STATISTICS WORKSHEET 4

Q1 TO 15 Descriptive Answers

1) The central limit theorem establishes that, in many situations, when independent random variables are summed up, their properly normalized sum tends toward a normal distribution even if the original variables themselves are not normally distributed. It is important to use in hypothesis testing and also to calculate confidence intervals.
2) Sampling is the action or process of taking samples of something for analysis. A technique to reflect the results of the entire population by studying the results of each sample taken from the population. There are two types of sampling methods: Probability sampling involves random selection, allowing to make strong inferences about the whole group. Non-probability sampling involves non-random selection based on convenience or other criteria which allows to collect data easily.
3) The difference between type1 and type2 error is that there is a rejection of reality by the researcher in type one error, whereas the researcher accepts the false reality in type two error.
4) The normal distribution is a type of continuous probability distribution for a real-valued random variable; symmetrical around its mean, most of the observations cluster around the central peak, and the probabilities for values further away from the mean taper off equally in both directions.

5) In statistics, Covariance is a measure of the joint variability of two random variables. Correlation means a mutual relationship or connection between two or more things.
6) The difference between univariate, Bivariate, and multivariate analysis: Univariate statistics summarize only one variable at a time. Bivariate statistics compare two variables. Multivariate statistics compare more than two variables.
7) Sensitivity analysis is a method for predicting the outcome of a decision if a situation turns out to be different compares to the key predictions. Calculate sensitivity:

   Sensitivity = $\frac{\text{Total Positive Tests}}{\text{Total True Positives + Total False Negatives}}$

8) Hypothesis testing allows to make probabilistic statements about population parameters. H0 is a null hypothesis while H1 is an alternative hypothesis. Two-tailed hypothesis tests are called as non-directional and two sided tests because test for effects in both directions. **$H_1: \mu < \mu_0$**; where a decrease is hypothesized and this is called a lower tailed test, **$H_1: \mu \neq \mu_0$** where a difference is hypothesized and this is called a two tailed test.

9) Quantitative data is countable or measurable, related to the numbers, lends itself to the statistics analysis and on the other hands qualitative data is descriptive, relating to words and grouped to categorize according to themes.

10) Calculation of Interquartile Range:

- Interquartile Range = 3rd Quartile – 1st Quartile

11) It refers the mathematical concept that means to describe normal distribution, sometimes referred to a Gaussian distribution.

Bell curve is the bell shape that is created when a line is plotted using the data points for an item that meets the criteria of normal distribution.

12) One method to find outliers:

Box plots are a visual method to identify outliers.

13) The level of marginal significance within the hypothesis testing that represents the probability of occurrence of the given event. This is called P-value in hypothesis testing.

14) Formula of Binomial Probability:

$$P(X) = {}^{n}C_{x}\, p^{x}\, q^{n-x}$$

15) ANOVA is a method to determine if the mean of groups are different. It is a statistical analysis tool, use it to analyse data and produce a table that presents the results of a statistical analysis and used in a wide variety of real-life situations, but the most common include: Retail: Store are often interested in understanding whether different types of promotions, store layouts, advertisement tactics, etc. lead to different sales.

THANK YOU