STATISTICS WORKSHEET- 3

Question) 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) All of the mentioned

Question) 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) Discrete

Question) Which of the following function is associated with a continuous random variable ?

a) pdf

b) pmv

c) pmf

d) all of the mentioned

Answer) pdf (Probability density function)

Question) The expected value or _____ of a random variable is the center of its distribution.

a) mode

b) median

c) mean

d) bayesian inference

Answer) Mean

Question) 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) Varience

Question) 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) Mode

Question) 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) 0 and 1

Question) Which of the following tool is used for constructing confidence intervals and calculating standard errors for difficult statistics ?

a) baggyer

b) bootstrap

c) jacknife

d) none of the mentioned

Answer) Bootstrap

Question) Data that summarize all observations in a category are called _____ data.

a) frequency

b) summarized

c) raw

d) none of the mentioned

Answer) Summarized

Question) What is the difference between a boxplot and histogram ?

Answer) Histograms are bar charts that show the frequency of a numerical variable’s values and are used to approximate the probability distribution of the given variable.

Boxplot gather other information like the quartiles, the range, and outliers. Boxplots are especially useful when you want to compare multiple charts at the same time because they take up less space than histograms.

Question) How to select metrics ?

Answer) The metric(s) chosen to evaluate a machine learning model depends on various factors:

a) To find a regression or classification task.

b) To find the business objective like precision vs recall

c) To know distribution of the target variable.

There are a number of metrics that can be used, including adjusted r-squared, MAE, MSE, accuracy, recall, precision, f1 score, and the list goes on.

Question) How do you assess the statistical significance of an insight ?

Answer) First, you would state the null hypothesis and alternative hypothesis. Second, you would calculate the p-value, the probability of obtaining the observed results of a test assuming that the null hypothesis is true. Last, you would set the level of the significance (alpha) and if the p-value is less than the alpha, you would reject the null — in other words, the result is statistically significant.

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

Answer) Any type of categorical data won't have a gaussian distribution or lognormal distribution. Exponential distributions — e.g. the amount of time that a car battery lasts or the amount of time until an earthquake occurs.

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

Answer) When there are a number of outliers that positively or negatively skew the data.

Question) What is the Likelihood ?

Answer) The probability of some of the observed outcomes under specific parameter values is regarded as the likelihood of the set of parameter values under certain observed outcomes.