Statistics worksheet – 1

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

2. Which of the following theorem states that the distribution of averages of iid variables, properly?
   a) Central Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution?
   b) Modeling bounded count data

4. Point out the correct statement.
   c) All of the mentioned

5. _____ random variables are used to model rates.
   c) Poisson

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

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

8. Normalized data are centered at_____and have units equal to standard deviations of the original data.
   a) 0

9. Which of the following statement is incorrect with respect to outliers?
   c) Outliers cannot conform to the regression relationship

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

- The mean, median, and mode are all equal.
- The curve is known to be symmetric at the center, which is around the mean.
- Exactly 1/2 of all the values are known to be to the left of center whereas exactly half of all the values are to the right of the center.
- The total area under the curve is 1.

11. How do you handle missing data? What imputation techniques do you recommend?
When dealing with missing data, we can use two methods: imputation or the removal of data.
1. Simple Data Imputation

2. Explicit Modeling
a. Mean Imputation
b. Regression Imputation
c. Stochastic Regression Imputation
3. Implicit Modeling
a. Hot deck Imputation
b. Substitution
c. Cold deck Imputation
d. Composite Method (hybrid)
4. Multiple Data Imputation
5. Composite Data Imputation
6. Cascading Data 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. A/B testing is a way to compare two versions of a single variable, typically by testing a subject's response to variant A against variant B, and determining which of the two variants is more effective.

13. Is mean imputation of missing data acceptable practice?

It is but it ought to be a last resort. It is a popular solution to missing data, despite its drawbacks.

14. What is linear regression in statistics?

Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things:

1. Does a set of predictor variables do a good job in predicting an outcome (dependent) variable?

2. Which variables in particular are significant predictors of the outcome variable, and in what way do they – indicated by the magnitude and sign of the beta estimates – impact the outcome variable?

These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula y = c + b*x, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

15. What are the various branches of statistics?

1. Descriptive Statistics
A. Measure of Central Tendency
a. Mean
b. Median
c. Mode

B. Measure of Dispersion
a. Range
b. Standard Deviation

c. Variance
d. Percentile

2. Inferential Statistics
A. Regression analysis
B. Analysis of variance (ANOVA)
C. Analysis of covariance (ANCOVA)
D. Statistical significance (t-test)
E. Correlation analysis