

## STATISTICS WORKSHEET-1 ANSWERS

**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. Usually replacing the standard error by its estimated value does change the CLT.**

- a) True
- b) False

**ANSWER: b) False**

**7. 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. 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**

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

**ANSWER:**

A normal distribution, sometimes called the bell curve, is a distribution that occurs naturally in many situations. The bell curve is symmetrical. Half of the data will fall to the left of the mean; half will fall to the right. The mean, mode and median are all equal. The curve is symmetric at the center (i.e. around the mean,  $\mu$ ). The total area under the curve is 1.

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

**ANSWER:**

When data is missing at random, we can use list-wise or pair-wise deletion of the missing observations. However, there are many reasons why it's not possible to delete the data:

- Data size may be small and there may not be enough observations with non-missing data to produce a reliable analysis
- In predictive analytics, missing data can prevent the predictions for those observations which have missing data
- External factors may require specific observations to be part of the analysis

In such cases, we impute values for missing data.

A common technique is to use the mean or median of the non-missing observations. This can be useful in cases where the number of missing observations is low. However, for large number of missing values, using mean or median can result in loss of variation in data and it is better to use other imputations. Depending upon the nature of the missing data, we use different techniques to impute data.

**12. What is A/B testing?**

**ANSWER:**

A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

For instance, let's say you own a fruits shop and want to increase the sales of your fruits. Here, either you can use random experiments, or you can apply scientific and statistical methods. A/B testing is one of the most prominent and widely used statistical tools.

In the above scenario, you may divide the products into two parts – A and B. Here A will remain unchanged while you make significant changes in B's packaging. Now, on the basis of the response from customer groups who used A and B respectively, you try to decide which is performing better.

It's a hypothetical testing methodology for making decisions that estimate population parameters based on sample statistics.

### **13. Is mean imputation of missing data acceptable practice?**

#### **ANSWER:**

The mean or median imputation of the non-missing observations can be useful in cases where the number of missing observations is low. However, for large number of missing values, using mean or median can result in loss of variation in data and it is better to use other imputation techniques.

### **14. What is linear regression in statistics?**

#### **ANSWER:**

Linear regression is a basic and commonly used type of predictive analysis. The main idea of regression is to examine two things:

- (1) does a set of independent variables do a good job in predicting an outcome/ dependent variable?
- (2) Which variables are significant predictors of the outcome variable, and in what way do they impact the outcome variable?

These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables.

The simplest form of the regression equation with one dependent and one independent variable is defined by the formula

$$y = c + b \cdot x,$$

where  $y$  = estimated dependent variable score,

$c$  = constant,

$b$  = regression coefficient, and

$x$  = score on the independent variable.

### **15. What are the various branches of statistics?**

#### **ANSWER:**

The two fundamental parts of statistics are descriptive statistics and inferential statistics.

Descriptive statistics manages the presentation and collection of information or data. This is typically the initial segment of statistical analysis. Descriptive statistics manages the collection of information, its presentation in different forms, like tables, charts, and outlines, and discovering averages, deviations and different measures which would help in analysing the data.

Inferential statistics includes reaching the right inferences from the statistical analysis that has been performed using the descriptive statistics. Eventually, the summarizing of data makes a significant aspect to study inferential statistics. Most forecasts of things to happen and analyse about a population occur by contemplating a fair sample of the population and deciding how it behaves. And then making assumptions based on this for the entire population. Inferential statistics deals with techniques used for the analysis of data, making estimates, and drawing conclusions from limited information obtained through sampling and testing the reliability of the estimates.