STATISTICS WORKSHEET – 1

CHOOSE THE CORRECT OPTION :-

1. Option(A) [Bernoulli random variables takes the value 1 and 0].
2. Option(A) [Central limit theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases].
3. Option(B) [Poisson distribution is used for modeling unbounded count data].
4. Option(D) [Many random variables, properly normalized, limit to a normal distribution].
5. Option(C) [Poisson distribution is used to model counts].
6. Option(B) [Usually replacing the standard error by its estimated value doesn't change the CLT].
7. Option(B) [Hypothesis testing is concerned with making decisions using data].
8. Option(A) [Normalized data are centered at 0 and have units equal to standard deviations of the original data].
9. Option(C) [Outliers can conform to the regression relationship].

SUBJECTIVE ANSWER TYPE QUESTION :-

10. The normal distribution is a continuous probability distribution that is symmetrical around its mean with most values near the central peak. also known as called Gaussian distribution and Bell curve.
11. Best techniques to handle missing data :-
    - Use deletion methods to eliminate missing data. The deletion methods only work for certain datasets where participants have missing fields.
    - Use regression analysis to systematically eliminate data.
    - Use data imputation techniques.
12. A/B testing is a randomized experiment where it aims to compare two variants/versions (A vs B) of something (webpage,

images, etc.) to determine which variation performs better for a given conversion goal based on user experience,

also known as split testing.

In statistics, an A/B test is an example of statistical hypothesis testing, a process whereby a hypothesis is made about the relationship between two data sets and those data sets are then compared against each other to determine if there is a statistically significant relationship or not.

13. Mean imputation does not preserve the relationships among variables. True, 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.
14. A regression model that estimates the relationship between one independent variable and one dependent variable using a straight line,called linear regression in statistics. Linear regression looks at various data points and plots a trend line.
15. The two main branches of statistics are descriptive statistics and inferential statistics. Both of these are employed in scientific analysis of data and both are equally important for the student of statistics.

Descriptive Statistics

Descriptive statistics deals with the presentation and collection of data. This is usually the first part of a statistical analysis.

Inferential Statistics

Inferential statistics, as the name suggests, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics.