ANSWER KEY

1. A – TRUE
2. A – CENTRAL LIMIT THEOREM
3. A – MODELLING EVENT/TIME DATA
4. D – ALL OF THE MENTIONED
5. C – POISSON
6. B – FALSE
7. B – HYPOTHESIS TESTING
8. A – 0
9. D – NONE OF THE MENTIONED

10. Normal Distribution is a type of continuous probability distribution which is symmetric about its mean and forms a bell curve. Values closer to the mean have a higher probability and values to the tails of the curve have lower.

11. There are multiple ways to deal with missing data. The easiest one is completely removing all rows containing missing data points. We can also replace the missing data values with mean/median/modal values of the entire data column. Time series data can be dealt with using Last Observation Carried Forward (LOCF) or Next Observation Carried Backward (NOCB). Some recommended imputation techniques are:
    i) Frequency Category imputation: replacing missing values with the categorical value of highest frequency.
    ii) Arbitrary Value Imputation: Adding arbitrary values like 'missing' or 999. May skew the data.
    iii) LOCF/NOCB: For time series data

12. A/B Testing is a randomised experiment of hypothesis testing comparing two variants (A and B) of the same variable. This is done by testing a subject's response to variant A against B to determine which is more effective. A/B testing is used to understand user's preference between the choices. For example: an e-commerce site might have the 'Add to Cart' option on the left(A) and right(B) side of the page and the user might like the left placement better than the right one. The website might then determine A is probably better than B.

13. Mean Imputation has a few drawbacks that could make it a bad practice. First, it doesn't take into account the relationship between variables. In imputing data in a table of weight according to height with mean weight 60kg, a man of height 6 feet 4 inches missing his expected weight datapoint might end up with a value of around 60kg even though the recommended value is around 85kg. Second, it also reduces the variance in the sample by creating a datapoint equal to mean.

14. Linear regression is the modelling of a relationship between two variables using a linear equation($y=mx+c$). On a graph we try to fit data points to a straight line describing the linear equation. It can be used for quantifying the strength of the relationship via Pearson's rank correlation and for error reduction.

15. There are two types of statistics:
    i) Descriptive Statistics: Describes important characteristics of data using central tendencies (mean, median, and mode) and measures of dispersion (standard deviation, variance). Distribution of data is shown using graphs and charts.
    ii) Inferential Statistics: Used to draw conclusions about the characteristics of a population from a sample and decide the reliability of those conclusions. Can calculate the probability that the statistics of the sample provide an accurate picture of the corresponding parameters of the population.