Handout - 1

Topic: Testing of Hypothesis

Objective: To make inferences about a population based on the information provided by samples selected from the population

Important Notations:

	For Population (large group) (Parameters)	For Sample (Subset of population) (Statistics)
Mean	μ	$\bar{X}$
Variance	σ^2	S^2
S.D	σ	S
Proportion	P	p or $\hat{p}$
Difference of Mean (Betn two things) comparison	$\mu_1 - \mu_2$	$x_1 - x_2$
Difference of Proportion	$P_1 - P_2$	$p_1 - p_2$ or $\hat{p}_1 - \hat{p}_2$

Important Def^n:

Population: It is a large group of objects/things/huma[n] for which result to be concluded.

Sample: It is a ~~p~~ small part of population, select[ed] randomly. using statistics of sample, w[e] derived result for population paramete[r]

Null Hypo.: (H_0): Set, Based on population data
- Tested for acceptance or rejecti[on]
- Normally It is a hypo. of no diff[erence]
- Research wants it to be reject[ed]

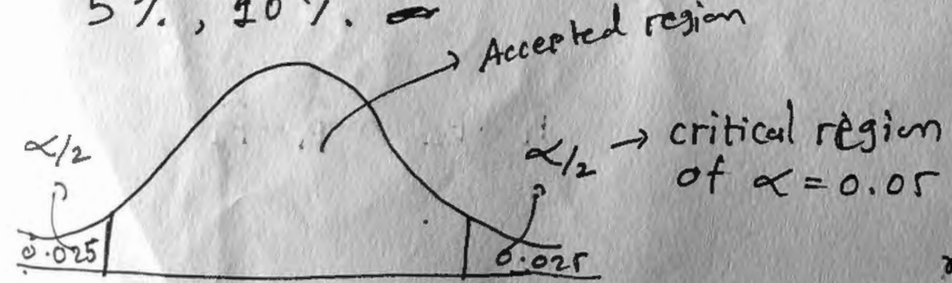
Alternative Hypo. : (H_1) :- Complementary of null hypo.
+ It is interest of researchers.

Note: - acceptance of null hypo does not mean that it is true only we do not have enough evidence to reject it
- Rejection of null hypo. means we have evidence that it is false

Type I error: If Null hypo is true and though reject it

Type II error: If Null hypo is false and though accept it (bcoz of not enough evidence)
Researchers do not want such situation so they try to minimize such errors.
& so increase % level of acceptance region.
& keep sample large.

Level of significance: (α) : Also known as critical region:
It is that area where ~~researchers~~ there is maximum probability of risk of Type I error. (It rejects a lot when it is good).
Normally this area is small. for eg 1%, 2%, 5%, 10%.



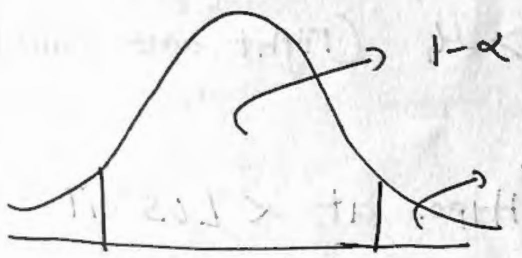
Explanation:
out of 100%. researchers ~~set~~ test null hypo. at 5% LOS. so he/she can accept hypo in large area i.e 95%.
- This 5% area indicates probable chance of rejection of null hypo.
- Means 5 chances out of 100 that null hypo. is rejected, when it is true.
- & we are 95% confident that our decision is right.
- Means hypo. has been rejected at a 5% LOS, that we could be wrong with only 0.05 chance.

level of confidence : $(1-\alpha)$: It is complimentary area of α
If the LOS is 1 % means ~~LOS in 99%~~ level of confidence is 99 %.

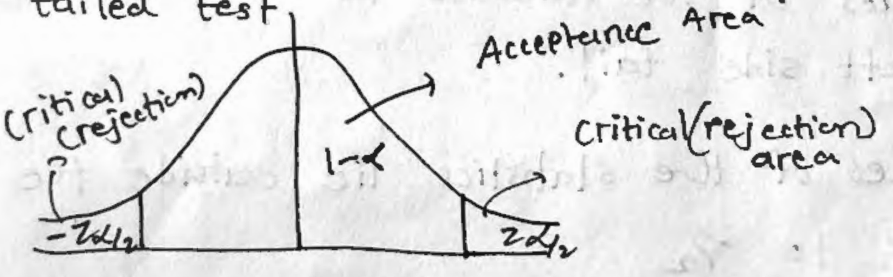
Degree of Freedom : It is no. of Indept. observations of the samples.

$$\nu = n - K \qquad n = \text{sample size}$$
$$K - \text{no of constraints}$$

for eg. In addition of five no whose total is 100, first four no we can select any but fift no have restriction. so here $n=5$ & $K=1$.



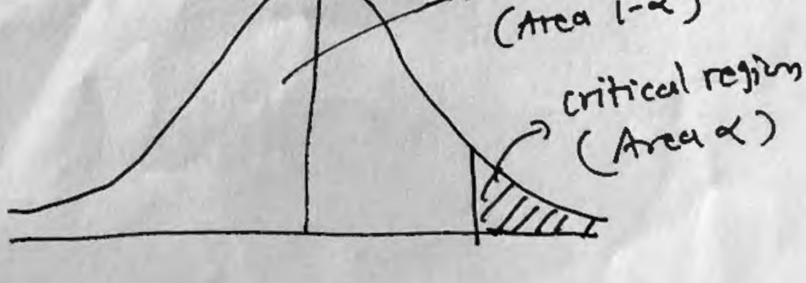
Two tailed test : when α value check at both part of curve (ie left and right area) then it is called two tailed test.



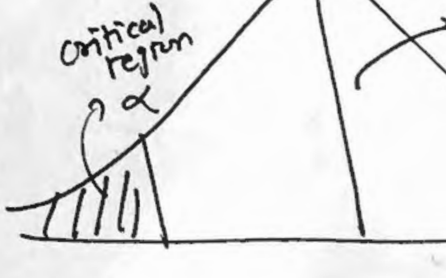
when Z value calculated using formula and if is in the critical region, Htpo. is rejected

One tailed test :

Right one tailed test



left one tailed t



Type of Hypothesis :

For Mean :

Case I :

$H_0 : \mu = \mu_1$ (some value)

$H_1 : \mu \neq \mu_1$ (Two tailed test)

Case II :

$H_0 : \mu = \mu_1$

$H_1 : \mu < \mu_1$ (left one tailed test)

Case III

$H_0 : \mu = \mu_1$

$H_1 : \mu > \mu_1$ (right one tailed test)

Use when to determine which one population is better (comparison of better)

Rejection rule : Reject Hypo. at α LOS if

(1) +ve values of statistics (refer handouts) lie outside α on on the right side tail

(2) −ve values of the statistics lie outside $(-\alpha)$ on the left side tail.

(3) The values of the statistics lie outside the range $-\alpha/2$ to $\alpha/2$

Handout 2 Date: 20/9/11

Topic: Algorithm of HYPO. and formulas.

<u>objective</u>: To know how to do testing of Hypot

<u>Steps For HYPO</u>:

Step 1: Set Null HYPO. H_0 & alternate HYPO. H_1

Step 2: Choose α (LOS) (any 5%, 1%, 2%)

Step 3: Determine Degree of freedom (only using t-test, χ^2-test & F-test)

Step 4: Use appropriate test out of z-test, t-test χ^2 test & F test.

For Large sample $(n \geq 30) \rightarrow$ use z-test

For Small " $(n < 30) \rightarrow$ use t-test

For eg. $Z = \cancel{calc}\ \frac{\bar{x} - \mu}{S.E}$ (for mean) ~~[illegible]~~

<u>Step 5</u>: Determine critical region

Step 6: Decision: Always it will be by comparing calculate statistic (z, F, t, χ^2) with critical value (from standard table).

Note: If calculated value < critical value

~~Accept~~ Null HYPO. is ~~Rejected~~ accepted

If calculated value > critical value

Null hypo. is rejected

$$Z = \frac{\bar{x} - \mu}{S.E} = \frac{\text{calculated value} - \text{Observed value}}{S.E}$$

(↳ Refer Diff formulas for S.E)

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

observed value is always what has been set for null hypo. (related to population)

Std. error: (S.E.) : It is a std. devi. of the sampling distribution of a statistic.

Statistics	Std. error (S.E)
Mean : $\bar{x}$	$\frac{\sigma}{\sqrt{n}}$
	$\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ when N is finite no. given also
	$\therefore \sqrt{\frac{N-n}{N-1}}$ is known as finite population correction factor.
Proportion : $p, \hat{p}$	$\sqrt{\frac{pq}{n}}$ where p - prob of success, q - prob of failure
	$\therefore \sqrt{\frac{pq}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$
Diff. of Mean : $\bar{x}_1 - \bar{x}_2$ when comparison is betn two things	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ when σ is known
	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ when σ is unknown calculate s_1, s_2 as std. deviation from given data normally we need to do so when small data (t test) apply
Diff. of proportion: $P_1 - P_2$ when comparison is bet^n two things	$\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Handout 2

	level of significance		
	α 1% (0.01)	α 5% (0.05)	α 10% (0.1)
Two tailed Test	$\lvert Z_\alpha \rvert = 2.58$	$\lvert z \rvert = 1.966$	$\lvert z \rvert = 0.645$
Right tailed	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$	$Z_\alpha = 1.28$
left tailed	$Z_\alpha = -2.33$	$Z_\alpha = -1.645$	$Z_\alpha = -1.28$

How to see above values from Z table (Normal table)

For Two tailed test

Step 1: For $\alpha = 0.01$ do first $\frac{\alpha}{2} = \frac{0.01}{2} = 0.005$

Step 2: Calculate: $0.5 - 0.005$

$= 0.495$

Step 3: Find 0.495 from z table and observe corresponding raw and col. $\boxed{\text{gives } 2.58}$ as z value

z	0.00	0.01	0.02	0.03	..	0.05	0.06	0.07	0.08	0.09
0.0										
0.1										
0.2										
0.3										
⋮										
2.1										
2.2										
2.3	0.4893	0.4896	0.49	$\boxed{0.4901}$	← one tailed					
2.4										
2.5	0.4938	0.4940	- - -				0.4948	0.4949	$\boxed{0.4951}$	0.4952
2.6									↑ two tailed	

For one tailed test

Step 1: For $\alpha = 0.01$

Step 2: Calculate: $0.5 - 0.01$

$= 0.49$

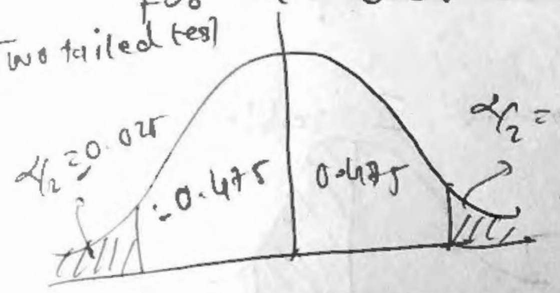
Step 3: Find 0.49 from z table and observe corresponding row & col. $\boxed{\text{gives } 2.33}$

Note:

Normally we ~~do~~ do ~~set~~ two tailed test, but specifically mentioned about interior, superior, less, greater than accordingly we do one tailed test (Right or Left

For $\alpha = 0.05$

Two tailed test



$$\begin{array}{r} 0.5 \\ -0.025 \\ \hline 0.475 \end{array}$$

from Z table Area = 0.475

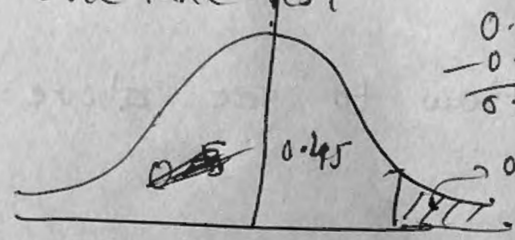
Z value is 1.96

so Interval is

[-1.96 to 1.96]

Right

One taile test



$$\begin{array}{r} 0.5 \\ -0.05 \\ \hline 0.45 \end{array}$$

Area = 0.45

Z value is 1.65

Left one tail test

Z value is -1.65

Date: 30/9/19

Handout - 3

Topic: Formula of statistics

Objective: To calculate S.E & then calculate value of Z, t, F, χ^2

For Large sample:

