STATISTICS WORKSHEET-4

ANSWERS:

1. Central Limit Theorem states that the sample distribution becomes more and more normal when the sample size increases. It's useful when analysing big datasets as it allows the assumption of normally distributed dataset besides outliers.
2. One of the primary steps for a statistical analysis where a predetermined number of observations are taken from a population set. There are 2 main types of sampling: 1. Probability sampling – involves random selection, helping make strong inferences about the whole group., 2. Non-Probability sampling – involves collecting data based on criteria or convenience for easy data collection.
3. Type 1 Error – If Ho is true and we reject Ho. Type 2 Error – If Ho is false and we fail to reject Ho.
4. Normal distribution (or bell curve distribution) is a distribution that is symmetric about the mean, where essentially data near mean are far more frequent than data far from mean (when symmetric). It can also be skewed to the left or skewed to the right.
5. Covariance is the measure of difference between 2 random variables and to what extent, they change together. Correlation is used to see relation between 2 random variables.
6. Univariate statistics summarizes only 1 variable at a time. Bivariate summarizes 2 variables at a time whereas multivariate summarizes 2 or more variables at a time.
7. Sensitivity analysis determines how changes in different independent variables affect a particular dependant variable. It can be calculated by leveraging formulas that reference different input cells.
8. Hypothesis testing I making a guess based on assumptions without scientific proof or explaining the situation based on reasonable assumptions. Ho is the Null Hypothesis or current assumption. H1 is the alternate hypothesis or the assumption opposite to Ho. In Two Tail test, Ho and H1 are not equal in any direction.
9. Quantitative data is measures of values or counts expressed in numbers. Qualitative data is measures of types and is expressed in alphanumerical values.
10. Range is calculated by subtracting the lowest value from the highest value. IQR is calculated by subtracting the 1^{st} quantile value from the 3^{rd} quantile value.
11. Bell curve distribution or normal distribution is a distribution that is symmetric about the mean, where essentially data near mean are far more frequent than data far from mean (when symmetric). It can also be skewed to the left or skewed to the right.
12. The simplest method to find outliers is to use iqr to set up lower and upper outlier limit.

13. P-value measures the probability of obtaining the observed results, assuming the null hypothesis is true. If p value lower than alpha level, we reject Ho and if it's greater than alpha level we fail to reject Ho.
14. Binomial distribution formula:
    - $P(x=r) = nCr\ p^r\ q^{n-r}$
    - Where, n-number of trials, r-number of success desired, p-probability of success in 1 trial, q = 1-p
15. ANOVA is another statistical tool which splits an observed variance data into different components for additional tests. ANOVA is most helpful for 3 or more variables. It's similar to the 2 sample t-test but results in less type 1 errors.