Module-5

Hypothesis Testing - I

Population and Sample :-

In statistics population refers to the data which is our study of interest.

A sample is the subset of the population which we select to make inference or draw conclusion about the entire population.

Terms related to tests of hypothesis :-

(1) Parameters :- The statistical constants of population are called parameters, such as mean (μ), S.D (σ), population proportion (P).

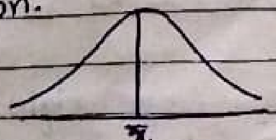
(2) Statistic :- The statistical constants of sample such as mean ($\bar{x}$), S.D (S), population proportion (p) are called statistic.

(3) Sampling Distribution :-

From a population of size N, consider all population samples of equal size n, Each sample collected, will have different value of statistic (say mean). Since sample is random, every statistic is random variable. As a r.v, it has mean, S.D, etc.

Lets consider mean of each n samples, if these means according to their probabilities, the prob. dist^n such formed is called sampling dist^n.

As size of sample is large ($n \geq 30$), the sample mean ($\bar{x}$) is approximately normally distributed, with mean $\bar{x}$ and S.D $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, σ is S.D of population.



classmate
Date
Page

Standard error:- The standard deviation of sampling distribution, is called standard error of statistic.

Hypothesis:-

when we have to make some decision about the population, based on sample information, we have to make assumption about the nature of the population or parameters of the population, such assumption which may or may not be true are called Hypothesis.

There are two type of Hypothesis:-

(1) **Null Hypothesis:-**

we set a hypothesis, which assumes that there is no significant difference between population parameter and sample statistic, such a hypothesis is called null hypothesis. We initially assumes that null hypothesis is true and then test it for possible rejection. It is denoted by H_0.

(2) **Alternate Hypothesis:-**

The hypothesis which is complementary to null hypothesis is called alternate hypothesis. It is denoted by H_1 or H_a.

Ex:- A university claims that an average weight of students is 50 kg.

Here, the null hypothesis is $H_0 : \mu = 50$. it's true that (ie we assume that average weight of student is 50 kg. Now it will be checked so we set Alternate hypothesis for that)

So Alternate hypothesis, could be any one of the following:-

$H_1 : \mu \neq 50$ (avg. weight is not 50 kg)

$H_1 : \mu < 50$ (avg. weight is less than 50 kg)

$H_1 : \mu > 50$ (avg " " more than 50 kg.)

ie If null Hypothesis is

$H_0 : \mu = \mu_0$

then alternate hypothesis could be any one of the following

$H_1 : \mu \neq \mu_0$ (Two tailed hypo)	(For decision making)
or $H_1 : \mu < \mu_0$ (left tailed hypo)	(For validity)
or $H_2 : \mu > \mu_0$ (right tailed hypo)	(For research hypo)

After setting the hypothesis, we test the hypothesis. It is a process, in which we decide whether to accept or reject the null hypothesis (hence to reject or accept the alternate hypothesis).

First we decide which test is to applied. As test is to be chosen depending on the sample size.
For sample size n greater or equal to 30 ie $n \geq 30$
Z - Test is applied.
Now we calculate Test statistic.

($n \geq 30$, is considerd as large sample size)

Test statistic :-

For Z - Test, test statistic is given by

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \quad \text{or} \quad Z = \frac{\bar{x} - \mu}{S.E \text{ of } \bar{x}}$$

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \sim N(0,1)$$

z is standard Normal Variate.

$\bar{x}$ = mean of sample
μ = mean of population
$\sigma_{\bar{x}}$ = S.E of $\bar{x}$

$\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$, where σ is S.D of population. ie if S.D of population is given

$\boxed{\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}}$, S = S.D of Sample, we use this formula if S.D of population is not given

Error in hypothesis :-

To draw valid conclusion about the population parameter, based on the sample result, finally after finding test statistic, we decide either to accept or reject the null hypothesis, our decision may or may not be true. ie based on this, there are two type of errors in hypothesis testing.

(1) Type I error
(2) Type II error

Type I error :- Type I error, occurs when ~~It is the error~~ we rejected the null hypothesis, H_0, when it is true.

ie when null hypothesis is true but difference b/w mean is significant and H_0 is rejected.

level of significance :-

The max probability of making type I error is called level of significance and denoted by α.

ie α gives probability of rejecting H_0, while H_0 is true. ie probability of making wrong decision is α.

ie probability of accepting H_0, when H_0 is true is $1-\alpha$. ie, prob. of making correct decision is $1-\alpha$.

The commonly used level of significance are 5% and 1% (will be provided in the question)

$$\begin{bmatrix} \alpha = 5\% & \text{ie} & p = .05 \\ \alpha = 1\% & \text{ie} & p = .01 \end{bmatrix}$$

ie at 5% level of significance means prob. of making Type I error is .05.

ie 5% chances are there that our test will reject the null hypothesis, when it is true

$P(\text{rejecting } H_0, \text{ when } H_0 \text{ is true}) = .05$

Type II error:-

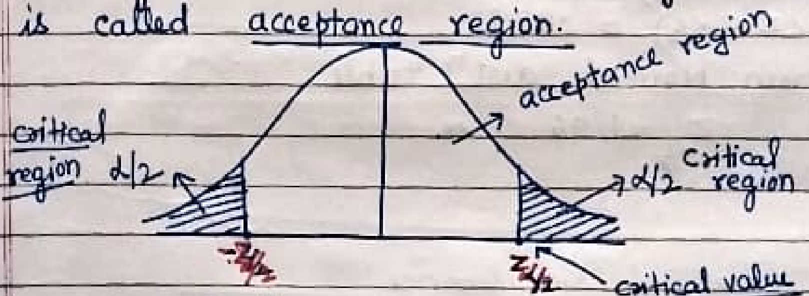
Type II error occurs, ~~when~~ if we accept the null hypothesis, when it is false. ie when difference of mean is insignificant and Ho is accepted. denoted by β.

	Accept Ho	Reject Ho
Ho True	correct decision	Type I error
H_0 False	Type II error	correct decision.

* Based on level of significance in the hypothesis two type of regions are there.

Critical region:- It is a region under standard normal curve, where null hypothesis is rejected. It is called rejection region.

The region which is not covered by the rejection region is called acceptance region.



at α level of significance.

The value of z, that separate the critical and acceptance region is called critical value.

Critical region and critical value for different test

(a) For two tailed test:-

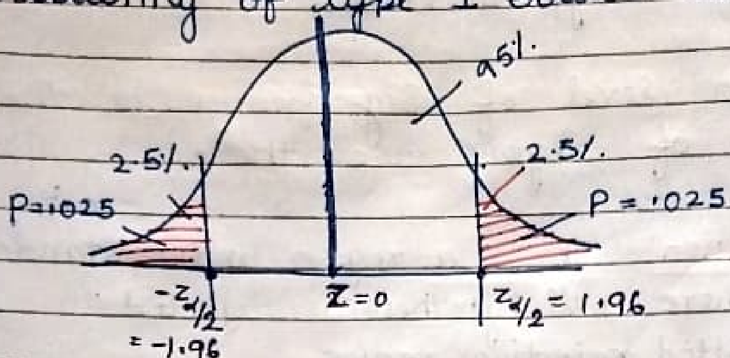
(1) If $\mu \neq \mu_0$, test is two tailed test i.e.

$$\begin{bmatrix} H_0 : \mu = \mu_0 & \text{(null hypo)} \\ H_1 : \mu \neq \mu_0 & \text{(alter. hypo)} \end{bmatrix}$$

The critical values are $z_{\alpha/2}, -z_{\alpha/2}$ at α level of significance.

* at 5% level of significance:-

Probability of type I error = $\cdot 05$



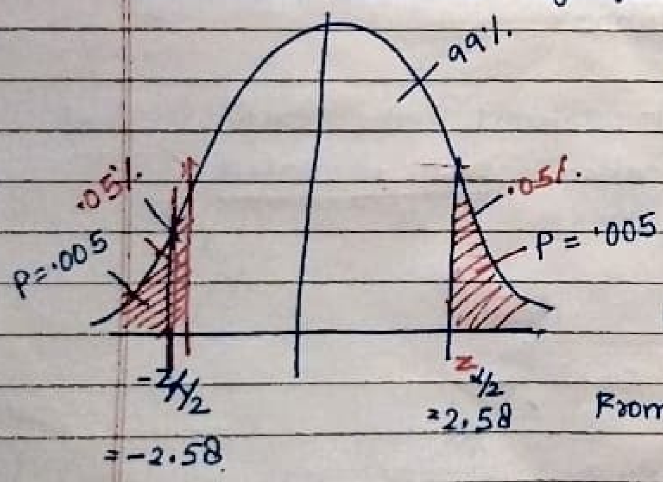
from Table area from 0 to

$$P(0 < z < z_{\alpha/2}) = \cdot 5 - \cdot 025$$
$$P(0 < z < \alpha/2) = \cdot 475$$

ie from Normal distⁿ table

$$z_{\alpha/2} = 1.96$$

* at 1% level of significance:



$$P(0 < z < z_{\alpha/2}) = \cdot 5 - P(z > z_{\alpha/2})$$

$$P(0 < z < z_{\alpha/2}) = \cdot 495$$

From Table

$$z_{\alpha/2} = 2.58$$

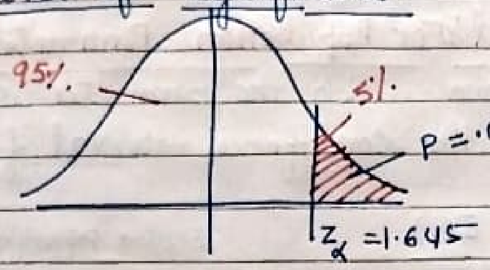
(b) One Tailed Test :-

(*) If $\mu > \mu_0$, then test is right tailed test.

ie $H_0 : \mu = \mu_0$

$H_1 : \mu > \mu_0$

at 5% level of significance :-



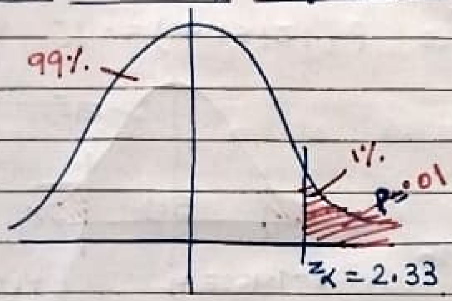
$P(0 < z < z_\alpha) = .5 - .05$

$= .45$

From Table

$\boxed{z_\alpha = 1.645}$

at 1% level of significance :-



$P(0 < z < z_\alpha) = .5 - .01$

$= .49$

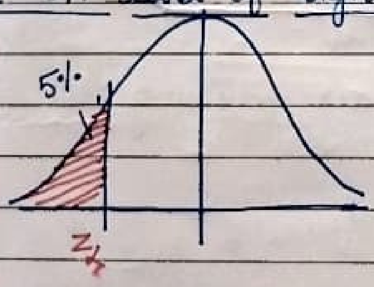
ie $\boxed{z_\alpha = 2.33}$

(*) If $\mu < \mu_0$ then test is left tailed :

ie $H_0 : \mu = \mu_0$

$H_1 : \mu < \mu_0$.

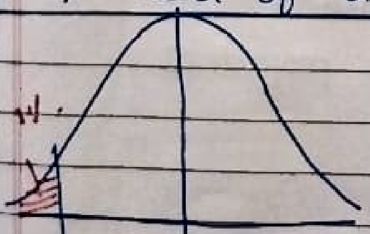
at 5% level of significance :-



by symmetery

$z_\alpha = -1.645$

at 1% level of significance



by symmetry

$\boxed{z_\alpha = -2.33}$

Critical value

Table for Z-Test

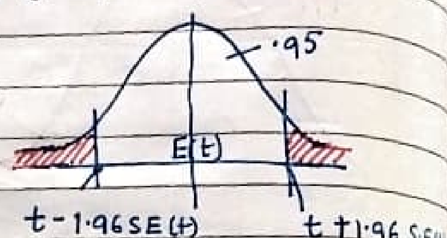
Test	1% (.01)	2% (.02) loss	5% (.05)	10% (.1)
Two Tailed	$\lvert z_{\alpha/2}\rvert = 2.58$	$\lvert z_{\alpha/2}\rvert = 2.33$	$\lvert z_{\alpha/2}\rvert = 1.96$	$\lvert z_{\alpha/2}\rvert = 1.6$
Right Tailed	$z_\alpha = 2.33$	$z_\alpha = 2.055$	$z_\alpha = 1.645$	$z_\alpha = 1.28$
Left Tailed	$z_\alpha = -2.33$	$z_\alpha = -2.055$	$z_\alpha = -1.645$	$z_\alpha = -1.28$

Confidence Interval: For Population Parameter:-
The interval within which the parameter is expected to lie is called the condifence interval for that Parameter.

Ex. Say $$z = \frac{t - E(t)}{S\cdot E(t)}$$ $E(t) \rightarrow$ Popuation para.
$t \rightarrow$ Sample Static

for 5% level of significance
$$|z| \le 1.96 \quad \text{ie} \quad P(|z| \le 1.96) = .95$$

ie $$\left|\frac{t - E(t)}{S\cdot E(t)}\right| \le 1.96$$



ie $$-1.96 \le \frac{t - E(t)}{S\cdot E(t)} \le 1.96$$

$$\boxed{t - 1.96 \times S\cdot E(t) \le E(t) \le t + 1.96 S\cdot E(t)}$$

Similarly at 1% loss, confidence interval for population parameter is:-

$$\boxed{t - 2.58\ S\cdot E(t) \le E(t) \le t + 2.58\ S\cdot E(t)}$$

Large Sample Tests:- we consider following test under large sample test:-

(1.) Test for singee mean
(2) Test for difference of means
(3) Test for single proportion
(4) Test for difference of proportions

Procedeure for Hypothesis Testing :-

Test 1:- Test for single Mean :-

(1) set up null Hypothesis $H_0 : \mu = \mu_0$
(2) Set up alternate Hypothesis $H_1 : \mu \neq \mu_0$
or $(\mu < \mu_0)$
or $(\mu > \mu_0)$

(3) select suitable level of significance

(4) Calculate test statistic

$$\left[Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \right]$$

$\bar{x}$ - sample mean
μ - population mean
σ S.D of population
n - sample size

Note - if σ is not known, sample S.D 'S' can be used in place of σ.
ie

$$\left[Z = \frac{\bar{x} - \mu}{S / \sqrt{n}} \right]$$

(5) If $|Z_{cal}| < |Z_\alpha|$, accept H_0,
other wise reject H_0 (ie accept H_1)

Note: Null Hypothesis is accepted at α level of significance, it means that difference b/w the population mean and sample mean is not significant at α% level of significance.