Mean μ:

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

Proportion P:

$$Z = \frac{\hat{p} - \mu_p}{\sigma_p} = \frac{p - P}{\sqrt{\frac{pq}{n}}}$$

$$\text{or} = \frac{p - P}{\sqrt{\frac{pq}{n}} \cdot \sqrt{\frac{N-n}{N-1}}}$$

when N is given

Difference of Mean: $\mu_1 - \mu_2$

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

when σ is unknown calculate as

$$\sigma = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}}$$

Difference of Proportion: $P_1 - P_2$.

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

Here $P_1 = \frac{x_1}{n_1}$ = favo... / Tot...

$P_2 = \frac{x_2}{n_2}$ = no of ... / Total

If $P_1 = P_2 = P$

Prob of succes is equal for both object then

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

wher $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ =

Small sample: T-test

(2)

Mean μ: $$t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$ where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

& $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

& D.F (Degree of freedom)
$v = n-1$

Diff. of Mean: Degree of freedom is n_1+n_2-2 $((n_1-1)+(n_2-1))$

If diff. of two population Mean is zero

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

where $$\sigma^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

or $$\sigma^2 = \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2 + (x_{2i}-\bar{x}_2)^2}{n_1+n_2-2}$$

If diff of two population Mean is nonzero (say δ)

$$t = \frac{(\bar{x}_1-\bar{x}_2)-\delta}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

Observed correlation coeffi.:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

r is observed correlation coeffi.
& ρ – population correlation coeffi

Here Hypo: $H_0: \rho = 0$ no correlation betⁿ variate
$H_1: \rho \neq 0$ or $\rho > 0$ or $\rho < 0$

Degree of freedom is $v = n-2$ for this test

Handout-4 Date: 31/9/19

F - test and χ^2 test

Objective: To learn about f-test & χ^2 test for any size of sample

F - test: Variance ratio

ratio of the variances of two Indepl. random sample is

$$F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$$

with Degree of freedom $v_1 = n_1 - 1$ & $v_2 = n_2 - 1$

n_1 – sample x

n_2 – another sample y

s_1^2 – Variance of sample x

s_2^2 – Variance of another sample y

If variance is same

$$F = \frac{s_1^2}{s_2^2}$$

If not given s_1 & s_2 calculate $\rightarrow$

$$\begin{cases} s_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\ s_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 \end{cases}$$

Note: F is always a +ve no.

χ^2 test: [Non parametric Test]

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

where O :- observed frequency

E :- Expected frequency

How to find value of χ^2

Step 1: calculate expected frequencies

$$E = \frac{RT \times CT}{N}$$

where E – Expected frequency

RT – The row total for the row containing the cell

CT – The column total for the col. containing cell

N – Total No. of observations

Step 2: Find $(O-E)^2$

Step 3: Find $\sum \frac{(O-E)^2}{E} = \chi^2$ It can be bet^n 0 to ∞

If it is zero then O & E coincides

Step 4:
Conclusion: compare calculated value of χ^2 with table value of χ^2 for given degree of freedom $(\nu = n-k)$ at a certain LOS (Level of significance)

- If calculated value of χ^2 is more than table value of χ^2, the difference bet^n theory & observation is considered as significant.
- If calculated value of χ^2 is less than the table value of χ^2, not considered as significant

Use: It is use for test of goodness of fit

Note: How to calculate E (expected frequency) from given observed O (frequency)?

eg

Given known as O ← observed freq

Treatment	Fever	No fever	Total
Quinine	40	1584	1624 → col. total
No quinine	440	4432	4872
Total	480	6016	6496

Row total → Row 1 + Row 2
40 + 440 = 480

- Now $E = \frac{RT \times CT}{N}$ RT = 480 or 6016
CT = 1624 or 4872
- ~~select~~ calculate E for first Row first col position
- so ~~sel~~ for that select RT = 480 & CT = 1624 as intersection position of first Row, first col, so

$E = \frac{480 \times 1624}{N} = 120$. ∴ Expected fre. E =

120	1504	1624
360	4512	487[illegible]

$(1504 = 1624 - 120)$