Statistical Inference

The process of drawing inferences or conclusions about population on the basis of sample information is called Statistical Inference.

Estimation

Hypothesis Testing

Estimation :-

Estimation is a procedure by which we obtain an estimate of population parameter using the sample information.
e.g $\bar{x}$ is an estimator of μ.

Testing of hypothesis :-

Testing of hypothesis is a procedure which enables us to decide on the basis of information obtained by sampling whether to accept or reject any specified statement or hypothesis regarding parameter. e.g a medical researcher may decide on the basis of experimental evidence whether smoking increase the risk of cancer or not.

Statistical Hypothesis :-

A Statement or assumption which may or may not be true is called hypothesis.

⇒ we accept the hypothesis as true if it is supported by sample data otherwise we reject the hypothesis. or Evidence from the sample that is inconsistent with the stated hypothesis leads to a rejection of hypothesis.

Null and alternate hypothesis :-

Null hypothesis is any hypothesis we wish to test for possible rejection under the assumption that it is true. Null hypothesis is denoted by H_0.

⇒ The term implies ussally that it is "no effect". e.g the drug is ineffective in curing the particular disease. or the coin is unbiased.

⇒ The null hypothesis should precise.

⇒ The null hypothesis should assign a numarical value. e.g $H_0 : \mu = 62$.

Alternate hypothesis is any other hypothesis which we accept when null hypothesis is rejected.

⇒ It is denoted by H_1 or H_A.

⇒ Null hypothesis is tested against alternate hypothesis. if null hypothesis is $H_0 : \mu = 62$ then alternate hypothesis
$H_1 : \mu \neq 62$ or $H_1 : \mu > 62$, $H_1 : \mu < 62$

Simple and composite hypothesis:-

A simple hypothesis is one in which all the parameters of the distribution are specified. e.g if heights of college students are normally distributed with $\sigma^2 = 4$, the hypothesis that its mean $\mu = 62$". As the mean and variance together specify a normal distribution completely. A hypothesis is composite if all the parameters are not specified. For instance if we hypothesize that $\mu > 62$ or $\sigma^2 < 4$ the hypothesis becomes a composite hypothesis. The concept of simple and composite hypothesis applies to both null & alternate

Test statistic:-

A sample statistic which provides basis for testing null hypothesis.

⇒ Every test statistic has a probability distribution which gives the probability of obtaining a specified value of the test statistic when null hypothesis is true.

⇒ The sampling distribution of most commonly used test statistic are Z, t, F and chisquare.

Acceptance and Rejection region:-

All possible values which a test - statistic may assume can be divided into two mutually exclusive groups.

⇒ one group consisting of values which appear to be consistant with null hypothesis, this group is called acceptance region.

⇒ other group having values which are unlikely to occur if Ho is true, this group is known as rejection region. The rejection region is also called critical region.

⇒ The values that separates the critical region from acceptance region are called the critical values.

Type I and Type II Errors:-

When we perform a hypothesis test, we derive the evidence from the sample in the form of a test-statistic. There is a possibility that the sample evidence may lead us to make a wrong decision.

⇒ Rejection of null hypothesis Ho, when it is actually true is called Type I error.

⇒ Acception or non-rejection of null hypothesis Ho, when it is actually false is Type II error.

True situation	Decision: Do not reject or Accept H_0	Decision: Reject H_0
H_0 is true	Correct Decision (No Error)	Wrong Decision (Type I Error)
H_0 is false	Wrong Decision (Type II Error)	Correct Decision (No Error)

Example:-

In a court trial the hypothesis is that the accused is innocent. After having heard the evidence presented during the trial, the Judge arrives at a decision. Suppose that the accused is, in fact innocent (i.e H_0 is true) but the findings of the judge has rejected the true null hypothesis and in doing so he has made type I Error.

⇒ If on the other hand, the accused is in-fact guilty (i.e H_0 is false) and the findings of judge is innocent, the judge has accepted a false null hypothesis and by accepting a false hypothesis, he has committed a type II error.

⇒ The probability of committing type I Error is also called level of significance and denoted by α.

⇒ Type II Error is denoted by β.

$$\alpha = P(\text{Type I Error}) = P\left(\text{Reject } H_0 / H_0 \text{ is true}\right)$$

$$\beta = P(\text{Type II Error}) = P\left(\text{Do not reject } H_0 / H_0 \text{ is false}\right)$$

General procedure for Testing hypothesis:-

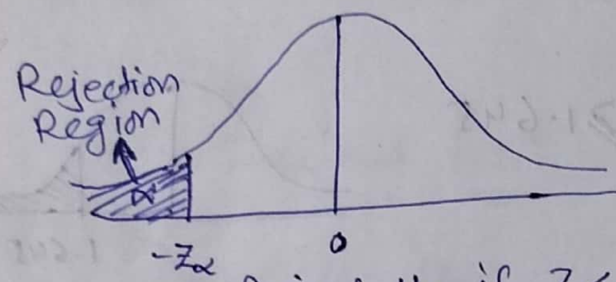
1) State the null and alternate hypothesis
2) Choose level of significance.
3) Choose an appropirate test-statistic and establish the C.R based on α.
4) Reject H_0 if computed value of test-Statistic is in the C.R, otherwise do not reject.
5) Draw conclusions.

Testing hypothesis about mean (Single Sample)

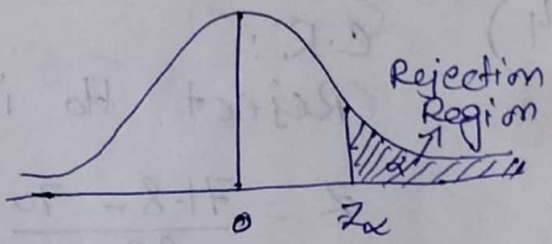
1) Testing on a single mean when variance known. (Z-test)
2) Testing on a single mean when variance is unknown. (t-test) and $n < 30$

One-tailed and two-tailed tests:-

A test for which the entire rejection region is located in only one of the two-tails, either in the right side or in the left tail of the sampling distribution of test statistic is called one tailed test. e.g If Z is the test-statistic



Reject H_0 if $Z < -Z_\alpha$



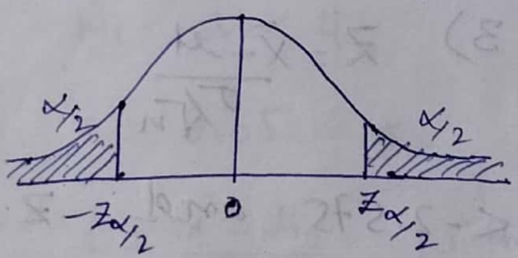
Reject H_0 if $Z > Z_\alpha$

A one tailed test is used when the alternative hypothesis is formulated as

$H_1: \mu > \mu_0$ or $H_1: \mu < \mu_0$

⇒ If the rejection region is divided equally between the two tails of the distribution of test-statistic, the test is referred to as a two tailed test. i.e

$H_1: \mu \neq \mu_0$



Reject H_0 if $Z < -Z_{\alpha/2}$ or $Z > Z_{\alpha/2}$

"The location of the C.R can be determined only after H_1 has been stated."

Example 10.3 (Case I when variance is known) (Walpole, Pg 338)

1) $H_0: \mu = 70$ or $\mu \leq 70$ (variance known)

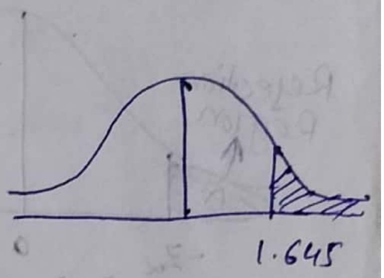
$H_1: \mu > 70$

2) $\alpha = 0.05$

3) $$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

4) C.R:

Reject Ho if $Z > 1.645$



$$Z = \frac{71.8 - 70}{8.9/\sqrt{100}}$$

$$Z = 2.02$$

5) As the calculated value of Z lies in rejection region so reject Ho and conclude that mean life span today is greater than 70 years.

Example 10.4

1) $H_0: \mu = 8$

$H_1: \mu \neq 8$

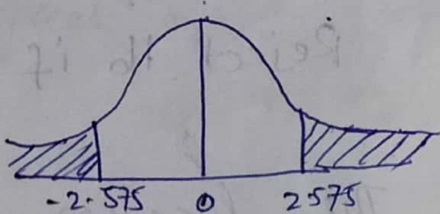
2) $\alpha = 0.01$

3) $$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

4) Critical region:-

Reject Ho if $Z < -2.575$ or $Z > 2.575$

$$Z = \frac{7.8 - 8}{0.5/\sqrt{50}} = -2.83$$



5) Decision: Reject Ho and conclude that the average breaking strength is not equal to 8.

Case II :-

Testing of mean when variance unknown.

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Two tailed $H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$

C.R : Reject H_0 if $t < -t_{(\alpha/2, n-1)}$ or
$t > t_{(\alpha/2, n-1)}$

one tailed test

1) $H_0: \mu = \mu_0$
$H_1: \mu > \mu_0$

Reject H_0 if $t > t_{\alpha, (n-1)}$

2) $H_0: \mu = \mu_0$
$H_1: \mu < \mu_0$

Reject H_0 if $t < -t_{\alpha, (n-1)}$

Example 10.5 (walpole, Pg 340)

1) $H_0: \mu = 46$
$H_1: \mu < 46$

2) $\alpha = 0.05$

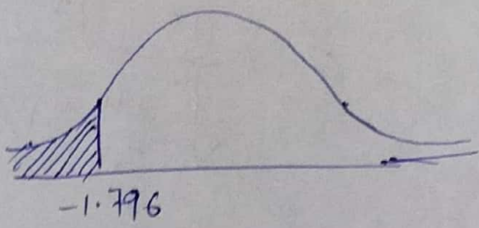
3) $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ with $n-1 = \overset{=12-1}{11}$ degrees of freedom.

4) C.R: Reject H_0 if $t < -1.796$ with 11 d.f.

$$t = \frac{42 - 46}{11.9/\sqrt{12}}$$

$t = -1.16$

5) Conclusion:-



Do not reject Ho.

Decision: The average number of killowatt hours used annually by home vacuum cleaners is not significantly less than 46.

Testing hypothesis about mean:- (Two Samples)

1) Two Sample with known population Variance. (Z - test)

Question:- A random Sample of size 36 from a normal population with variance 24 gave $\bar{x}_1 = 15$. A Second sample of size 28 from another normal population with variance 80 gave $\bar{x}_2 = 13$. Test $H_0: \mu_1 - \mu_2 = 0$ against $H_1: \mu_1 - \mu_2 \neq 0$, Let $\alpha = 0.05$.

Solution:-

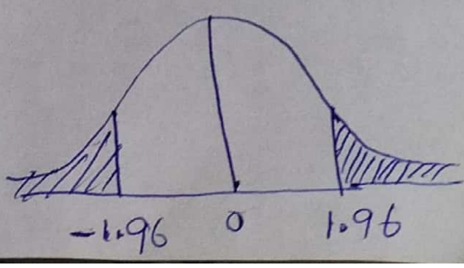
1) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$
   $H_1: \mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$

2) $\alpha = 0.05$

3) $$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

4) Critical region:-

Reject Ho if $Z < -1.96$ or $Z > 1.96$



$$Z = \frac{(15-13)-0}{\sqrt{\frac{24}{36}+\frac{80}{28}}}$$

$$Z = \frac{2}{1.88} = 1.06$$

5) conclusion:-

Do not reject Ho or accept Ho and conclude that $\mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$.

Degrees of freedom:- Number of items that are free from restrictions.

have only 7 hats. Yet you want to wear a different hat every day of the week.

On the first day, you can wear any of the 7 hats. On the second day, you can choose from the 6 remaining hats, on day 3 you can choose from 5 hats, and so on.

When day 6 rolls around, you still have a choice between 2 hats that you haven't worn yet that week. But after you choose your hat for day 6, you have no choice for the hat that you wear on Day 7. You *must* wear the one remaining hat. You had 7-1 = 6 days of "hat" freedom—in which the hat you wore could vary!