STATISTICS WORKSHEET1

1. (a)

2. (a)

3. (b)

4. (d)

5. (c)

6. (b)

7. (b)

8. (a)

9. (c)

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

The normal distribution, also known as the Gaussian or standard normal distribution, is the continuous probability distribution that plots all of its values in a symmetrical fashion, and most of the results are situated near the central peak. Values are equally likely to plot either above or below the mean. The normal distribution is a probability distribution that (roughly) describes many common datasets in the real world. It is the most common type of distribution, and it arises naturally in statistics through random sampling techniques.
Nowadays, it is more common to show up as a model for the "lifespan" of a product, like a bulb, or the outcome of standardized tests, like IQ. Biological measurements, like height or weight are often estimated with normal distribution.

11. How do you handle missing data? What imputation techniques do you recommend?

Missing data can be dealt in a variety of ways. One is list wise deletion. Another common strategy to handle missing data is imputation. Imputation is the process of substituting an estimate for missing values and analysing the entire data set as if the imputed values were the true observed values.

The most prevalent methods to estimate missing values are

Mean imputation

 Mean imputation is the replacement of a missing observation with the mean of the non-missing observations for that variable but it has a slew of drawbacks

Hot deck imputation

Selecting all the sample participants who are comparable on other factors, then choose one of their missing variable values at random.

Cold deck imputation

A value picked deliberately from an individual with similar values on other variables. In most aspects, this is comparable to Hot Deck, but without the random variance. As an example, under the same experimental condition and block, you can always select the third individual.

Regression imputation

With regression imputation the information of other variables is used to predict the missing values in a variable by using a regression model. Commonly, first the regression model is estimated in the observed data and subsequently using the regression weights the missing values are predicted and replaced.

Stochastic regression imputation

The predicted value of a regression plus a random residual value. This has all of the benefits of regression imputation plus the random component's benefits. The majority of multiple imputation is based on stochastic regression imputation.

Interpolation and extrapolation

An estimate based on other observations made by the same person. It generally only works with data that is collected over time. Proceed with caution, though. For a variable like height in children–one that cannot be reduced through time–interpolation would make more sense. Extrapolation entails estimating beyond the data's true range, which necessitates making more assumptions than is necessary.

Single or Multiple imputation

Single imputation usually identifies a particular record for a subject, e.g. baseline or just the previous non-missing value and repeats it for the missing data points. Multiple imputation uses a predicted value for a given subject and time point using statistical modelling of available data.

12. What is A/B testing?

A/B testing, also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics.

Essentially, A/B testing eliminates all the guesswork out of website optimization and enables experience optimizers to make data-backed decisions. In A/B testing, A refers to 'control' or the original testing variable. Whereas B refers to 'variation' or a new version of the original testing variable.

The version that moves your business metric(s) in the positive direction is known as the 'winner.' Implementing the changes of this winning variation on your tested page(s) / element(s) can help optimize your website and increase business ROI.

The metrics for conversion are unique to each website. For instance, in the case of ecommerce, it may be the sale of the products. Meanwhile, for B2B, it may be the generation of qualified leads.

A/B testing is one of the components of the overarching process of Conversion Rate Optimization (CRO), using which you can gather both qualitative and quantitative user insights. You can further use this collected data to understand user behavior, engagement rate, pain points, and even satisfaction with website features, including new features, revamped page sections, etc. If you're not A/B testing your website, you're surely losing out on a lot of potential business revenue.

13. Is mean imputation of missing data acceptable practice?

No, mean imputation has a lot of drawbacks.

1. Mean imputation does not preserve the relationships among variables.

2. Mean Imputation Leads to an Underestimate of Standard Errors

14. What is linear regression in statistics?

Linear regression analysis is used to predict the value of a variable based on the value of another variable. In statistics, linear regression is an approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variable) denoted X. The case of one explanatory variable is called simple linear regression.

15. What are the various branches of statistics?

There are two branches of statistics: descriptive statistics and inferential statistics

Inferential Statistics:
Inferential statistics used to make inference and describe about the population. These stats are more useful when it is not easy or possible to examine each member of the population.

Descriptive Statistics:
Descriptive statistics are use to get a brief summary of data. You can have the summary of data in numerical or graphical form.