

1. **Bernoulli random variables take (only) the values 1 and 0.**

Ans= True

2. **Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?**

Ans=Centroid Limit Theorem.

3. **Which of the following is incorrect with respect to use of Poisson distribution?**

Ans=Modeling bounded count data.

4. **Point out the correct statement.**

Ans=All of the mentioned

5. **_____ random variables are used to model rates.**

Ans=Poisson

6. **10. Usually replacing the standard error by its estimated value does change the CLT.**

Ans=False

7. **1. Which of the following testing is concerned with making decisions using data?**

Ans=Hypothesis

8. **4. Normalized data are centered at _____ and have units equal to standard deviations of the original data.**

Ans=0

9. **Which of the following statement is incorrect with respect to outliers?**

Ans=Outliers cannot conform to the regression relationship

10. **What do you understand by the term Normal Distribution?**

Ans= 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. **How do you handle missing data? What imputation techniques do you recommend?**

Ans=There are some ways to handle it.

1. **Listwise Deletion:** Delete all data from any participant with missing values. If your sample is large enough, then you likely can drop data without substantial loss of statistical power. Be sure that the values are missing at random and that you are not inadvertently removing a class of participants.
2. **Recover the Values:** You can sometimes contact the participants and ask them to fill out the missing values. For in-person studies, we've found having an additional check for missing values before the participant leaves helps.

Imputation

•

Educated Guessing: It sounds arbitrary and isn't your preferred course of action, but you 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.

Please rate your level of agreement to the following statements about the Chipotle website.

	Strongly Disagree 1	2	3	4	Strongly Agree 5
The Chipotle website has a clean and simple presentation.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
The Chipotle website is easy to use.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
I find the Chipotle website to be attractive.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
The Chipotle website is trustworthy.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
It is easy to navigate within the Chipotle website.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The Chipotle website's capabilities meet my requirements.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
The information on the Chipotle website is credible.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
I will likely return to the Chipotle website in the future.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

- **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 more risky 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, we had enough data to create stable regression equations and predict the missing values [automatically in the calculator](#).
 - **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](#) [pdf], software creates plausible values based on the correlations for the missing data and then averages the simulated datasets by incorporating random errors in your predictions. It is one of a number of examples where computers continue to change the statistical landscape. Most statistical packages like SPSS come with a multiple-imputation feature.

12. What is A/B testing?

Ans= A/B testing (also known as bucket testing or split-run 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. Is mean imputation of missing data acceptable practice?

Ans= Mean imputation is the practice of replacing null values in a data set with the mean of the data. Mean imputation is generally bad practice because it doesn't take into account feature correlation.

14. What is linear regression in statistics?

Ans=The measure of the extent of the relationship between two variables is shown by the **correlation coefficient**. The range of this coefficient lies between -1 to +1. This coefficient shows the strength of the association of the observed data for two variables.

A linear regression line equation is written in the form of:

$$Y = a + bX$$

where X is the independent variable and plotted along the x-axis

Y is the dependent variable and plotted along the y-axis

The slope of the line is b, and a is the intercept (the value of y when x = 0).

15. What are the various branches of statistics?

Ans=The two main branches of statistics are [descriptive statistics](#) and [inferential statistics](#).

a. 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](#).

b. Inferential Statistics

[Inferential statistics](#), as the name suggests, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics. In the end, it is the inferences that make studies important and this aspect is dealt with in inferential statistics.