Chapter 8

Statistical Inference

Statistical Inference

Statistical Inference makes use of information from a sample to draw conclusions (inferences) about the population from which the sample was taken. It has two branches; Estimation and Testing of Hypothesis.

Estimator

An estimator is any quantity calculated from the sample data, which is used to give information about an unknown quantity in the population. In other words, any statistic that is used to estimate a population parameter is called estimator. For example, the sample mean is an estimator of the population mean.

Estimate

An estimate is a specific value or range of values used for indication of the value of an unknown quantity based on observed data. More formally, an estimate is the particular value of an estimator that is obtained from a particular sample of data and used to indicate the value of a parameter.

Estimation

Estimation is the process by which sample data are used to indicate the value of an unknown quantity in a population.

Results of estimation can be expressed as a single value, known as a Point Estimate, or a range of values, known as an Interval Estimate.

Properties of a Good Point Estimator

Unbiasedness:

An estimator $\hat{\theta}$ is said to be unbiased estimator of parameter θ if the mean of sampling distribution of the values of $\hat{\theta}$ is equal to θ i.e.

$$E(\hat{\theta}) = \theta.$$

Efficiency:

Efficiency refers to the size of standard error of the statistic. If we compare two statistics from a sample of the same size then the estimator with smaller standard error is said to be more efficient.

Particularly, an estimator $\hat{\theta}_1$ is said to be efficient estimator than $\hat{\theta}_2$ if Var $(\hat{\theta}_1)$ < Var$(\hat{\theta}_2)$. The variane is calculated if the estimators are unbiased. If estimators are biased then mean squared errors will be compared.

Consistency:

An estimator is said to be consistent estimator of a population parameter if as the sample size increases, it becomes almost certain that the value of the statistic comes very close to the value of population parameter.

In other sense an estimator $\hat{\theta}_n$ (as n is the sample size) is a consistent estimator for parameter θ if and only if, for all $\varepsilon > 0$, no matter how small, we have;

$$P(|\hat{\theta}_n - \theta| < \varepsilon) = 1, \text{ when } n \to \infty.$$

Sufficiency:

An estimator is called sufficient estimator if it makes so much use of the sample information that no other estimator could extract form the sample additional information about the population parameter being estimated.

Hypothesis

It is a supposition or assumption, which acts as a foundation or as a starting point in an investigation, irrespective of its probable truth or falsity. For example, average body temperature of adults is 98.6°F, procedure *A* of cultivation is better than that of *B*, etc.

Statistical Hypothesis

A statistical hypothesis is a statement about parameter(s) of population(s). For example, average body temperature of adults is 98.6°F, more than 70% voters are in favour of a particular party, etc.

Testing of Hypothesis

Hypothesis testing begins with an assumption, called a hypothesis that we make about a population parameter. Then we collect sample data, produce sample statistic, and use this as information to decide how likely it is that our hypothesized population parameter is correct. The purpose of this type of inference to determine whether enough statistical evidence exists to enable us to conclude that a belief or hypothesis about a parameter is supported by the data.

Null Hypothesis

A hypothesis to be tested for possible rejection under the assumption that it is true, is called null hypothesis and is denoted by H_0. For example, in a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug. We would write

H_0: there is no difference between the two drugs on average.

We give special consideration to the null hypothesis. This is due to the fact that the null hypothesis relates to the statement being tested.

Alternative Hypothesis

The alternative hypothesis, denoted by H_1, is to be considered as an alternate to the null hypothesis. It is also known as Research Hypothesis. For the above example, we would write

H_1: the two drugs have different effects, on average.

The alternative hypothesis might also be that the new drug is better, on average, than the current drug. In this case we would write

H_1: the new drug is better than the current drug, on average.

Simple Hypothesis

A simple hypothesis is a hypothesis, which specifies the population distribution completely.

For example,

1. H_0: $p = 0.5$, i.e., p is specified
2. H_0: $X \sim N(5, 20)$, i.e., μ and σ^2 are specified

Composite Hypothesis

A composite hypothesis is a hypothesis, which does not specify the population distribution completely.

For example,

1. $H_1: p > 0.5$, i.e., p is not completely specified
2. $H_1: X \sim N(5, \sigma^2)$, i.e., σ^2 is not completely specified

Type-I Error

In a hypothesis test, a type-I error occurs when the null hypothesis is rejected when it is in fact true; that is, H_0 is wrongly rejected. The probability of committing type-I error is denoted by α.

