**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

**ANS: 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

**ANS: 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

**ANS: 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

**ANS: 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

**ANS: 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

**ANS: 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

**ANS: c) 0 and 1**

8. **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

**ANS: 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

**ANS: b) summarized**

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

**Ans**: Histograms and box plots are**graphical representations for the frequency of numeric data values.**

**A frequency distribution shows how often each different value in a set of data occurs. A histogram is the most commonly used graph to show frequency distributions. This helpful data collection and analysis tool is considered one of the seven basic quality tools.**

Histograms and box plots are graphical representations for the frequency of numeric data values. They aim to describe the data and explore the central tendency and variability before using advanced statistical analysis techniques. In this article, we will further discuss the similarities and differences between these two tools.

a): Both histograms and box plots allow to visually assess the central tendency, the amount of variation in the data as well as the presence of gaps, outliers or unusual data points.

b): Both histograms and box plots are used to explore and present the data in an easy and understandable manner. Histograms are preferred to determine the underlying probability distribution of a data. Box plots on the other hand are more useful when comparing between several data sets. They are less detailed than histograms and take up less space.

You can use histograms and box plots to verify whether an improvement has been achieved by exploring the data before and after the improvement initiative. Both tools can be helpful to identify whether variability is within specification limits, whether the process is capable, and whether there is a shift in the process over time.

**11. How to select metrics?**

**ANS:**  The five easy steps listed below will enable them to systematically arrive at the appropriate metrics.

**Step 1** Why is the measurement required?

**Step 2** What needs to be measured?

**Step 3** What is the precision of measurement required?

**Step 4** How will it be measured?

**Step 5** What use will the measurement be put to? By whom?

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

**ANS: statistical significance:**

In regards to business, statistical significance is important because it helps you know that the changes you've implemented can be positively attributed to various metrics.

For example, if you've recently implemented a new application to help your office work more efficiently, statistical significance provides you with the confidence in knowing that it made a positive impact on your company's overall workflow. That is, the app's impact was statistically significant and provided value. If it turns out the app wasn't statistically significant, this means your business dollars and the app are at risk. Make sure to measure the statistical significance for every result to get a more comprehensive calculation and result.

Calculating the statistical significance is rather extensive if you calculate it by hand and this is why it's typically calculated using a calculator. When you calculate it by hand, however, it will help you more fully understand the concept. Here are the steps for calculating statistical significance:

1. Create a null hypothesis.
2. Create an alternative hypothesis.
3. Determine the significance level.
4. Decide on the type of test you'll use.
5. Perform a power analysis to find out your sample size.
6. Calculate the standard deviation.
7. Use the standard error formula.
8. Determine the t-score.
9. Find the degrees of freedom.
10. Use a t-Table

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

**ANS:** Many random variables have distributions that are asymptoticallyGaussian but may be significantly non-Gaussian for small numbers. For example the Poisson Distribution, which describe the number of unlikely events occurring after providing a sufficient opportunity for a few events to occur. It is pretty non-Gaussian unless the mean number of events is very large. The mathematical form of the distribution is still Poisson, but a histogram of the number of events after many trials with a large average number of events eventually looks fairly Gaussian.

What generally never becomes Gaussian, however, is the Uniform Distribution. A histogram of the stars’ right ascensions or declinations (the azimuthal and elevation angles used in astronomy) looks a lot like a step function, i.e., flat within the image boundaries. The positions are not uniformly spaced, but they are distributed in the same way as a uniformly distributed random variable for any size postage stamp, including the entire image.

Another example is the location of the centers of raindrop ripples on a pond; they are not uniformly spaced in the east-west direction, but they are uniformly distributed.

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.**

**ANS:**  The median is generally considered to be a 'better' measure of central tendency in a skew distribution or one with extreme values, such as one might expect for workers' salaries (a small proportion of individuals have a much higher salary than the majority). The median lies closer to the majority of values, is insensitive to extreme values and exactly half the values are smaller (greater) than the median.

However, in the context of salaries the mean does have the advantage that when multiplied by the number of individuals it gives the total cost of salaries in a given set of workers.

A further advantage is that means can be modelled using general linear models and similar techniques, although these methods rely on approximation to the normal distribution and this may not be satisfied.

The Mean is one of the measures of central tendency which is very sensitive to extreme observations, as a result the Median is preferred since more than 50% of the data have to be outlying before it will be affected. Hence the median is robust and it gives better estimate in highly perturbed data sets.

However, the Median does not use all the information in the data set. The mean does use all the information and it is preferred when the data follows the classical assumptions.

**15. What is the Likelihood?**

**ANS:** Likelihood is the**hypothetical probability that an event that has already occurred would yield a specific outcome**. The concept differs from that of a probability in that a probability refers to the occurrence of future events, while a likelihood refers to past events with known outcomes.

Only the likelihoods associated with the outcomes that actually occurred are used. By sampling the trained neural network, approximations of likelihoods were obtained for design variable values, and, consequently, constraint satisfaction problemsolving capability was achieved.

If likelihoods can be calculated for alternative models, then a likelihoodratio test may be used to discriminate between hypotheses.

The acceptance of the block proposal is verified by using the ratio of likelihoods evaluated at the new and current values, respectively.