FLIP ROBO TECHNOLOGIES WORK SHEET for STATISTICS

1st Set: Questions with single answers

Questions	Answers
Q1	A) True
Q2	A) Central Limit Theorem
Q3	B) Modelling bounded count data
Q4	D) All of the mentioned
Q5	C) Poisson
Q6	B) False
Q7	B) Hypothesis
Q8	A) 0
Q9	C) Outliers cannot conform to the regression relationship

2nd Set: Questions with subjective answers

Question 10:

Answer: Normal Distribution is a probability distribution that is symmetric around the mean, showing data frequency near the mean is higher. Graphically, the normal distribution appears as a 'Bell Curve'. In an ideal case scenario, the normal distribution can also define a Symmetrical Distribution, when where a dividing line produces two mirror images.

Question 11:

Answer: Missing data can be dealt in various ways. The most common reaction is to ignore it. Choosing to make no decision, on the other hand, indicates that our statistical programme will make the decision for us, that is removing things in a listwise sequence most of the time. Depending on why and how much data is gone, listwise deletion may or may not be a good idea.

Another common strategy is imputation. Imputation is the process of substituting an estimate for missing values and analysing the entire data set as if the imputed values were the true observed values.

There are number of imputation methods for missing data dealing like Mean/Median Imputation, Multivariate Imputation by Chained Equations (MICE), Random Forest using decision trees to estimate missing values, impact is different basis on the size of the dataset. For example Random forest works best on huge dataset and if used on a small dataset, increases the chances of overfitting

Question 12:

Answer: A/B testing, or split testing, refers to a randomized experimentation process wherein two versions of a variable shown to different segments of visitors at the same time to determine which version leaves the maximum impact and drives business metrics. In other words, it is like testing whether version A or version B, that is more appealing.

The concept is like the scientific method. If you want to find out what happens when you change one thing, you must create a situation where only that one thing changes.

Think about the experiments we conducted in elementary school. If we put 2 seeds in 2 cups of dirt and put one in the closet and the other by the window, we will see different results. This kind of experimental setup is A/B testing.

Question 13:

Answer: Mean imputation is considered to be a terrible practice for missing data. The reason being:

1. It does not preserve the relationship among variables
2. It leads to a reduction of variance of the data

Question 14:

Answer: Linear regression models the relationships between at least one explanatory variable and an outcome variable. These variables are known as the independent and dependent variables, respectively. When there is one Independent Variable, the procedure is known as simple linear regression. Where there are more Independent Variables, we call it multiple regression.

Suppose we use linear regression to model how the outside temperature in Celsius and Insulation thickness in centimetres, our two independent variables, relate to air conditioning costs in dollars (dependent variable).

Let's interpret the results for the following multiple linear regression equation:

Air Conditioning Costs (in INR) = 2 * Temperature (in Celsius) – 1.5 * Insulation (in cm)

The coefficient sign for Temperature is positive (+2), which indicates a positive relationship between Temperature and Costs. As the temperature increases, so does air condition costs. More specifically, the coefficient value of 2 indicates that for every 1 Celsius increase, the average air conditioning cost increases by two Rupees.

On the other hand, the negative coefficient for insulation (–1.5) represents a negative relationship between insulation and air conditioning costs. As insulation thickness increases, air conditioning costs decrease. For every 1 cm increase, the average air conditioning cost drops by INR 1.50.

We can also enter values for temperature and insulation into this linear regression equation to predict the mean air conditioning cost.

Question 15:

Answer: The two main branches of statistics are descriptive statistics and inferential statistics. Both of these are employed in scientific analysis of data, and both are equally important.

Descriptive statistics deals with the presentation and collection of data. Whereas, Inferential statistics, as the name suggests, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics.