Statistical Inference

Statistical inference is mainly concerned with making inferences regarding the unknown aspects of the distribution of the population based on samples taken from it. The unknown aspect may be the form of the distribution or values of the parameters involved or both. Statistical inference is broadly classified into two

1. Estimation of parameters
2. Testing of hypotheses

Estimation deals with methods of determining numbers which may be taken as the values of the unknown parameters (called the point Estimation) as well as with the determination of intervals which will contain the unknown parameters with a specified probability (known as interval estimation), based on samples taken from the population.

Testing of hypotheses deals with the methods for deciding either to accept or reject the hypotheses based on samples taken from the population, with the degree of validity of the decision indicated in terms of probability.

Statistics

Any function of the sample values is known as a statistics.

Point Estimation

Any value or function of the sample values suggested as the value of the parameters is known as the ***Estimator***.

Interval Estimation

In point estimation a number calculated from the sample is suggested as the estimate of the unknown parameter. In interval estimation we find out two statistics t_1 and t_2 ($t_1<t_2$) such that the probability that the interval (t_1,t_2) contains the true value of the unknown parameter has a pre-assigned value α called the confidence coefficient of the interval. Such an interval is called a confidence interval with confidence coefficient α.

Testing of Hypothesis

A hypothesis is an assertion about the form of the distribution or the value of the parameters of statistical populations.

Examples

1. The weight of the students follows the normal distribution
2. Height of the college students follows the normal distribution with mean 5 feet
3. The given data follows normal distribution with mean 10 and standard deviation 4.

Simple and Composite Hypothesis

If the population specifies the population completely then the hypothesis is called a simple hypothesis and otherwise it is called composite hypothesis.

In the above example 1 and 2 are composite and 3 is simple.

Statistical test

A procedure by which we may accept or reject the hypothesis based on sample taken from the population is called a statistical test.

Null hypothesis and Alternative hypothesis

The hypothesis that is tested is called the '*null hypothesis*' and is usually denoted by H_0. The hypothesis which we will accept or reject according as we reject or accept H_0. This hypothesis is called the '*alternative hypothesis*' and is usually denoted by H_1.

Two Types of Errors

It is impossible to assert whether a hypothesis is correct or wrong by a statistical test, as the decision is based on a sample only. A true hypothesis may be rejected and a false hypothesis accepted I a test. In short we admit that our procedure may result in committing one or the other of the following two types of errors

1. Rejecting H_0 when it is true
2. Accepting H_0 when it is false

The first is called Type I error and the second is called Type II error.

Action taken based on sample data	State of Nature	
	H_0 is true	H_0 is false
Reject H_0	Type I error	No error
Accept H_0	No error	Type II error

Test Statistics, Acceptance Region and Critical Region

The function of the sample observations is chosen to take the decision either to accept or reject the hypothesis is called the test statistics.

We divide the range of variation of the test statistics into two regions, acceptance regions and rejection region or critical region such that the probabilities of the two types of errors are not very large

Significance level and Power of the test

The probability of the test statistics falling in the critical region when the hypothesis is true is called the significance level or size of the test. That is the significance level is the probability of the first type of error and is denoted by α.

ie significance level = α = Prob.{Rejecting $H_0 \mid H_0$}

Probability of correct decision is called the power of the test. That is power of the test is the probability of rejecting the null hypothesis when the alternative hypothesis is true and is denoted by β.

Power = β = Prob.{Rejecting $H_0 \mid H_1$}
= 1 - Prob.{Accepting $H_0 \mid H_1$}
= 1 – Prob.{second type of error}.

Steps in a Statistical Test Procedure

The different steps in testing of hypothesis is as follows

1. Define the population and formulate the hypothesis
2. Choose an appropriate test statistics
3. Divide the range of variation of the test statistics into two regions, acceptance region (*A*) and rejection or critical region (*C*) so that probability of type I error significance level has a pre-assigned value
4. Take a sample, calculate the test statistics and decide whether to accept or reject the hypothesis.

Large Sample Test Concerning the mean of Normal Population

Sl. No.	Null Hypothesis	Alternative Hypothesis	Test Statistics (t)	Critical Region	Distribution of t
1	H_0: $\mu=\mu_0$ σ known	H_1: $\mu \neq \mu_0$	t = $\frac{(\bar{x}-\mu_0)\sqrt{n}}{\sigma}$	$\lvert t\rvert \geq t_{\alpha/2}$	Normal
2	H_0: $\mu=\mu_0$ σ known	H_1: $\mu < \mu_0$	t = $\frac{(\bar{x}-\mu_0)\sqrt{n}}{\sigma}$	$t < -t_{\alpha}$	Normal
3	H_0: $\mu=\mu_0$ σ known	H_1: $\mu > \mu_0$	t = $\frac{(\bar{x}-\mu_0)\sqrt{n}}{\sigma}$	$t > t_{\alpha}$	Normal
4	H_0: $\mu=\mu_0$ σ unknown	H_1: $\mu \neq \mu_0$	t = $\frac{(\bar{x}-\mu_0)\sqrt{n}}{s}$	$\lvert t\rvert \geq t_{\alpha/2}$	Normal
5	H_0: $\mu=\mu_0$ σ unknown	H_1: $\mu < \mu_0$	t = $\frac{(\bar{x}-\mu_0)\sqrt{n-1}}{s}$	$t < -t_{\alpha}$	Normal
6	H_0: $\mu=\mu_0$ σ unknown	H_1: $\mu > \mu_0$	t = $\frac{(\bar{x}-\mu_0)\sqrt{n-1}}{s}$	$t > t_{\alpha}$	Normal
7	H_0: $\mu_1=\mu_2$ σ known	H_1: $\mu_1 \neq \mu_2$	t = $\frac{(\bar{x}_1-\bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$\lvert t\rvert \geq t_{\alpha/2}$	Normal
8	H_0: $\mu_1=\mu_2$ σ known	H_1: $\mu_1<\mu_2$	t = $\frac{(\bar{x}_1-\bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$t < -t_{\alpha}$	Normal
9	H_0: $\mu_1=\mu_2$ σ known	H_1: $\mu_1>\mu_2$	t = $\frac{(\bar{x}_1-\bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$t > t_{\alpha}$	Normal
10	H_0: $\mu_1=\mu_2$ σ unknown	H_1: $\mu_1 \neq \mu_2$	t = $\frac{(\bar{x}_1-\bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$	$\lvert t\rvert \geq t_{\alpha/2}$	Normal
11	H_0: $\mu_1=\mu_2$ σ unknown	H_1: $\mu_1<\mu_2$	t = $\frac{(\bar{x}_1-\bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$	$t < -t_{\alpha}$	Normal
12	H_0: $\mu_1=\mu_2$ σ unknown	H_1: $\mu_1>\mu_2$	t = $\frac{(\bar{x}_1-\bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$	$t > t_{\alpha}$	Normal

Small Sample Test Concerning the mean of Normal Population

Sl. No.	Null Hypothesis	Alternative Hypothesis	Test Statistics (t)	Critical Region	Distribution of t
1	H_0: $\mu=\mu_0$ σ known	H_1: $\mu \neq \mu_0$	t = $\frac{(\bar{x}-\mu_0)\sqrt{n}}{\sigma}$	$\lvert t\rvert \geq t_{\alpha/2}$	Normal
2	H_0: $\mu=\mu_0$ σ known	H_1: $\mu < \mu_0$	t = $\frac{(\bar{x}-\mu_0)\sqrt{n}}{\sigma}$	$t < -t_\alpha$	Normal
3	H_0: $\mu=\mu_0$ σ known	H_1: $\mu > \mu_0$	t = $\frac{(\bar{x}-\mu_0)\sqrt{n}}{\sigma}$	$t > t_\alpha$	Normal
4	H_0: $\mu=\mu_0$ σ unknown	H_1: $\mu \neq \mu_0$	t = $\frac{(\bar{x}-\mu_0)\sqrt{n-1}}{s}$	$\lvert t\rvert \geq t_{\alpha/2}$	Students t_{n-1}
5	H_0: $\mu=\mu_0$ σ unknown	H_1: $\mu < \mu_0$	t = $\frac{(\bar{x}-\mu_0)\sqrt{n-1}}{s}$	$t < -t_\alpha$	Students t_{n-1}
6	H_0: $\mu=\mu_0$ σ unknown	H_1: $\mu > \mu_0$	t = $\frac{(\bar{x}-\mu_0)\sqrt{n-1}}{s}$	$t > t_\alpha$	Students t_{n-1}
7	H_0: $\mu_1=\mu_2$ σ known	H_1: $\mu_1 \neq \mu_2$	t = $\frac{(\bar{x}_1-\bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$\lvert t\rvert \geq t_{\alpha/2}$	Normal
8	H_0: $\mu_1=\mu_2$ σ known	H_1: $\mu_1<\mu_2$	t = $\frac{(\bar{x}_1-\bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$t < -t_\alpha$	Normal
9	H_0: $\mu_1=\mu_2$ σ known	H_1: $\mu_1>\mu_2$	t = $\frac{(\bar{x}_1-\bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$t > t_\alpha$	Normal
10	H_0: $\mu_1=\mu_2$ σ unknown	H_1: $\mu_1 \neq \mu_2$	t = $\frac{(\bar{x}_1-\bar{x}_2)}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\left(\frac{n_1s_1^2+n_2s_2^2}{n_1+n_2-2}\right)}}$	$\lvert t\rvert \geq t_{\alpha/2}$	Students $t_{n_1+n_2-2}$
11	H_0: $\mu_1=\mu_2$ σ unknown	H_1: $\mu_1<\mu_2$	t = $\frac{(\bar{x}_1-\bar{x}_2)}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\left(\frac{n_1s_1^2+n_2s_2^2}{n_1+n_2-2}\right)}}$	$t < -t_\alpha$	Students $t_{n_1+n_2-2}$
12	H_0: $\mu_1=\mu_2$ σ unknown	H_1: $\mu_1>\mu_2$	t = $\frac{(\bar{x}_1-\bar{x}_2)}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\left(\frac{n_1s_1^2+n_2s_2^2}{n_1+n_2-2}\right)}}$	$t > t_\alpha$	Students $t_{n_1+n_2-2}$