

STATISTICS

1. a) True.
2. a) Central limit theorem.
3. b) Modeling bounded count data.
4. d) All of the above.
5. c) Poission.
6. b) False.
7. b) Hypothesis.
8. a) 0(Zero).
9. d) None of mentioned.
10. In Statistics, a normal distribution or Gaussian distribution is a type of continuous probability distribution for a real-valued random variable. Their importance is partly due to the central limit theorem. It states that, under some conditions, the average of many samples (observations) of a random variable with finite mean and variance is itself a random variable, whose distribution converges to a normal distribution as the number of samples increases. Therefore, physical quantities that are expected to be the sum of many independent processes, such as measurement errors, often have distributions that are nearly normal.
A normal distribution is sometimes informally called a bell curve. However, many other distributions are bell-shaped (such as the Cauchy, Student's t, and logistic distributions).

11. Missing data can be dealt with in a variety of ways. The most common reaction is to ignore it. Choosing to make no decision, on the other hand, indicates that your statistical programme will make the decision for you. Your application will remove things in a listwise sequence most of the time. Depending on why and how much data is gone, listwise deletion may or may not be a good idea. Another common strategy among those who pay attention 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. And how would you choose that estimate? The following are some of the most prevalent methods:

Mean imputation:

Calculate the mean of the observed values for that variable for all non-missing people. It has the advantage of maintaining the same mean and sample size, but it also has a slew of drawbacks. Almost all of the methods described below are superior to mean imputation.

Substitution:

Assume the value from a new person who was not included in the sample. To put it another way, pick a new subject and employ their worth instead.

Hot deck imputation:

A value picked at random from a sample member who has comparable values on other variables. To put it

another way, select 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:

The result of regressing the missing variable on other factors to get a predicted value. As a result, instead of utilising the mean, you're relying on the anticipated value, which is influenced by other factors. This keeps the associations between the variables in the imputation model, but not the variability around the anticipated values.

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 and multiple imputation are the two forms of imputation. When people say imputation, they usually mean single.
- The term "single" refers to the fact that you only use one of the seven methods to estimate the missing number outlined above.
- It's popular since it's simple to understand and generates a sample with the same number of observations as the complete data set.
- When listwise deletion eliminates a considerable amount of the data set, single imputation appears to be a tempting option. It does, however, have certain restrictions.
- Unless the data is Missing Completely at Random, certain imputation processes, such as means, correlations, and regression coefficients, result in skewed parameter estimations. The bias is frequently worse than with listwise deletion, which is most software's default.
- The level of the bias is determined by a number of factors, including the imputation technique, the missing

data mechanism, the fraction of missing data, and the information in the data set.

12. 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.

13.

- Bad practice in general
- If just estimating means: mean imputation preserves the mean of the observed data
- Leads to an underestimate of the standard deviation
- Distorts relationships between variables by "pulling" estimates of the correlation toward zero

14. Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things:
1. Does a set of predictor variables do a good job in predicting an outcome variable?
 2. Which variables in particular are significant predictors of the outcome variable?

The simplest form of the regression equation with one dependent and one independent variable is defined by the formula $y=c+b*x$, where y =estimated dependent variable score, c =constant, b =regression coefficient, and x =score on the independent variable.

15. The two main branches of Statistics are Descriptive statistics and Inferential statistics.

Descriptive statistics: It 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.

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.