

# STATISTICS WORKSHEET-1

## ANSWERS

1. True
2. Central limit theorem
3. Modeling bounded count data
4. All of the mentioned
5. Poisson
6. False
7. Hypothesis
8. 0
9. Outliers cannot conform to the regression relationship

### 10. NORMAL DISTRIBUTION

Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve.

#### KEY TAKEAWAYS

- A normal distribution is the proper term for a probability bell curve.
- In a normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis of 3.
- Normal distributions are symmetrical, but not all symmetrical distributions are normal.
- In reality, most pricing distributions are not perfectly normal.

11.

### 12. A/B Testing

At its most basic, A/B testing, also known as split testing, is a way to compare different versions of something to see which performs better. In these experiments, you define a conversion goal to measure, like clicks or completed transactions. Two variations of the same marketing asset (like a web page or email) are then shown to different users at random while measuring the difference in performance.

13. The process of replacing null values in a data collection with the data's mean is known as mean imputation. Mean imputation is typically considered terrible practice since it ignores feature correlation.  
mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower
14. Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things: (1) does a set of predictor variables do a

good job in predicting an outcome (dependent) variable? (2) Which variables in particular are significant predictors of the outcome variable, and in what way do they—indicated by the magnitude and sign of the beta estimates—impact the outcome variable? These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. .

## **15. BRANCHES OF STATISTICS**

The two main branches of statistics are **descriptive statistics** and **inferential statistics**.

Descriptive statistics implies a simple quantitative summary of a data set that has been collected. It helps us understand the experiment or data set in detail and tells us everything we need to put the data in perspective.

Inferential statistics, unlike **descriptive statistics**, is the attempt to apply the conclusions that have been obtained from one experimental study to more general populations. This means inferential statistics tries to answer questions about populations and samples that have not been tested in the given experiment.