STATISTICS WORKSHEET- 6

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

1. Which of the following can be considered as random variable?
   a) The outcome from the roll of a die
   b) The outcome of flip of a coin
   c) The outcome of exam
   d) All of the mentioned

Answer: d) All of the mentioned

2. Which of the following random variable that take on only a countable number of possibilities?
   a) Discrete
   b) Non Discrete
   c) Continuous
   d) All of the mentioned

Answer: a) Discrete

3. Which of the following function is associated with a continuous random variable?
   a) pdf
   b) pmv
   c) pmf
   d) all of the mentioned

Answer: a) pdf

4. The expected value or_____of a random variable is the center of its distribution.
   a) mode
   b) median
   c) mean
   d) bayesian inference

Answer: c) mean

5. Which of the following of a random variable is not a measure of spread?
   a) variance
   b) standard deviation
   c) empirical mean
   d) all of the mentioned

Answer: a) variance

6. The_____of the Chi-squared distribution is twice the degrees of freedom.
   a) variance
   b) standard deviation
   c) mode
   d) none of the mentioned

Answer: a) variance

7. The beta distribution is the default prior for parameters between _____
   a) 0 and 10
   b) 1 and 2
   c) 0 and 1
   d) None of the mentioned

Answer: c) 0 and 1

8. Which of the following tool is used for constructing confidence intervals and calculating standard errors fordifficult statistics?
   a) baggyer
   b) bootstrap
   c) jacknife
   d) none of the mentioned

Answer: b) bootstrap

9. Data that summarize all observations in a category are called_____data.
   a) frequency
   b) summarized
   c) raw
   d) none of the mentioned

Answer: b) summarized

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

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

Answer: **Histograms** and **box plots** are graphical representations for the frequency of numeric data values. Major differences between them are Histograms are preferred to determine the underlying probability distribution of a data, while Box plots on the other hand are more useful when comparing between several data sets

11. How to select metrics?

Answer: The key point is to choose metrics that clearly indicate where we are now in relation to our goals. Good metrics can be improved. Good metrics measure progress, which means there needs to be room for improvement.

To select metrices, we need to:

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

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

Answer:
Create a null hypothesis.
Create an alternative hypothesis.
Determine the significance level.
Decide on the type of test you'll use.
Perform a power analysis to find out your sample size.
Calculate the standard deviation.
Use the standard error formula.
Determine the t-score.
Find the degrees of freedom.
Use a t-table.

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

Answer: Many random variables have distributions that are asymptotically Gaussian but may be significantly non-Gaussian for small numbers. For example, the Poisson Distribution, which describes (among other things) the number of unlikely events occurring after providing a sufficient opportunity for a few events to occur.

The simplest example is the distribution of numbers that show up on the top of a fair die after a large number of throws. Each number from 1 to 6 will occur with approximately 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.

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

Answer: For a skewed distribution, the median is often the best measure of central tendency. For example,if we want to know the average person salary in a city then the mean would be highly biased because of thelarge concentration of high salary people towards the right. The median on the other hand gives the 50-percentile data that best explains the central tendency in this case.

Therefore, the median in this case gives a better understanding of an average person’s salary compared to the mean which happens to move closer to the richer people's salary and is not a good indicator mostly.

15. **What is the Likelihood?**

Answer: In statistics, the **Likelihood** function (often simply called the likelihood) measures the goodness of fit of a statistical model to a sample of data for given values of the unknown parameters