# STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1. Bernoulli random variables take (only) the values 1 and 0.
a) True
b) False

***Answer: a) True***

---

2. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?
a) Central Limit Theorem
b) Central Mean Theorem
c) Centroid Limit Theorem
d) All of the mentioned

***Answer: a) Central Limit Theorem***

---

3. Which of the following is incorrect with respect to use of Poisson distribution?
a) Modeling event/time data
b) Modeling bounded count data
c) Modeling contingency tables
d) All of the mentioned

***Answer: b) Modeling bounded count data***

---

4. Point out the correct statement.
a) The exponent of a normally distributed random variables follows what is called the log-normal distribution
b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent
c) The square of a standard normal random variable follows what is called chi-squared distribution
d) All of the mentioned

***Answer: d) All of the mentioned***

---

5. _____ random variables are used to model rates.
a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned

***Answer: c) Poisson***

---

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
a) True
b) False

***Answer: b) False***

---

7. 1. Which of the following testing is concerned with making decisions using data?
a) Probability
b) Hypothesis
c) Causal
d) None of the mentioned

***Answer: b) Hypothesis***

---

8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.
a) 0
b) 5
c) 1
d) 10

***Answer: a) 0***

---

9. Which of the following statement is incorrect with respect to outliers?
a) Outliers can have varying degrees of influence
b) Outliers can be the result of spurious or real processes
c) Outliers cannot conform to the regression relationship
d) None of the mentioned

***Answer: c) Outliers cannot conform to the regression relationship***

---

## WORKSHEET

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

10. What do you understand by the term Normal Distribution?

***Answer:*** Normal distribution is a type of distribution where that has most of the data in the center with decreasing amounts evenly distributed to the left and the right.

99.7 % of the data points lie between the 1st standards deviation.
95.5 % of the data points lie between the 2st standards deviation.
68.3 % of the data points lie between the 3rd standards deviation.

---

11. How do you handle missing data? What imputation techniques do you recommend?

***Answer:*** We can handle the missing data by the below masures/methods:

a. Mean/Median/Mode
b. Ignore the records with missing values (If not much data information is lost)
c. Predict missing values (Linear Regression, Random forest etc)

---

12. What is A/B testing?

***Answer:*** A/B testing is means by which users are randomly given two or more experiences. This is used to test success of variation between two things.

---

13. Is mean imputation of missing data acceptable practice?

***Answer:*** It depends on case to case. For example, we the column which has age, mean can be used to replace the null/missing values. This will not affect much on the trend. While if the missing/null values are in the salary column then taking mean as measure to replcae null values can represent the trend wrongly. for example CEO/Director of a company can have big amounted salary while a fresher/intern can have comparatively very less salary.

---

14. What is linear regression in statistics?

***Answer:*** A simple linear regression model consists of 2 variable that is dependent and independent variable. We find out the behaviour of the dependent variable from independent variable by studying the pattern/trends of the data points.

simple linear regression equation : $Y = mX + c$

Y = Independent variable
X = Dependent variable
m = Slope (Y/X)
c = intercpet made by the line on Y-axis

---

15. What are the various branches of statistics?

***Answer:*** There are mainly 2 branches of statistics as below:

a. Descriptive statistics: This branch deals with collection and describing the data. Ex - Mean, Median, Mode.
b. Inferential statistics: This branch deals with making inferences from the dataset.

---