Statistic Worksheet

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

a) True

b) False

ANS: a)

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: a)

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)

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: c)

5. _____ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

ANS: c)

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

a) True

b) False

ANS: b)

7. 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)

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

ANS: a)

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

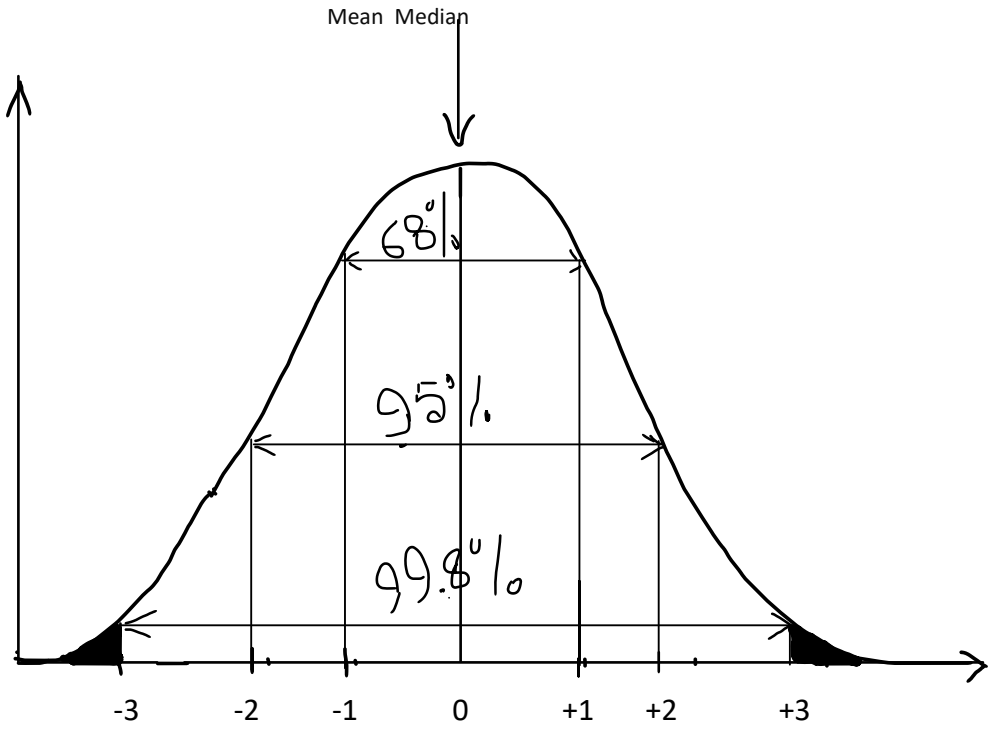
c) Outliers cannot conform to the regression relationship

d) None of the mentioned

ANS: c)

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

ANS: Normal distribution, also known as Gaussian distribution, is a continuous probability distribution that is symmetric around its mean, forming a bell-shaped curve when plotted.In a normal distribution, the mean values of a dataset are clustered around the median values,which implies that most values are allowed to be symmetrical.



Normal Distribution Plot

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

ANS: To handle missing data from a dataset there are several way to treat the missing values:

1)Deletion: Remove rows or columns with missing values. This approach is good but may lead to loss of valuable information.

2)Imputation: Fill in missing values with their mean, median, or mode values. This maintain sample size but can introduce bias if not done carefully.

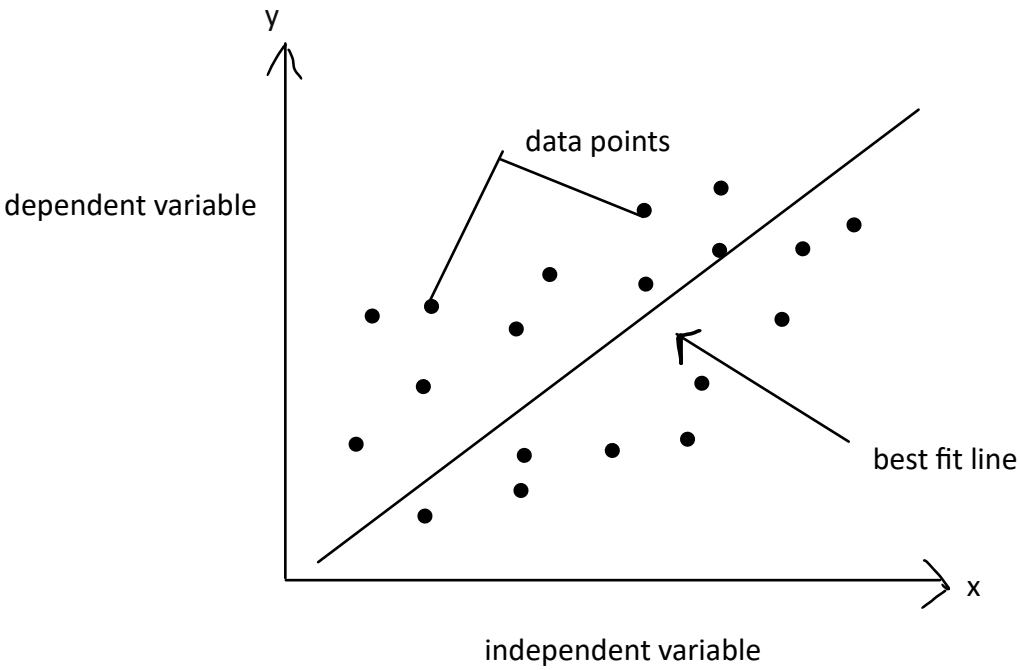
I recommend using mean or median imputation for numerical variables and mode imputation for categorical variables.

13. Is mean imputation of missing data acceptable practice?

ANS: Yes, mean imputation of missing data is a commonly accepted practice. It involves replacing missing values with the mean of the observed data for that variable.

14. What is linear regression in statistics?

ANS: Linear regression is a statistical method that establishes a straight line between two variables in a general correlation coefficient.It is commonly used to predict value of a single dependent or multiple dependent variables.



15. What are the various branches of statistics?

ANS: