Unit-3

Estimation and Testing of Hypothesis

Statistical Inference: The process of drawing inference about population based on the samples data is called statistical inference.

Statistical inferences are of 2 types

1) Estimation
2) Testing of Hypothesis

Estimation: Estimation, the process to obtain the values of unknown population parameter with the help of sample data is called estimation.

Estimate: Estimate is numerical value of Estimation (or) An estimate is a statement made to find an unknown population parameter

Estimator: Estimator is the rule/formula/function that tells us how to estimate.

Types of Estimation

1) **Point Estimation:** When an estimate for the unknown population parameter is expressed by a single value. It is called point estimate

Ex: Weight of all students of cmrit is 40 kgs

2) **Interval Estimation:** When an estimate about unknown population parameter is expressed by a range of values within which the population parameter is expected to occur is called interval estimation

Ex: Weight of students of cmrit is 40 ± 20 kgs

Properties of Estimation:

1) An Estimator is not expected to estimate the population parameter without error.
2) An Estimator should be close to the value of unknown population parameter

Point Estimation:

Note: A point estimate of a parameter θ is a single numerical value which is computed from a given Sample and Serves as an approximation of the unknown exact value of parameter 'θ'.

Parameter: The Statistical Constants like mean, S.D, Co-relation Coefficient etc obtained for the population are Called Parameters.

Statistics: Constants for Sample mean($\bar{x}$), Variance (S^2), S.D (S) are Called Statistics

Difference b/w Parameter & Statistics

Parameter	Statistics
1) parameter is a Statistical measure based on all units of Population	1) Statistics is a stastical measure based on all units of Sample
2) μ, σ^2 are parameters	2) $\bar{x}$, S^2 are all Statistics
3) parameter refers to population	3) Statistics refers to Sample.

Unbiased Estimator: A Statistic is Said to be an unbiased estimator of the Corresponding parameter if the mean of Sampling distribution of the Statistic is equal to the Corresponding population parameter (or) A Statistic or point Estimator ($\hat{\theta}$) is Said to be an Unbiased Estimator of θ if $E(\hat{\theta})=\theta$. That means

E(Statistic) = parameter

Interval estimation: By using point estimation we may not get derived degree of accuracy in estimating a parameter. Therefore it is better to replace point estimation (Single numerical value) by interval estimation i.e an estimate of a population parameter given by two numbers between which the population parameter may be supposed to lie, with a reasonable degree of certainity

definition: An interval estimate of an unknown parameter θ is an interval of the form $L \le \theta \le U$ where the endpoints L and U depend on the numerical value of the statistic $\hat{\theta}$ for a particular sample on the Sampling distribution.

Imp The advantage of an interval estimate over a point estimate is that the interval estimate is Constructed in Such a way that we can assess the confidence that the interval Contains the parameter. For this reason interval estimates are called Confidence intervals.

Hypothesis and hypothesis testing

We can find either a single number for the parameter (a point estimate) or an interval of values (an interval estimate) However there are many problems in which rather than estimating the value of a parameter we need to decide whether to accept or reject a Statement is called "hypothesis" and the decision making procedure about the hypothesis is called "hypothesis testing". In other words A "statistical hypothesis" is a statement about the parameter of one or more populations.

Types of hypothesis:

There are two types of hypothesis

1) Null hypothesis
2) Alternate hypothesis

Null hypothesis: For applying tests of significance we first set up hypothesis - a definite Statement about population parameter such a hypothesis is usually a hypothesis of no difference called null hypothesis. It is denoted by H_0.

Ex: $H_0 : \mu = \mu_0$

Confidence interval: $(1-\alpha)100\%$ is confidence interval.

The quantities L and U are called the lower and upper confidence limits respectively and $1-\alpha$ is called the Confidence coefficient or degree of Confidence.

properties of Good estimator: An estimator is said to be good estimator if it is

1) Unbiased (no partiality)

2) Consistent

3) efficient and sufficient

Unbiased estimator: An estimator $\hat{\theta}$ is said to be unbiased if mean of the sampling distribution of estimator is equal to θ i.e $E(\hat{\theta}) = \theta$

Consistency: If an estimator $\hat{\theta}$ approaches to parameter θ closer and closer as sample size n increases the $\hat{\theta}$ is consistent estimator of θ.

Most efficient: If ① $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators of θ.

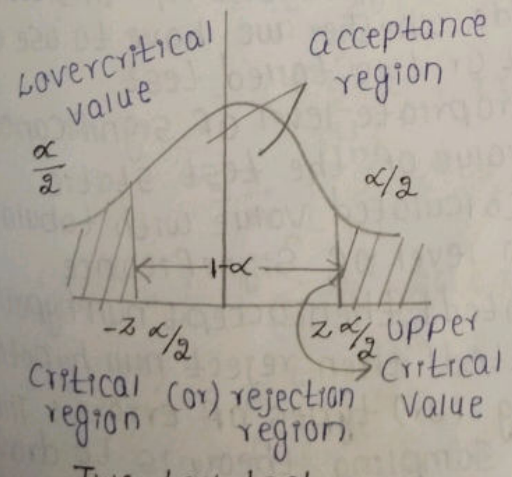
2) σ_1^2, σ_2^2 are two variances of sampling distributions and $\sigma_1^2 < \sigma_2^2$ then $\hat{\theta}_1$ is said to be most efficient unbiased estimator of θ.

Alternative hypothesis: Any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis. It is denoted by H_1.

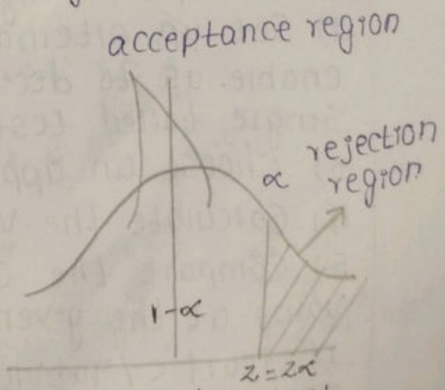
Types of alternative hypothesis: If we want to test null hypothesis that the population has a specified mean μ_0 (say) i.e $H_0 : \mu = \mu_0$ then alternative hypothesis would be

1) Two tailed alternative $H_1 : \mu \neq \mu_0$
2) right tailed alternative $H_1 = \mu > \mu_0$
3) left tailed alternative $H_1 : \mu < \mu_0$

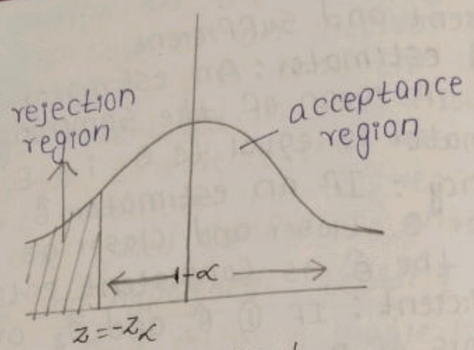
The setting of alternative hypothesis is very important to decide whether we have to use single tailed test (left or right) or two tailed test.



Two tail test



right tail test



left tail test

Critical region or rejection region: A region in the sample space which leads to rejection of null hypothesis H_0 is called critical region or rejection region.

Acceptance region: A region in the sample space which leads to acceptance of null hypothesis H_0 is called acceptance region.

Critical value or Significant value: The value of the statistic which seperates the critical region (rejection region) and acceptance region is called critical value or significant value.

This value depends on

1) level of significance used and
2) alternative hypothesis one tailed or two tailed test

Procedure for Testing of hypothesis:

1) Set up null hypothesis H_0
2) Set up alternative hypothesis H_1. This will enable us to decide whether we have to use a single tailed test or two tailed test.
3) Choose an appropriate level of significance 'α'.
4) Calculate the value of the test static.
5) Compare the calculated value with tabulated value at the given level of significance

If $|cal| < |Tabulated|$ then accept null hypothesis

If $|cal| > |Tabulated|$ then reject null hypothesis

errors in sampling (or) types of errors: The main objective in sampling theory is to draw valid inferences about the population parameters

on the basis o
we decide to a
examining a s
two types of

Corr

H_0 is true
H_0 is false

Type I error:
it is type I
probability of
$(1-\alpha)100 =$

Type II error
it is type II
P(Type II
α and β ar
(or) producer

Theorem: Pr
size n x_1, x_2
population

estimator
unbiased e
proof: s^2

$s^2 = \frac{1}{n}$
$E(s^2)$

ce

: A region in
ejection of null
egion or rejec-

he sample
f null hypothesis

ue : The value
the critical
ptance region
ficant value.

d or two tailed

sis :

H_1 . This will
have to use a
test .
f significance 'α'.
st static .
e with tabulated
nificance
ept null hypothesis
ect null hypothesis
f errors : The
ory is to draw
lation parameters

on the basis of sample results. In practice we decide to accept or to reject the lot after examining a sample from it. As such there are two types of errors

	Accept	reject
H_0 is true	Correct decision	Type I error
H_0 is false	Type II error	Correct decision

Type I error: If H_0 is true, but we reject then it is type I error.

probability of Type I error = P(Type I error) $= \alpha$

$(1-\alpha)100$ = Confidence.

Type II error: If H_0 is false, but we accept then it is type II error.

P(Type II error) $= \beta$

α and β are called sizes of Type I, Type II errors (or) producer's risk, consumer's risk respectively.

Theorem: Prove that for a random sample of size n $x_1, x_2 \ldots x_n$ taken from an infinite population $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ is not unbiased estimator of population parameter σ^2 but $\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ unbiased estimator of σ^2

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

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

$$= \frac{1}{n}\left[\sum_{i=1}^{n} x_i^2 - 2\bar{x}_1\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\bar{x}^2\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} x_i^2 - 2\bar{x}\, n \frac{\sum_{i=1}^{n} x_i}{n} + \bar{x}^2\sum_{i=1}^{n} 1\right]$$

$$s^2 = \frac{1}{n}\left[\sum_{i=1}^{n} x_i^2 - 2n\bar{x}\bar{x} + \bar{x}^2(n)\right] = \frac{1}{n}\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right]$$

$$E(s^2) = \frac{1}{n}E\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) \quad \text{— (1)}$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} E(x_i^2) - nE(\bar{x}^2)\right]$$

$V(x_i) = E(x_i^2) - [E(x_i)]^2$ $\quad \sigma^2 = \sum P_i x_i - u^2 = E(x_i^2) - [E(x_i)]^2$

$\sigma^2 = E(x_i^2) - u^2$ — ② $\quad (\because E(x_i) = u)$

$\therefore E(x_i^2) = \sigma^2 + u^2$ — ②

$V(\bar{x}) = E(\bar{x}^2) - E(\bar{x})^2 \quad \left(\because Var(\bar{x}) = Var\left(\frac{\sum x_i}{n}\right)\right)$

$\frac{\sigma^2}{n} = E(\bar{x}^2) - u^2$

$E(\bar{x}^2) = \frac{\sigma^2}{n} + u^2$ — ③

Sub ② and ③ in ①

$E(S^2) = \frac{1}{n}\left[\sum_{i=1}^{n}(\sigma^2 + u^2) - n\left(\frac{\sigma^2}{n} + u^2\right)\right]$

$= \frac{1}{n}\left[(n)(\sigma^2 + u^2) - \sigma^2 - nu^2\right]$

$= \frac{1}{n}\left[n\sigma^2 + nu^2 - \sigma^2 - nu^2\right]$

$E(S^2) = \frac{1}{n}\left[(n-1)\sigma^2\right] \neq \sigma^2$

$\therefore S^2$ is not unbiased estimator of σ^2

2) $s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]$

$= \frac{1}{n-1} \cdot n\left[\sum_{i=1}^{n}\frac{(x_i - \bar{x})^2}{n}\right] = \frac{n}{n-1} E\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)$

$= \frac{n}{n-1} \cdot S^2$

$E(s^2) = E\left(\frac{n}{n-1} \cdot S^2\right)$

$= \frac{n}{n-1} E(S^2)$

$= \frac{n}{n-1} \cdot \frac{(n-1)}{n}\sigma^2$

$E(s^2) = \sigma^2$

$\therefore s^2$ is unbiased estimator of σ^2

1. If x is a random variable and k is a constant then $E(x+k) = E(x) + k$

$E(x) = \sum P_i x_i$ where $x_i \in X$

$E(x+k) = \sum P_i (x_i + k)$

$= \sum P_i x_i + \sum P_i k = E(x) + k\sum P_i = E(x) + k(1)$

$= u + k \quad (\because \sum P_i = 1)$

$E(x+k) = E(x) + k$

2) $E(k) = k$

proof: E(

$E(k) = \sum$

$\therefore E(k)$

3) $E(kx) = \sum$

$= k$

$E(kx) = k$

4) If x is a

constants

proof: E(

E(

$\therefore$ E(

Bayesian

in Bayes

probability

variable in

of populat

whose dis-

feelings' or

possible v

definition

with the

u with d

Sample evi

in Bayes

normal dis

$u_1 = n$

Here u_1 a

the posteri

u_1 and σ_1

σ^2 is unk

variance.