Note:- 95% confidence limit for μ are

$$\left[\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \right]$$

If σ is not known then sample S.D s will be used ie 95% confidence limit is

$$\boxed{\bar{x} - 1.96\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + 1.96\frac{s}{\sqrt{n}}}$$

Test 2 Test for difference of mean b/w two Samples

1. Set up $H_0: \bar{x}_1 = \bar{x}_2$ or $(\mu_1 = \mu_2)$

(2) Set up $H_1: \bar{x}_1 \neq \bar{x}_2$ or $\mu_1 \neq \mu_2$ (two tailed)
(or) $\bar{x}_1 < \bar{x}_2$ or $\mu_1 < \mu_2$ (left tailed)
(or) $\bar{x}_1 > \bar{x}_2$ or $\mu_1 > \mu_2$ (right tailed)

(3) Select α (l.o.s)

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

$\bar{x}_1, \bar{x}_2$ are means of two large sampless of size n_1 & n_2, drawn from two populations with variances σ_1^2 & σ_2^2, with mean μ

(5) find Z tabulated value at α l.o.s ie z_α

(6) If $|z_{cal}| < |z_\alpha|$ accept H_0

otherwise accept H_1. (H_0 will be rejected)

... σ is known

classmate

Date

Page

* If two samples are drawn from same population with S.D σ ie $\sigma_1 = \sigma_2 = \sigma$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

ie

$$\left[Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right] \quad \text{———} \; (1)$$

* If σ_1 & σ_2 are not known and $\sigma_1 \neq \sigma_2$ then σ_1, σ_2 can be approximated by santard deviation of ~~popul~~ Sample.
ie S_1 and S_2 are S.D of Sample then

$$\left[Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \right] \quad \text{———} \; (2)$$

* If σ_1 & σ_2 are not known by equal $\sigma_1 = \sigma_2$ then $\sigma = \sigma_1 = \sigma_2$ can be approximated by.

$$\sigma^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}$$

ie

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\left[Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_2} + \frac{S_2^2}{n_1}}} \right] \text{———} \; (3)$$

Note :- denominators of (2) & (3) are different

classmate
Date
Page

Test-3 Test For Single Proportion

(1) set up $H_0 : P = P_0$ (null Hypo)

(2) set up $H_1 : P \neq P_0$
(or) $P < P_0$
or $P > P_0$

(3) select α (l.o.s)

(4) calculate

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \quad , \text{ where } Q = 1 - P$$

p = sample proportion

P = population proportion

n = sample size

(5) Find the z tabulated value at α level of significance i.e. Z_α.

(6) check if $|Z| < |Z_\alpha|$, H_0 is accepted and H_1 rejected

otherwise accept H_1 & reject H_0.

Test-4 Test for difference of proportions b/w two samples

Let p_1 and p_2 be proportions of successes in two large samples of size n_1 and n_2 resp. drawn from samp population or from two different population having same proportion P.

(1) Set up $H_0 : p_1 = p_2$

(2) Set $H_1 : p_1 \neq p_2$
(or) $p_1 < p_2$
or $p_1 > p_2$

(3) Choose suitable level of significance α.

(4) Calculate

$$z = \frac{p_1 - p_2}{\sqrt{\frac{PQ}{n_1} + \frac{PQ}{n_2}}}$$

$$\left[z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right]$$

Note:- If P is not known then an unbiased estimate of P base on both samples is given by

$$\left[P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} \right]$$

(5) if $|z| < |z_{\alpha}|$, accept H_0 and reject H_1
otherwise accept H_1 and reject H_0.

* ~~If P is~~

* 95% confidence limits for P are given by

$|Z| \le 1.96$

$$\left| \frac{p - P}{\frac{\sqrt{pq}}{\sqrt{n}}} \right| \le 1.96$$

$\sqrt{\frac{pq}{n}}$ is taken as in the denominator, as P is assumed to be unknown, for which we are trying to find confidence limit and P is neary equal to p.

ie $$\boxed{p - 1.96\sqrt{\frac{pq}{n}} \le P \le p + 1.96\sqrt{\frac{pq}{n}}}$$