A type-I error is often considered to be more serious, and therefore more important to avoid, than a type II error. The hypothesis test procedure is therefore adjusted so that there is a guaranteed 'low' probability of rejecting the null hypothesis wrongly; this probability is never zero.

Type-II Error

In a hypothesis test, a type-II error occurs when the null hypothesis H_0, is not rejected when it is in fact false, H_0 is wrongly accepted. The probability of committing type-II error is denoted by β.

A type-II error would occur if it was concluded that the two drugs produced the same effect, i.e. there is no difference between the two drugs on average, when in fact they produced different ones. A type-II error is frequently due to sample sizes being too small.

Significance Level

The significance level of a statistical hypothesis test is a fixed probability of wrongly rejecting the null hypothesis H_0, if it is in fact true. It is the probability of a type I error and is set by the investigator in relation to the consequences of such an error. That is, we want to make the significance level as small as possible in order to protect the null hypothesis and to prevent, as far as possible, the investigator from inadvertently making false claims. The significance level is usually denoted by α.

Test Statistic

A test statistic is a quantity calculated from the sample of data. Its value is used to decide whether or not the null hypothesis should be rejected in our hypothesis test. The choice of a test statistic will depend on the assumed probability model and the hypotheses under question.

Critical Value

The critical value for a hypothesis test is a threshold to which the value of the test statistic in a sample is compared to determine whether or not the null hypothesis is rejected.

The critical value for any hypothesis test depends on the significance level at which the test is carried out, and whether the test is one-sided or two-sided (described below).

Critical Region

The critical region (CR), or rejection region (RR), is a set of values of the test statistic for which the null hypothesis is rejected in a hypothesis test. That is, the sample space for the test statistic is partitioned into two regions; one region (the critical region) will lead us to reject the null hypothesis H_0, the other will not. So, if the observed value of the test statistic is a member of the critical region, we conclude "Reject H_0"; if it is not a member of the critical region then we conclude, "Do not reject H_0".

One-Sided (One-Tailed) Test

The tails in a distribution are the extreme regions bounded by critical values. A test is said to be one-sided or one-tailed test if its entire critical region lies on just one (right or left) tail of the distribution under H_1.

In other words, the critical region for a one-sided test is the set of values less than the critical value of the test, or the set of values greater than the critical value of the test.

Example:

If μ = average body temperature of adults;

H_0: $\mu = 98.6$,

Against;

H_1: $\mu < 98.6$ or H_1: $\mu > 98.6$.

Two-Sided Test

A test is said to be two-sided or two-tailed test if its entire critical region lies on both tails of the distribution under H_1. In other words, the critical region for a two-sided test is the set of values less than a first critical value of the test and the set of values greater than a second critical value of the test.

Example :

Suppose, we want to test a manufacturers claim that there are, on average, 50 sticks in a match-box. We could set up the following hypothesis

H_0: $\mu = 50$,

Against;

H_1: $\mu \neq 50$

The choice between a one-sided and a two-sided test is determined by the purpose of the investigation or prior reasons for using a one-sided test.

P-Value

The probability value (*p*-value) of a statistical hypothesis test is the probability of getting a value of the sample test statistic that is at least as extreme as the one found from the sample data assuming that the null hypothesis is true.

It is the probability of wrongly rejecting the null hypothesis if it is in fact true. It is, actually, observed significance level of the test for which we would only just reject the null hypothesis. The *p*-value is compared with the presumed significance level of our test and, if it is smaller, the result is significant. That is, if the null hypothesis were to be rejected at the 5% significance level, this would be reported as "$p < 0.05$".

Small *p*-values suggest that the null hypothesis is unlikely to be true. The smaller it is the more convincing is the rejection of the null hypothesis.

Power of the Test

The power of a statistical hypothesis test measures the test's ability to reject the null hypothesis when it is actually false, that is, to make a correct decision.

In other words, the power of a hypothesis test is the probability of not committing a type- II error. It is calculated by subtracting the probability of a type- II error from 1, usually expressed as:

Power = $1 - P$ (type II error) = $1 - \beta$

The maximum power, a test can have is 1and the minimum is 0. Ideally, we want a test to have high power, close to 1.