2) $E(k) = k$

proof: $E(x) = \sum P_i x_i$ where $x_i \in X$

$E(k) = \sum P_i k = k \sum P_i = k(1) = k \quad (\because \sum P_i = 1)$

$\therefore E(k) = k$

3) $E(kx) = \sum P_i (k x_i)$

$= k \sum P_i x_i$

$E(kx) = k E(x)$

$E(kx) = k E(x)$

4) If x is a random variable and a and b are constants then $E(ax+b) = aE(x) + b$

proof: $E(x) = \sum P_i x_i$

$E(ax+b) = \sum P_i (a x_i + b)$

$= \sum P_i a x_i + \sum P_i b$

$= a \sum P_i x_i + b \sum P_i$

$= aE(x) + b(1) \quad (\because \sum P_i = 1)$

$\therefore E(ax+b) = aE(x) + b$

Bayesian estimation: The new concept introduced in Bayesian methods is personal or subjective probability. parameters are considered as random variable in Bayesian method. To estimate mean of population u is treated as random variable whose distribution is indicative of "strong feelings" or "assumption of a person about the possible values of u.

definition of Bayesian estimation: Combining with the prior feelings about possible values of u with direct possible values of u with direct sample evidence, the "posterior" distribution of u in Bayesian estimation is approximated by normal distribution with

$$u_1 = \frac{n\bar{x}\sigma_0^2 + u_0\sigma^2}{n\sigma_0^2 + \sigma^2} \quad \text{and} \quad \sigma_1 = \sqrt{\frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}}$$

Here u_1 and σ_1 are known as mean and S.D of the posterior distribution. In the computation of u_1 and σ_1, σ^2 is assumed to be known, when σ^2 is unknown it is replaced by sample variance. s^2 provided $n \geq 30$

Bayesian interval: A $(1-\alpha)100\%$ Bayesian interval μ is given by

$(\mu_1 - z_{\alpha/2}\sigma_1 < \mu < \mu_1 + z_{\alpha/2}\sigma_1)$

A professor's feelings about the mean mark in final examination in "probability" of a large group of students is expressed subjectively by normal distribution with $\mu_0 = 67.2$ and $\sigma_0 = 1.5$

a) If the mean mark lies in the interval (65, 70) determine the prior probability the professor should assign to the mean mark

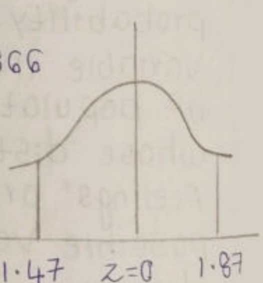
b) Find the professor posterior mean μ_1 and posterior S.d σ_1 if the examinations are conducted in a random sample of 40 students yielding mean 74.9 and S.D 7.4. Use 7.4 as an estimate σ

c) Determine the posterior probability which he will thus assign to the mean mark being in the interval (65, 70) using results obtained in (b)

$\mu_0 = 67.2$, $\sigma_0 = 1.5$, $n = 40$

a) when $x = 65 \Rightarrow z_1 = \frac{x - \mu_0}{\sigma_0} = \frac{65 - 67.2}{1.5} = -1.466$

when $x = 70 \Rightarrow z_2 = \frac{70 - 67.2}{15} = 1.866$



prior probability

$P(65 < x < 70) = P(-1.47 < z < 1.87)$

$= 0.4292 + 0.4693$

$= 0.8985$ (prior probability)

b) $\bar{x} = 74.9$ $\sigma = s = 7.4$

Posterior mean $\mu_1 = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$

$\therefore \mu_1 = \frac{(40)(74.9)(1.5)^2 + (67.2)(7.4)^2}{40(1.5)^2 + (7.4)^2}$

$\mu_1 = 71.987 \simeq 72$

Posterior Standard deviation

$\sigma_1 = \sqrt{\frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2}} = \sqrt{\frac{(7.4)^2(1.5)^2}{40(1.5)^2}} = 0.9225 \simeq 0.923$

c) Here $\mu_1 = 72$, $\sigma_1 = 0.923$

when $x =$

$x =$

$\therefore$ poster

$=$

$=$

$=$

d) 95%.

