STATISTICS WORKSHEET 6 ASSIGNMENT

Q-1:- (D) All of the mentioned.

Q-2:- (A) Discrete.

Q-3:- (A) pdf.

Q-4:- (C) Mean.

Q-5:- (C) Empirical Mean.

Q-6:- (B) Standard deviation.

Q-7:- (C) 0 and 1.

Q-8:- (B) Bootstrap.

Q-9:- (B) Summarized.

Q-10:- Histograms and box plots are similar as they both help to visualize and describe numeric data.

Histograms are better in determining the underlying distribution of the data.

Box plots allow you to compare multiple data sets better than histograms as they are less detailed and take up less space.

Q-11:- Metrics are measures of quantitative assessment commonly used for assessing, comparing, and tracking performance or production.

Generally, a group of metrics will typically be used to build a dashboard that manages or analyse review on a regular basis to maintain performance assessments, opinions, and business strategies.

In order to establish a useful metric, a manager must first assess its goals. From there, it is important to find the best outputs that measure the activities related to these goals. The final step is to set goals and targets for KPI metrics that are integrated with business decisions.

Q-12:- Statistical significance can be accessed using hypothesis testing:-

1. Stating a null hypothesis which is usually the opposite of what we wish to test (classifiers A and B perform equivalently, Treatment A is equal of treatment B) .
2. Then, we choose a suitable statistical test and statistics used to reject the null hypothesis.
3. Also, we choose a critical region for the statistics to lie in that is extreme enough for the null hypothesis to be rejected (p-value) .
4. We calculate the observed test statistics from the data and check whether it is lying in critical region or not.

Some common Tests are:- Chi-squared and ANOVA.

Q-13:- Some of the examples are as follows:-

1. Allocation of wealth among individuals.
2. Values of oil reserves among oil fields (many small ones, a small number of large ones).

3.The simplest example is the distribution of numbers that shows up on the top of a fair die after a large number of throws.

Each number from 1 to 6 will occur approx. equal frequency. Increasing the number of throws will not tend to produce a bell-shaped histogram, in fact the fractional occurrence will approach a constant 1/6 over the possible numbers.

Q-14:- For ex:- You execute a customer satisfaction survey with a sample of 9 and rate their overall satisfaction scores on a scale of 1 to 10. You get an average of 5.2. You know that in general, you tend to retain customers with a score over 3, so you're satisfied, because this indicates that you're still above where you want to be. But then, suddenly, you lose 6 of those 9 customers. You go back to look at your data, and you find the scores as : 1, 3, 3, 3, 3, 5, 9, 10..

The median of this group is 3, indicating that at least half of your customers were unhappy. The scores became smaller because of the unexpected 10's, and you missed out on an important part of your data – the midpoint that indicated that as many as half of your customers or more were dissatisfied with the company.

Median can play a major role in things like income level, research as well because a few millionaires may make it look like a socio-economic status of your sample being higher than what it really is.

Q-15:- In statistics, the likelihood function measures the goodness of fit of a statistical model to a sample of data for given values of the unknown parameters. It is formed from the joint probability distribution of the sample, but viewed and used as a function of the parameters only Thus, treating the random variables as fixed at the observed values.

The likelihood function describes a hypersurface whose peak, if it exists, represents the combination of model parameter values that maximize the probability of drawing the sample obtained.

The procedure for obtaining these arguments of the maximum of the likelihood function is known as maximum likelihood estimation, which for computational convenience is usually done using the natural logarithm of the likelihood, known as the log-likelihood function also.