

STA Past-Papers Notes

Write short note on the following:

1. **Simple Random Sampling:** A simple random sample of size n element is a sample drawn in such a way that every combination of n element has an equal chance of being the sample selected. The probability is $1/n$
2. **Stratified Sampling:** In stratified random sampling the population is classified into actually exclusive sub-groups or strata and probability samples are drawn independently from each of these strata. A sample from each strata maybe obtained by simple random sampling.
3. **Cluster Sampling:** this is a technique in which the population is subdivided into groups or clusters, then a probability sample of this clusters is drawn and studied. The samples in a clusters do not necessarily have homogeneity with each other.
4. **Systematic Sampling:** you select every k th elements of the population until you have all the sample in the population k can be any number. Thus if you wish to select a sample of 100 persons from total population of 10,000, you would take every hundredth individual ($K=N/n = 10,000/100 = 100$).

What is the null and alternative hypotheses:

Null hypotheses: A claim that there is no effect in the population

Alternative hypotheses: A claim that there is an effect in the population

What is a simple hypotheses:

A hypothesis that, when true, completely specifies the population distribution is called a *simple* hypothesis; one that does not is called a *composite* hypothesis.

A simple hypothesis is one which specifies a single specific prediction about the relationship between two variables

A composite hypothesis is made up of multiple simple hypothesis that are related to each other.

What are type I and II errors:

A type I error (false-positive) occurs if an investigator **rejects** a null hypothesis that is actually **true** in the population;

a type II error (false-negative) occurs if the investigator fails to reject a null hypothesis that is actually **false** in the population.

Difference between correlation and regression:

Correlation quantifies the strength of the linear relationship between a pair of variables, whereas regression expresses the relationship in the form of an equation.

Differentiate between One and Two tailed test

In a one-tailed test, the alternative hypothesis specifies the direction of the effect.

In a two-tailed test, the alternative hypothesis does not specify the direction of the effect. Instead, it only states that the effect is different from the null hypothesis.

For example, in the drug example above, the alternative hypothesis might be that the drug will have any effect on blood pressure (either increase or decrease). This would be a two-tailed test because the alternative hypothesis does not specify the direction of the effect.

For example, if we are testing the effect of a new drug on blood pressure, the alternative hypothesis might be that the drug will lower blood pressure, but not that it will increase blood pressure. This would be a one-tailed test because we are only interested in the direction of the effect that is specified in the alternative hypothesis.

Standard error of mean is a measure of **precision**

Standard deviation of a sampling distribution is called the **standard error**

As a general rule, researchers tend to use **95%** confidence interval

Standard error of mean represents the variability of the **mean** itself

With samples less than 30, **standard error** of mean is an estimate of population standard deviation

With samples more than 30, **sample standard deviation** is an estimate of population standard deviation.

The confidence intervals are quoted as a **mean** and **range**.

List and explain any four properties of a good estimator

1. **Unbiasedness:** The average of the estimator's results over many samples should be close to the true population parameter.
2. **Consistency:** A consistent estimator is one that becomes increasingly accurate as the sample size increases.
3. **Efficiency:** In other words, it provides more precise estimates with the same amount of data.
4. **Sufficiency:** This means that no other estimator based on the same sample data can provide more information about the population parameter

Define Regression analysis: The study of the dependency of a continuous independent variable on a continuous dependent variable.

what are the characteristics of good sample

1. **Representativeness:** A good sample should be representative of the population in terms of relevant characteristics such as age, gender, income, education, etc.
2. **Randomness:** The sample should be randomly selected from the population, without bias or prejudice, so that each individual in the population has an equal chance of being selected.
3. **Large size:** A sample that is large enough will produce more accurate and reliable results than a small sample. The larger the sample size, the more representative it is likely to be of the population.
4. **Independence:** The observations in the sample should be independent of each other, meaning that the inclusion or exclusion of one individual should not affect the results for the rest of the sample.

5. **Relevance:** The sample should include only individuals or data that are relevant to the research question or study being conducted.

what is hypothesis testing:

Hypothesis testing is the process of testing the validity of postulated value or researcher's claim

5 reasons why sample may be preferred to a complete enumeration.

1. Cost-effective & Less time consuming
2. It is more efficient and convenient for researcher
3. It reduces the risk of data overload
4. It can provide representative cross-section of the population
5. Allows for studying of subgroups within a population, providing insights about said subgroups that may not be apparent from studying the population as a whole

What is the population in statistics:

A population is defined as the entire group of individuals, items or data that we are interested in studying.

What is sample in statistics:

It is a subset of the population that is selected for study or analysis.

Steps in hypothesis testing:

1. Determine the test statistics
2. Determine the critical region
3. Compare the calculated and tabulated values
4. Make a decision based on the values

Define the following terms;

- (i) Interval Estimation
- (ii) Point Estimate
- (iii) Critical Region
- (iv) Significance Level-