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

a. True
b. False

Answer: 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

Answer: 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

Answer: 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

Answer: d) All of the mentioned

5. _____random variables are used to model rates.

a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned

Answer: c) Poisson

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

a) True
b) False

Answer: b) False (Does not change the CLT)

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

a) Probability
b) Hypothesis
c) Causal
d) None of the mentioned

Answer: b) Hypothesis

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

a) 0
b) 5
c) 1
d) 10

Answer: a) 0

9.Which of the following statement is incorrect with respect to outliers?

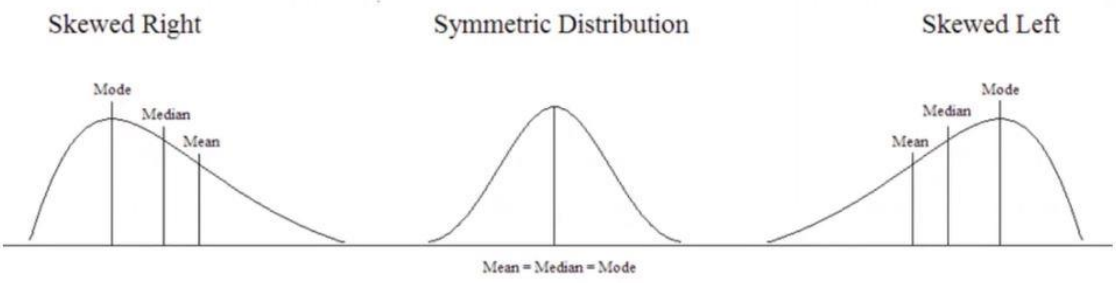
a) Outliers can have varying degrees of influence
b) Outliers can be the result of spurious or real processes
c) Outliers cannot conform to the regression relationship
d) None of the mentioned

Answer: 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.
- Normal Distributions are symmetrical, but not all the symmetrical distributions are normal distribution.
- Many naturally occurring phenomena tend to approximate the normal distribution.



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

Ans:

By imputing or deleting the missing values, we can handle the missing data. There are many imputing techniques to handle the data.

The following are common methods:

- Mean imputation. Simply calculate the mean of the observed values for that variable for all individuals who are non-missing. ...
- Substitution. ...
- Hot deck imputation. ...
- Cold deck imputation. ...
- Regression imputation. ...
- Stochastic regression imputation. ...
- Interpolation and extrapolation.

12. What is A/B testing?

Ans:

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?
Ans:

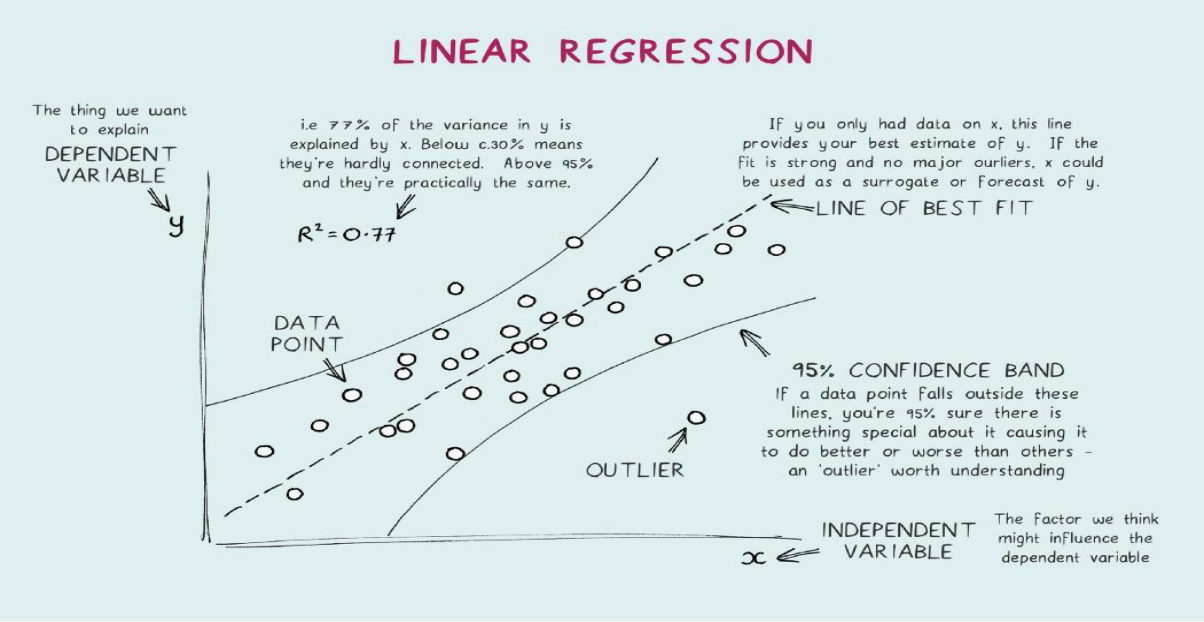
In mean imputation we will take the mean of all the variables of one column and fill the missing data with the mean value. Although Mean imputation is very simple and to apply, But yes there is some drawbacks of using this imputation:

1. Mean substitution leads to bias in multivariate estimates such as correlation or regression coefficient. Values that are imputed by a variable's mean have, in general, a correlation of zero with other variables. Relationship between variables are therefore biased towards zero.

2. Standard errors and variance of imputed variables are biased. For instance, Let's assume that we would like to calculate the standard error of a mean estimation of an imputed variable. Since all imputed values are exactly the mean of our variable, we would be so sure about the correctness of our mean estimate. In other words, the confidence interval around the point estimation of our mean would be too narrow.

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:

The real branches of statistics are:

- Descriptive Statistics
- Inferential Statistics

Descriptive Statistics

1. Measure of central tendency (mean, median, mode).

2. Measure of dispersion(spread) (Variance and Standard deviation).

Measure of Dispersion

$$\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$$

x = data, Mu = Mean, N = Occurrences

Inferential Statistics: Sampling Data and infer the result to describe entire population.