**Statistics worksheet solution**

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: c)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:Poisson**

6. 10. Usually replacing the standard error by its estimated value does change the CLT.

a) True b) False

**Ans:a)True**

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**

8. 4. Normalized data are canteredat \_\_\_\_\_\_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:Normal distribution is also known as the Gaussian distribution. It 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. In graph form, normal distribution will appear as a bell curve.

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

Ans:Missing data appear when no value is available in one or more variables of an individual.

1. Deletions. Pairwise Deletion. Listwise Deletion/ Dropping rows. Dropping complete columns.
2. Basic Imputation Techniques. Imputation with a constant value. Imputation using the statistics (mean, median, mode)
3. K-Nearest Neighbour Imputation.

Imputation Techniques

* Complete Case Analysis(CCA)
* Arbitrary Value Imputation.
* Frequent Category Imputation.

12. What is A/B testing?

Ans:A/B testing also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drive business metrics.

13. Is mean imputation of missing data acceptable practice?

Ans: If the number of missing values in a variable is large, and these values are imputed by the observed sample mean, then the resulting variance estimated for that variable can be severely underestimated.

14. What is linear regression in statistics?

Ans: Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

15. What are the various branches of statistics?

Ans:There are three real branches of statistics:

 data collection, descriptive statistics and inferential statistics.