STATISTICS

ASSIGNMENT NO – 4

Q1. D)

Q2. A)

Q3. A)

Q4. C)

Q5. C)

Q6. A)

Q7. C)

Q8. B)

Q9. A)

Q10. A histogram is a graphical representation of the distribution of a dataset, showing the number of observations that fall within specified intervals or bins. It shows the shape, centre, and spread of the data.

A boxplot, on the other hand, is a standardized way of displaying the distribution of data based on five number summary (minimum, first quartile, median, third quartile, and maximum). It provides a more concise summary of the distribution and identifies outliers.

In summary, a histogram displays the frequency distribution of the data, while a boxplot summarizes the distribution and highlights its skewness and outliers.

Q11. prioritize objectives, examine which metric consistently predicts their achievement, and identify which activities influence predictors, in that order. And continuously re-evaluate this process to keep up with the times.

Q12. Here are some steps to assess the statistical significance of an insight:

1. Choose a hypothesis test: Determine the appropriate hypothesis test for the type of data you have and the question you are trying to answer.
2. Define the null and alternative hypothesis: The null hypothesis states that there is no effect or relationship, while the alternative hypothesis states that there is an effect or relationship.
3. Choose a significance level: The significance level, typically denoted by alpha (α), is the threshold for rejecting the null hypothesis. Common significance levels are 0.05 or 0.01.
4. Calculate the test statistic: Use the appropriate formula for the chosen hypothesis test to calculate the test statistic.
5. Determine the p-value: The p-value is the probability of observing a test statistic as extreme or more extreme than what was actually observed, assuming the null hypothesis is true.
6. Compare the p-value to the significance level: If the p-value is less than the significance level, reject the null hypothesis and conclude that the result is statistically significant.
7. Interpret the results: Finally, interpret the results in the context of the problem and make conclusions about the significance of the insight.

8. It is important to keep in mind that statistical significance does not guarantee practical significance, and that other factors such as sample size and data quality should also be considered.

Q13. Examples of data that does not have a Gaussian (normal) distribution nor log-normal distribution include:

1. Exponential distribution: often used to model waiting times between events in a Poisson process.
2. Pareto distribution: used to model the distribution of wealth, income, and other quantities that follow a power law.
3. Weibull distribution: used to model the time to failure for mechanical and engineering systems.
4. Poisson distribution: used to model the number of events that occur in a fixed interval of time or space.
5. Gamma distribution: used to model the time between events in a Poisson process, as well as various other continuous positive-valued random variables.
6. Cauchy distribution: often used to model the distribution of residual errors in regression analysis.

These are just a few examples, and there are many other distributions that could be used to model different types of data, depending on the characteristics of the data and the problem being addressed.

Q14. The median is a better measure than the mean in the presence of outliers or skewed distributions. For example:

1. Income distribution: The income distribution of a country may be heavily skewed, with a small percentage of the population earning much more than the majority. In this case, the mean income would be significantly higher than the median income, and the median would be a better representation of the typical income.
2. Housing prices: Housing prices in a city may have a few extremely expensive properties, which would drive the mean higher and make it a poor representation of the typical housing price. In this case, the median would be a better measure of the typical housing price.
3. Medical costs: The cost of medical procedures may have a few outliers, such as rare and expensive treatments, which would drive the mean higher and make it a poor representation of the typical medical cost. In this case, the median would be a better measure of the typical medical cost.

These are just a few examples, and the choice between the mean and median as a measure of central tendency will depend on the specific data and problem being addressed.

Q15. Likelihood is a concept in statistical modelling and estimation that quantifies the probability of observing the data given a particular set of parameters for a statistical model. The likelihood function is a function of the parameters of the model and the observed data, and it provides a way to evaluate the goodness-of-fit of the model to the data. The maximum likelihood estimate (MLE) is the set of parameter values that maximize the likelihood function, and it is often used as a point estimate of the parameters.

The likelihood is important in Bayesian statistics, where it is used in conjunction with prior information to update beliefs about the parameters of a model. The likelihood is also used in

frequentist statistics, where the MLE is used as an estimator of the parameters, and the likelihood ratio test is used to test hypotheses about the parameters.

In summary, the likelihood is a key tool for evaluating the fit of a statistical model to the data and for making inferences about the parameters of the model.