STATISTICS WORKSHEET

1. True
2. Central Limit Theorem
3. Modelling 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, 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.

11

An analysis is only as good as its data, and every researcher has struggled with dubious results because of missing data. In this article, I will cover three ways to deal with missing data.

Understanding the nature of missing data is critical in determining what treatments can be applied to overcome the lack of data. Data can be missing in the following ways:

- **Missing Completely At Random (MCAR):** When missing values are randomly distributed across all observations, then we consider the data to be missing completely at random. A quick check for this is to compare two parts of data – one with missing observations and the other without missing observations. On a t-test, if we do not find any difference in means between the two samples of data, we can assume the data to be MCAR.
- **Missing At Random (MAR):** The key difference between MCAR and MAR is that under MAR the data is not missing randomly across all observations, but is missing randomly only within sub-samples of data. For example, if high school GPA data is missing randomly across all schools in a district, that data will be considered MCAR. However, if data is randomly missing for students in specific schools of the district, then the data is MAR.
- **Not Missing At Random (NMAR):** When the missing data has a structure to it, we cannot treat it as missing at random. In the above example, if the data was missing for all students from specific schools, then the data cannot be treated as MAR.

12

STATISTICS WORKSHEET

A/B testing is a user experience research methodology. A/B tests consist of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing or "two-sample hypothesis testing" as used in the field of statistics

13

True, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased. That's a good thing.

Plus, by imputing the mean, you are able to keep your sample size up to the full sample size. That's good too.

This is the original logic involved in mean imputation.

Mean **reduces a variance of the data**

As we can see, the variance was reduced (that big change is because the dataset is very small) after using the Mean Imputation. Going deeper into mathematics, a smaller variance leads to the narrower confidence interval in the probability distribution[

14

In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables. The case of one explanatory variable is called simple linear regression; for more than one, the process is called multiple linear regression.

15

Descriptive Statistics

Descriptive statistics deals with the presentation and collection of data. This is usually the first part of a statistical analysis. It is usually not as simple as it sounds, and the statistician needs to be aware of designing experiments, choosing the right focus group and avoid biases that are so easy to creep into the experiment.

Inferential Statistics

Inferential statistics, as the name suggests, involves drawing the right conclusions from the statistical analysis that has been performed using

STATISTICS WORKSHEET

descriptive statistics. In the end, it is the inferences that make studies important and this aspect is dealt with in inferential statistics.