Answers

1. True
2. A
3. B
4. D
5. C
6. False
7. B
8. A
9. C

10. A normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme. The middle of the range is also known as the mean of the distribution.

11. When dealing with missing data, data scientists can use two primary methods to solve the error: imputation or the removal of data. The imputation method develops reasonable guesses for missing data. It's most useful when the percentage of missing data is low.

- Mean imputation. ...
- Substitution. ...
- Hot deck imputation. ...
- Cold deck imputation. ...
- Regression imputation

12. When you run an A/B test, you compare one page against one or more variations that contain one major difference in an element of the control page. After a set amount of time, or visits, you compare the results to how the change affected your results. A/B testing is actually a nickname for a variety of testing that involves single element changes across multiple variations, so it might actually be A/B/C/D testing, but it is still called A/B testing and if you see A/B/n or Split Test that is just a more accurate expression of one control with many variations.

13. The process of replacing null values in a data collection with the data's mean is known as mean imputation.

Mean imputation is typically considered terrible practice since it ignores feature correlation. Mean imputation decreases the variance of our data while increasing bias.

Mean imputation reduces the variance of the imputed variables. Mean imputation shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence interval.

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