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?
   - The outcome from the roll of a die
   - The outcome of flip of a coin
   - The outcome of exam
   - All of the mentioned

ANS :- All of the mentioned

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

ANS :- Discrete

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

ANS :- pdf

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

ANS :- mean

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

ANS :- variance

6. The__of the Chi-squared distribution is twice the degrees of freedom.
   - variance
   - standard deviation
   - mode
   - none of the mentioned

ANS :- variance

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

- 0 and 10
- 1 and 2
- 0 and 1
- None of the mentioned

ANS :- 0 and 1

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

- baggyer
- bootstrap
- jacknife
- none of the mentioned

ANS :- bootstrap

9. Data that summarize all observations in a category are called_____data.

- frequency
- summarized
- raw
- none of the mentioned

ANS :- 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 :- 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.

11. How to select metrics?

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

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

ANS :- Statistical significance can be accessed using hypothesis testing:
– Stating a null hypothesis which is usually the opposite of what we wish to test (classifiers A and B perform equivalently, Treatment A is equal of treatment B)
– Then, we choose a suitable statistical test and statistics used to reject the null hypothesis
– Also, we choose a critical region for the statistics to lie in that is extreme enough for the null hypothesis to be rejected (p-value)
– We calculate the observed test statistics from the data and check whether it lies in the critical region

13. Give examples of data that doesnot have a Gaussian distribution, nor log-normal.
ANS :- Allocation of wealth among individuals
Values of oil reserves among oil fields (many small ones, a small number of large ones)
Share this:

14. Give an example where the median is a better measure than the mean.
ANS :- Let's say you run a customer satisfaction survey with a sample of 9 and rate their overall satisfaction scores on a scale of 1 to 10. You get an average of 5.22. You know that in general, you tend to retain customers with a score over 3, so you're satisfied, because this indicates that you're still above where you want to be. But then, suddenly, you lose 6 of those 9 customers. You go back to look at your data, and you find these scores:

1, 3, 3, 3, 3, 5, 9, 10, 10

The median of this group is a 3, indicating that at least half of your customers or more were unhappy. The scores became lopsided because of the unexpected 10's, and you missed out on an important part of your data – the midpoint that indicated that as many as half of your customers or more were dissatisfied with your company.

Median can play a major role in things like income level research as well, because a few millionaires may make it look like the socio-economic status of your sample is higher than it really is.

Whenever a graph falls on a normal distribution, using the mean is a good choice. But if your data has extreme scores (such as the difference between a millionaire and someone making 30,000 a year), you will need to look at median, because you'll find a much more representative number for your sample

15. What is the Likelihood?
ANS :- 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. It is formed from the joint probability distribution of the sample, but viewed and used as a function of the parameters only, thus treating the random variables as fixed at the observed values.

FLIP ROBO