WORKSHEET

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

2. Which of the following random variable that take on only a countable number of possibilities?

a) Discrete

3. Which of the following function is associated with a continuous random variable?

a) pdf

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

c) mean

5. Which of the following of a random variable is not a measure of spread?

a) variance

6. The _____ of the Chi-squared distribution is twice the degrees of freedom.

a) variance

7. The beta distribution is the default prior for parameters between _____

c) 0 and 1

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

b) bootstrap

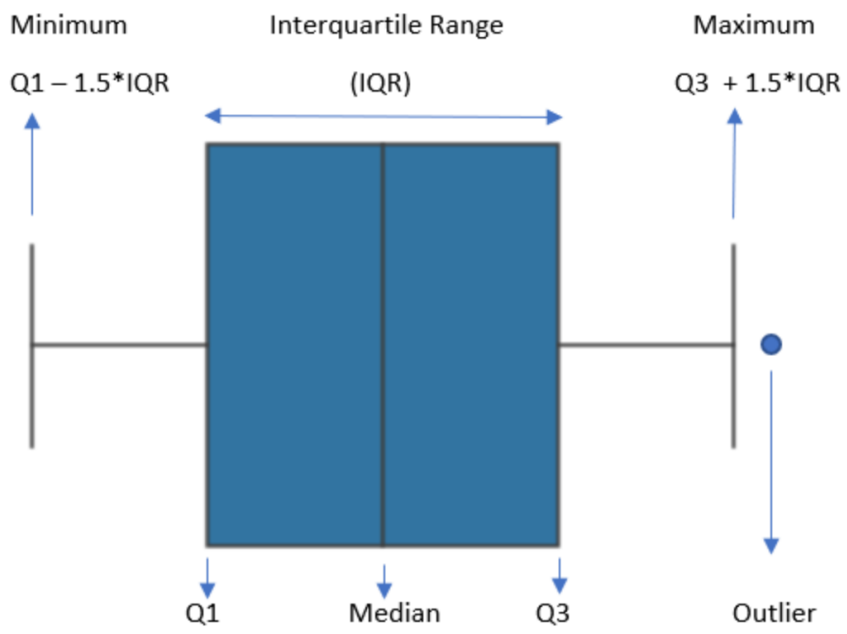
9. Data that summarize all observations in a category are called _____ data.

b) summarized

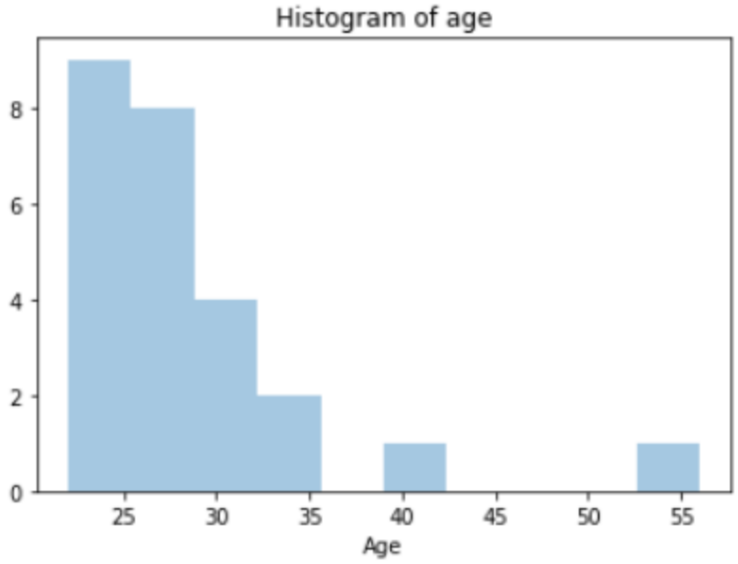
Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

ANS: A boxplot shows the distribution of the data with more detailed information. It shows the outliers more clearly, maximum, minimum, quartile(Q1), third quartile(Q3), interquartile range(IQR), and median. You can calculate the middle 50% from the IQR. Here is the picture:



Histogram takes only one variable from the dataset and shows the frequency of each occurrence. I will use a simple dataset to learn how histogram helps to understand a dataset.



11. How to select metrics?

ANS: Step 1: Articulate Your Goals. This is obvious, but you should always start by defining your goals for your product. ...

Step 2: List the Actions That Matter. ...

Step 3: Define Your Metrics. ...

Step 4: Evaluate your Metrics.

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

ANS: State the Research Hypothesis.

State the Null Hypothesis.

Select a probability of error level (alpha level)

Select and compute the test for statistical significance.

Interpret the results.

13. Give examples of data that doesnot have a Gaussian distribution, nor log-normal.

ANS: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. Example: Duration of a phone car, time until the next earthquake, etc.

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

ANS: Income is the classic example of when to use the median instead of the mean because its distribution tends to be skewed.

15. What is the Likelihood?

ANS: The likelihood is the probability that a particular outcome is observed when the true value of the parameter is , equivalent to the probability mass on ; it is not a probability density over the parameter . The likelihood, , should not be confused with , which is the posterior probability of given the data .