

STATISTICS WORKSHEET-1

- 1 = a) True
- 2 = a) Central Limit Theorem
- 3 = b) Modeling bounded count data
- 4 = d) All of the mentioned
- 5 = c) Poisson
- 6 = b) False
- 7 = b) Hypothesis
- 8 = a) 0
- 9 = c) Outliers cannot conform to the regression relationship

SUBJECTIVE ANSWER

10 = Normal distribution also known as the gaussian distribution is a probability distribution that is symmetric that data near the mean are more frequent in occurrence than data far from the mean. The normal distribution appears as a “bell curve”. When graphed. The middle of the range is also known as the mean of the distribution

11 = There are 3 types of imputation. 1) mean imputation 2) median imputation
3) mode imputation

Replace missing values with the mean of the variable. Replace missing values with the median of the variable. Replace missing values with the most frequent value of the variable.

When dealing with missing data data scientists to solve the error “imputation or data removal”. The imputation method substitutes reasonable guesses for missing data. It’s most useful when the percentage of missing data is low.

12 = AB testing is comparing two versions of a web page to figure out the better performing variation it tests a control (A) version against a variant (B) version to measure which one is

most successful based on your key metrics .A\B testing is a controlled experiment in which you run two different product or website versions and see which performs better.

13 = imputing the mean preserves the mean of the observed data so if the data are missing completely at random the estimate of the mean remains unbiased. that's a good thing mean imputation is typically considered terrible practice since it ignores feature correlation mean imputation decreases the variance of our data while increasing bias.

14 = 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 linear regression is graphically depicted using a straight line of best fit with the slope defining how the change in one variable impacts a change in the other.

15 = There are two main branch of statistics descriptive branch of statistics and inferential. statistics both of these are employed in scientific analysis of data and both are equally important for the student of statistics.