

## STATISTICS WORKSHEET

1. A

2. D

3. D

4. D

5. B

6. B

7. A

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.

11. a) Replacing With Arbitrary Value. ...

b) Replacing With Mode. ...

c) Replacing With Median. ...

d) Replacing with previous value – Forward fill. ...

e) Replacing with next value – Backward fill. ...

f) Interpolation. ...

g) Impute the Most Frequent Value.

12. A/B testing is a type of experiment in which you split your web traffic or user base into two groups, and show two different versions of a web page, app, email, and so on, with the goal of comparing the results to find the more successful version.

13. Mean imputation is typically considered terrible practice since it ignores feature correlation.

14. Linear regression is the most basic and commonly used predictive analysis. Regression estimates are used to describe data and to explain the relationship.

15. There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.