Answer 1: Option A

Answer 2: Option A

Answer 3: Option B

Answer 4: Option D

Answer 5: Option C

Answer 6: Option B

Answer 7: Option B

Answer 8: Option A

Answer 9: Option C

Answer 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

Answer 11: Generally speaking, there are three main approaches to handle missing data: **(1) Imputation**—where values are filled in the place of missing data, **(2) omission**—where samples with invalid data are discarded from further analysis and **(3) analysis**—by directly applying methods unaffected by the missing values. A common technique is to use the **mean or median** of the non-missing observations

Answer 12: A/B testing is one of the most popular controlled experiments used to optimize web marketing strategies. It allows decision makers to choose the best design for a website by looking at the analytics results obtained with two possible alternatives A and B.

Answer 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

Answer 14: Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. A linear regression line has an equation of the form **Y = a + bX,** where **X** is the explanatory variable and **Y** is the dependent variable. The slope of the line is **b**, and **a** is the intercept (the value of **y** when **x = 0**).

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