$(\mu_1 - z$

$=$

The mea
entrance
variatio
distribu
1) What
mean m
for the
2) Cons
the te
100 stu
yielding
3) wha
to the
$\mu_0 =$

a) $x =$

$x =$

Let x
comm
prior

when $x = 65 \Rightarrow z_1 = \dfrac{65-72}{0.923} = -7.5839$

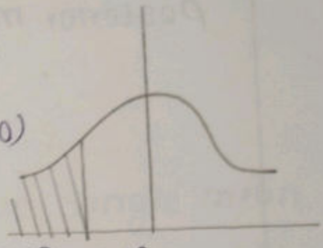
$x = 70 \Rightarrow z_2 = \dfrac{70-72}{0.923} = -2.166$

$\therefore$ posterior probability $P(65 < x < 70)$

$= P(-7.5839 < x < -2.167)$

$= 0.5 - 0.4850$

$= 0.0150$



-7.58 -2.16

d) 95% Bayesian interval limits are

$(\mu_1 - z_{\frac{\alpha}{2}}\sigma_1 , \mu_1 + z_{\frac{\alpha}{2}}\sigma_1) = (71.987 - 1.96(0.923), 71.987 + 1.96(0.923))$

$= (70.178, 73.795)$

The mean mark in mathematics in common entrance that will vary from year to year. If this variation of mean mark is subject by a normal distribution with mean $\mu_0 = 72$ and variance $\sigma_0^2 = 5.76$

1) What probability can we assign to the actual mean mark being somewhere b/w 71.8 and 73.4 for the next years test?

2) Construct a 95% Bayesian interval for μ if the test is conducted for a random sample of 100 students from the next incoming class yielding a mean mark of 70 with S.D of 8.

3) What posterior probability should we assign to the event of part 1

$\mu_0 = 72$, $\sigma_0 = \sqrt{5.76} = 2.4$, $n = 100$

a) $x = 71.8 \Rightarrow z = \dfrac{x - \mu_0}{\sigma_0} = \dfrac{71.8 - 72}{2.4} = -0.0833$

$x = 73.4 \Rightarrow z = \dfrac{x - \mu_0}{\sigma_0} = \dfrac{73.4 - 72}{2.4} = 0.58333$

Let x be the mean mark obtained in the common entrance test. Then

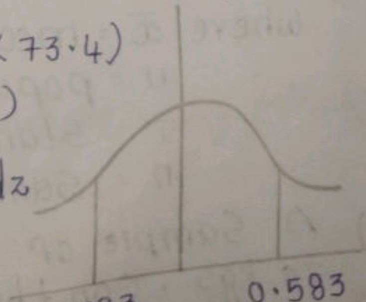
Prior probability $P(71.8 < x < 73.4)$

$= P(-0.083 < x < 0.5833)$

$= \int_{-0.083}^{0} \phi(z)\,dz + \int_{0}^{0.583} \phi(z)\,dz$

$= 0.0319 + 0.2190$

$= 0.2509$



-0.083 0.583

2) $n=100,\ \bar{x}=70,\ \sigma_0=2\cdot4,\ u_0=72,\ \sigma=8$

Posterior mean $u = \dfrac{n\bar{x}\sigma_0^2 + u_0\sigma^2}{n\sigma_0^2+\sigma^2}$

$= \dfrac{(100)(70)(2\cdot4)^2 + 72\,(8^2)}{100\,(2\cdot4)^2+8^2}$

$= 70\cdot2$

Posterior standard deviation

$\sigma_1 = \sqrt{\dfrac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}}$

$= \sqrt{\dfrac{8^2(2\cdot4)^2}{100(2\cdot4)^2+8^2}} = 0\cdot7589 \simeq 0\cdot76$

$\therefore$ 95% Bayesian interval for u is given by

$u_1 \pm Z_{\frac{\alpha}{2}}\sigma_1 = 70\cdot2 \pm (1\cdot96)(0\cdot7589)$

$= (68\cdot71,\ 71\cdot69)$

$u_1 = 70\cdot2,\ \sigma_1 = 0\cdot76$

when $x = 68\cdot71 \Rightarrow z = \dfrac{68\cdot71-70\cdot2}{0\cdot76} = -1\cdot960$

when $x = 71\cdot69 \Rightarrow z = \dfrac{71\cdot69-70\cdot2}{0\cdot76} = 1\cdot960$

posterior probability $= P(68\cdot71 < x < 71\cdot69)$

$= P(-1\cdot960 < x < 1\cdot960) = \int_{-1\cdot960}^{0}\phi(z)dz + \int_{0}^{1\cdot960}\phi(z)dz$

$= \int_{0}^{1\cdot960}\phi(z)dz + \int_{0}^{1\cdot960}\phi(z)dz = 0\cdot4750 + 0\cdot4750 = 0\cdot95$

11/11 Test of Significance for large Samples

Single mean:

Test Statistic $= Z_{cal} = \dfrac{\bar{x}-u}{\frac{\sigma}{\sqrt{n}}}$

where $\bar{x}$ = mean of the Sample
u = population mean
σ = Standard deviation
n = Sample Size

1) A Sample of 64 students have mean weight 70 kgs. Can this be regarded as a Sample from a Population with mean weight 56 kgs

and S.D 2
$n = 64$ (lar
$\bar{x}$ = Sampli
u = popula
S.D of po
$\therefore$ The prob
problem
null hypoth
alternate h
level of S
Test Stati
$|Z_{cal}$
$Z_{tab} = Z_{\frac{\alpha}{2}}$
$|Z_{cal}| > Z_t$
reject null
accept alte
$\therefore u \neq 5$

2) An oceanog
of ocean in
as hadd bein
Conclude a
readings to
given regio
with a S.D
$\alpha = 0\cdot05$
$n = 40$ (larg
$\bar{x} = 59\cdot1$
Given pro
Null hypot
alternate
level of S

and S.D 25 kgs

$n = 64$ (large sample $\because n \geq 30$)

$\bar{x}$ = Sampling mean = 70 kgs

μ = population mean = 56 kgs

S.D of population = σ = 25 kgs

$\therefore$ The problem is large sample single mean problem

null hypothesis $H_0 : \mu = 56$ (given value)

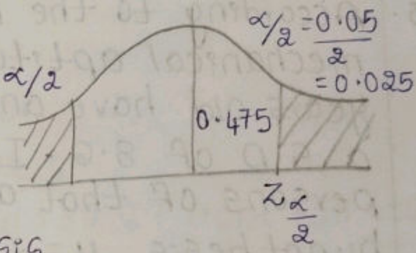
alternate hypothesis $H_1 : \mu \neq 56$ (two tail test)

level of significance $\alpha = 0.05$ (standard LOS)

Test Statistic : $Z_{cal} = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{70 - 56}{\frac{25}{\sqrt{64}}} = 4.48$

$|Z_{cal}| = |4.48| = 4.48$ — ①

$Z_{tab} = Z_{\frac{\alpha}{2}} = 1.96$ — ②



$|Z_{cal}| > Z_{tab}$

reject null hypothesis

accept alternate hypothesis

$\therefore \mu \neq 56$

8 An oceanographer wants to check wheter depth of ocean in a certain region is 57.4 fathoms as hadd being previously recorded. What can he conclude at 0.05 level of significance if the readings taken at 40 random location in the given region yielded with a mean of 59.1 fathoms with a S.D of 5.2 fathoms

$\alpha = 0.05$ $S/\sigma = 5.2$

$n = 40$ (large) $\mu = 57.4$

$\bar{x} = 59.1$

Given problem is large sample single mean

Null hypothesis $H_0 : 57.4 = \mu$

alternate hypothesis $H_1 : \mu \neq 57.4$

level of significance = 0.05

Test Statistic : $Z_{cal} = \dfrac{\bar{x} - u}{\frac{\sigma}{\sqrt{n}}} = \dfrac{\bar{x} - u}{\frac{s}{\sqrt{n}}}$

$= \dfrac{59.1 - 57.4}{\frac{5.2}{\sqrt{40}}}$

$= 2.067$

$|Z_{cal}| = |2.067| = 2.06$ — (1)

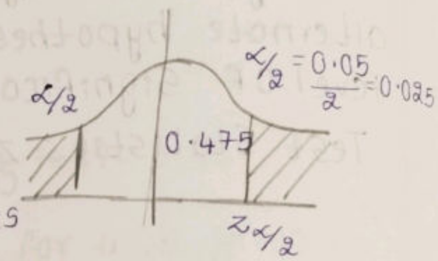
$Z_{tab} = Z_{\frac{\alpha}{2}} = 1.96$ — (2)

$|Z_{cal}| > Z_{tab}$

reject null hypothesis.

Accept alternate hypothesis

$\therefore u \neq 57.4$



3. According to the norms established for a mechanical aptitude test persons who are 18 years old have an average height of 73.2 with a S.D of 8.6. If fourty randomly selected persons of that age averaged 76.7. Test the hypothesis $u = 73.2$ against alternate hypothesis $u > 73.2$ at 0.01 level of significance.

$n = 40$ (large sample) > 30

mean $= u = 73.2$

$\sigma = 8.6$

$\bar{x} = 76.7$

$\alpha = 0.01$

Given problem is large sample single mean

null hypothesis $H_0 : u = 73.2$

Alternate hypothesis $H_1 : u > 73.2$ (right tail test)

level of significance $\alpha = 0.01$

Test Statistic : $Z_{cal} = \dfrac{\bar{x} - u}{\frac{\sigma}{\sqrt{n}}} = \dfrac{76.7 - 73.2}{\frac{8.6}{\sqrt{40}}}$

$= 2.57$

$|Z_{cal}| = |2.57| = 2.57$ — (1)

Z_{tab} = 2

$|Z_{cal}|$ >

reject nu

accept al

∴ u

4. In a ra
time take
with a s
hypothesis
hypothesi

$n = 60$

$\bar{x} = 33$

$s = 6$

$u = 32$

$\alpha = 0.$

large s

Null hy

alternat

Test s

$|Z_{cc}$

Z_{tab}

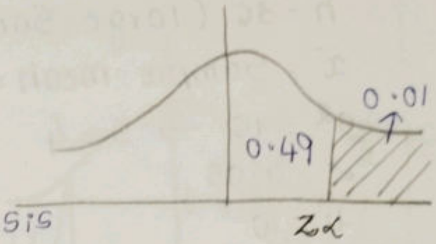
Z_{tab}

accept

u

5. An amb
on avg
ation in
as a me

$z_{tab} = z_\alpha = 2.33$ — (2)



$|z_{cal}| > z_{tab}$

reject null hypothesis

accept alternate hypothesis

$\therefore \mu \neq 73.2$

4. In a random sample of 60 workers, the avg time taken by them to get to work is 33.8 with a S.D of 6.1. Can we reject the null hypothesis $\mu = 32.6$ min in favour of alternate hypothesis >32.6 at $\alpha = 0.025$ level of Significance.

$n = 60$ (large Sample)

$\bar{x} = 33.8$

$s = 6.1$

$\mu = 32.6$

$\alpha = 0.025$

large Sample Single mean

Null hypothesis $H_0 : \mu = 32.6$

alternate hypothesis $H_1 : \mu > 32.6$ (right tail test)

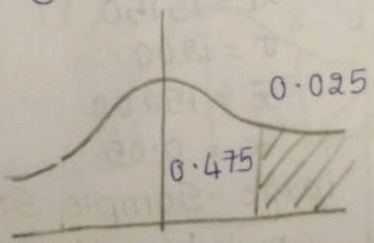
Test Statistic : $z_{cal} = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$

$= \dfrac{33.8-32.6}{\frac{6.1}{\sqrt{60}}} = 1.523$

$|z_{cal}| = |1.521| = 1.52$ — (1)

$z_{tab} = z_\alpha = 1.96$ — (2)

$z_{tab} > |z_{cal}|$



accept null hypothesis

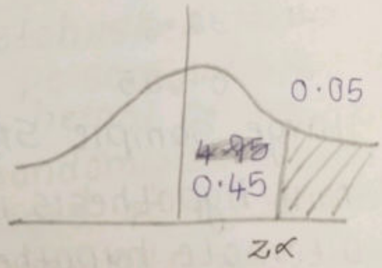
$\mu = 32.6$

5. An ambulance service claims that it takes on avg less than 10 min to reach its destination in emergency calls. A sample of 36 calls as a mean of 11 min and variance of 16 min test the Calaim at 0.05 level of Significance

$n = 36$ (large sample)
$\bar{x}$ = Sample mean = 11
$s^2 = 16 \rightarrow s = 4$
$\alpha = 0.05$
$\mu = 10$
large Sample Single mean
null hypothesis $H_0 : \mu = 10$
alternate hypothesis $H_1 : \mu > 10$ (right tail)
Test Staticstic : $Z_{cal} = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$

$= \dfrac{11-10}{\frac{4}{\sqrt{36}}} = 1.5$

$|Z_{cal}| = |1.5| = 1.5$
$Z_{tab} = Z_\alpha = 1.65$
null hypothesis is accepted
$\mu = 10$



6. It is claimed that a random Sample of 49 tyres has mean life of 15200 kms. This Sample was drawn from a population whose mean of 15150 kms and a S.D 1200 kms. Test the Significance at 0.05 level

$n = 49$
$\mu = 15150$
$\sigma = 1200$
$\bar{x} = 15200$
$\alpha = 0.05$

large Sample Single mean
null hypothesis $H_0 : \mu = 15150$
alternate hypothesis $H_1 : \mu \neq 15150$ (two tail)
Test Statistic : $Z_{cal} = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{15200-15150}{\frac{1200}{\sqrt{49}}}$

$= 0.29$

$|Z_{cal}| = |0.291| = 0.29$ — (1)

