

**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 for difficult statistics?
- a) baggyer
  - b) bootstrap**
  - c) jackknife
  - 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**

**Q10 and Q15 are subjective answer type questions, Answer them in your own words briefly.**

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

Answer - Boxplots may also depict values that are far outside of the normal range of responses (referred to as outliers). A histogram is a graphical representation of the spread of data points

11. How to select metrics?

Answer - Classification. This algorithm will predict data type from defined data arrays. For example, it may respond with yes/no/not sure.

Regression. The algorithm will predict some values. For example, weather forecast for tomorrow.

Ranking. The model will predict an order of items.

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

Answer - 1) State the Research Hypothesis.

2) State the Null Hypothesis.

3) Select a probability of error level (alpha level)

4) Select and compute the test for statistical significance.

5) Interpret the results

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

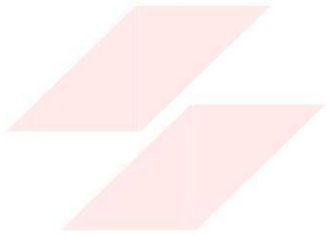
Answer - Exponential distributions do not have a log-normal distribution or a Gaussian distribution. In fact, any type of data that is categorical will not have these distributions as well.

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

Answer - Income is the classic example of when to use the median instead of the mean because its distribution tends to be skewed. The median indicates that half of all incomes fall below 27581, and half are above it. For these data, the mean overestimates where most household incomes fall.

15. What is the Likelihood?

Answer - Likelihood Function in Machine Learning and Data Science is the joint probability distribution(jpd) of the dataset given as a function of the parameter. Think of it as the probability of obtaining the observed data given the parameter values.



**FLIP ROBO**