

Statistics– WORKSHEET 1

1. A
2. A
3. B
4. D
5. C
6. B
7. B
8. A
9. B
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. **Listwise Deletion:** Delete all data from any participant with missing values. If sample is large enough, then you likely can drop data without substantial loss of statistical power.
Educated Guessing: It sounds arbitrary and isn't preferred course of action, but we can often infer a missing value. For related questions, for example, like those often presented in a matrix, if the participant responds with all "4s", assume that the missing value is a 4.
Average Imputation: Use the average value of the responses from the other participants to fill in the missing value. If the average of the 30 responses on the question is a 4.1, use a 4.1 as the imputed value. This choice is not always recommended because it can artificially reduce the variability of your data but in some cases makes sense.
Common-Point Imputation: For a rating scale, using the middle point or most commonly chosen value. For example, on a five-point scale, substitute a 3, the midpoint, or a 4, the most common value (in many cases). This is a bit more structured than guessing, but it's still among the riskier options. Use caution unless you have good reason and data to support using the substitute value.
Regression Substitution: You can use multiple-regression analysis to estimate a missing value. We use this technique to deal with missing SUS scores. Regression substitution predicts the missing value from the other values. In the case of missing SUS data.
Multiple Imputation: The most sophisticated and, currently, most popular approach is to take the regression idea further and take advantage of correlations between responses. In multiple imputation, software creates plausible values based on the correlations for the missing data and then averages the simulated datasets by incorporating random errors in predictions.
12. A/B testing (also known as split testing) is a process of showing two variants of the same web page to different segments of website visitors at the same time and comparing which variant drives more conversions.
13. Bad practice in general, If just estimating means: mean imputation preserves the mean of the observed data, Leads to an underestimate of the standard deviation, Distorts relationships between variables by "pulling" estimates of the correlation toward zero
14. In statistics, linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression.
15. **Descriptive statistics** - It organizes raw data into meaningful information. An house hold articles manufacturing company would like to know what people feel about their products.

Statistics– WORKSHEET 1

For that purpose, the company forms a team of people and tries to collect information from the public. The team of people formed by the company is trying to collect data from the public directly. The data which is being collected directly from the public will always not be meaning full. Hence, the data which is being collected directly from the public has to be converted in to meaningful information.

Inferential statistics: It analyses sample data to draw conclusion about population. It analyses sample data to draw conclusion about population. Marketing research team of a company wants to know how far the people need a particular product manufactured by the company. There is one hundred thousand population in a particular city. It is bit difficult to go and ask all one hundred thousand people, due to time consumption and other factors. Hence, it takes a sample of 1000 people to draw conclusion for the whole population. That is making general statement from the study of particular cases or any treatment of data, which leads to prediction or inference concerning a larger group of data.