cluster around the central peak.it is also called as Bell like curve. When mean is equal to median is equal to mode is equal to zero.and STD is +-1.*

11. How do you handle missing data? What imputation techniques do you recommend?
**Handle missing data are as follows:**
Listwise or case deletion. ...
Pairwise deletion. ...
Mean substitution. ...
Regression imputation. ...
Last observation carried forward. ...
Maximum likelihood. ...
Expectation-Maximization. ...
Multiple imputation.

**Imputation Technique**
**Complete Case Analysis(CCA):-** This is a quite straightforward method of handling the Missing Data, which directly removes the rows that have missing data i.e we consider only those rows where we have complete data i.e data is not missing. ...
**Arbitrary Value Imputation. ...**
**Frequent Category Imputation**.

12. What is A/B testing?
*A/B testing (also known as bucket testing or split-run testing) is a user experience research methodology. A/B tests consist of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing or "two-sample hypothesis testing" as used in the field of statistics.*

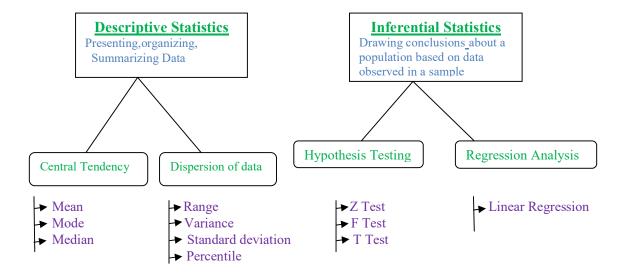13. Is mean imputation of missing data acceptable practice?

*Mean imputation is typically considered terrible practice since it ignores feature correlation. mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.*

14. What is linear regression in statistics?
*In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables).*

15. What are the various branches of statistics?
Ans. *There are mainly two types of statistics:*

| **Descriptive Statistics**<br>Presenting, organizing,<br>Summarizing Data | **Inferential Statistics**<br>Drawing conclusions about a<br>population based on data<br>observed in a sample |
|---|---|

| Central Tendency | Dispersion of data | Hypothesis Testing | Regression Analysis |
|---|---|---|---|

| Central Tendency | Dispersion of data | Hypothesis Testing | Regression Analysis |
|---|---|---|---|
| → Mean | → Range | → Z Test | → Linear Regression |
| → Mode | → Variance | → F Test | |
| → Median | → Standard deviation | → T Test | |
| | → Percentile | | |

**FLIP ROBO**