STATISTICS WORKSHEET 6

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1. D

2. A

3. A

4. C

5. D

6. A

7. C

8. B

9. B

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What is the difference between a boxplot and histogram?

A. Histograms are sometimes called **Frequency Plots while boxplots are referred to as Box-and-Whisker Plots**. Histograms and boxplots can be drawn either vertically or horizontally. There are many graphical tools that can generate histograms and boxplots quickly and easily (such as Minitab). Histograms are often confused with bar charts.

11. How to select metrics?

A. : selection metrics Performance metrics are defined as information and project-specific data used to characterize and assess an organization’s quality, capabilities, and skills. Performance metrics are defined differently in every industry and can change based on a company’s services and products.

12. How do you assess the statistical significance of an insight?

A. Statistical significance can be accessed using hypothesis testing:
– 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)
– Then, we choose a suitable statistical test and statistics used to reject the null hypothesis
– 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)
– We calculate the observed test statistics from the data and check whether it lies in the critical region
Common tests:
– One sample Z test
– Two-sample Z test
– One sample t-test
– paired t-test
– Two sample pooled equal variances t-test
– Two sample unpooled unequal variances t-test and unequal sample sizes (Welch's t-test)
– Chi-squared test for variances
– Chi-squared test for goodness of fit

13. Give examples of data that does not have a Gaussian distribution, nor log-normal.

A. The techniques that assume Gaussian or Gaussian-like distribution are listed below: Unfortunately, many real-life data are not normal. Permeability distribution of rock samples is lognormal.

14. Give an example where the median is a better measure than the mean.

A. You have a large dataset filled with information, but if you use different analysis methods, you may get different results. One of the most common errors when it comes to survey research comes from this issue, and interestingly it is one of the most basic types of analysis: Mean vs Median

15. What is the Likelihood?

A. The likelihood function describes a Hyper surface 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**. Additionally, the shape and curvature of the likelihood surface represent information about the stability of the estimates, which is why the likelihood function is often plotted as part of a statistical analysis.