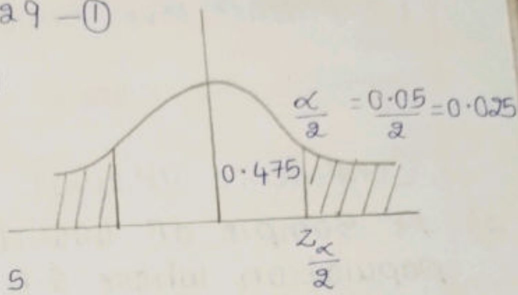
$Z_{tab} = Z_{\frac{\alpha}{2}} = 1.96$ — (2)

$|Z_{cal}| < Z_{tab}$

$Z_{tab} > |Z_{cal}|$

accept null hypothesis

$\mu = 15150$



(4/III) A sample of 900 students has a mean of 3.4 cm & S.D 2.61 cm. Is this sample has been taken from a large population of mean 3.25 cm & S.D 2.61 cm. If the population is normal & it's mean is unknown. Find 95% fudicial limits (confidence limits) of true mean (standard error, max error, confidence limits & confidence interval)

$n = 900 \quad \bar{x} = 3.4 \quad s = 2.61$

$\mu = 3.25 \quad \sigma = 2.61$

Given problem is large sample single mean

Null hypothesis $H_0 : \mu = 3.25$

Alternate hypothesis $H_1 : \mu \neq 3.25$ (Two tail)

$\alpha = 0.05$

Test Statistics : $Z_{cal} = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{3.4 - 3.25}{\frac{2.61}{\sqrt{900}}}$

$= 1.724$

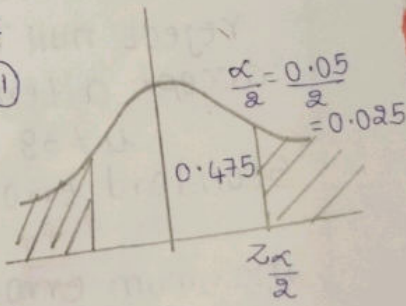
$|Z_{cal}| = |1.721| = 1.72$ — (1)

$Z_{tab} = Z_{\frac{\alpha}{2}} = 1.96$ — (2)

$Z_{tab} > |Z_{cal}|$

accept null hypothesis

$\mu = 3.25$



Standard error $= \dfrac{\sigma}{\sqrt{n}} = \dfrac{2.61}{\sqrt{900}} = 0.087$

Maximum error $= Z_{\frac{\alpha}{2}} \times S.E = 1.96 \times 0.087$

$= 0.17052$

Confidence level limits $= (\bar{x} - m \cdot E), (\bar{x} + m \cdot E)$

$= (3.4 - 0.17), (3.4 + 0.17)$

$= (3.23, 3.57)$

Confidence interval $= (3.23, 3.57)$

2) A Sample of 400 items is taken from a population whose S.D is 10. The mean of the Sample is 40. Test whether the Sample has come from a population with mean 38 also Calculate 95% confidence interval

$n = 400$ $\quad \sigma = 10$ $\quad \bar{x} = 40$ $\quad \mu = 38$

Given problem is large sample Single mean

Null hypothesis $(H_0) : \mu = 38$

Alternate hypothesis $(H_1) : \mu \neq 38$ (Two tail)

level of significance $(\alpha) = 1 - \alpha = \frac{95}{100} = 0.95$

$= 1 - 0.95 = \alpha$

$\alpha = 0.05$

Test statistic: $Z_{cal} = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{40 - 38}{\frac{10}{\sqrt{400}}}$

$= 4$

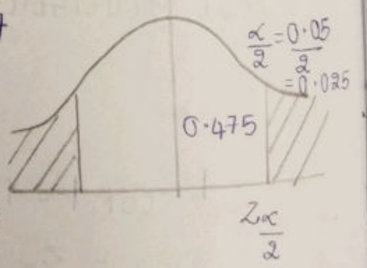
$Z_{tab} = Z_{\frac{\alpha}{2}} = 1.96$ & $Z_{cal} = 4$

$Z_{cal} > Z_{tab}$

reject null hypotesis

accept alternate hypothesis

$\mu \neq 38$



Standard error $= \dfrac{\sigma}{\sqrt{n}} = \dfrac{1}{2} = 0.5$

maximum error $= Z_{\frac{\alpha}{2}} \times S.E = 1.96 \times 0.5 = 0.98$

Confidence limits $= (\bar{x} - ME, \bar{x} + m \cdot E)$

$= (40 - 0.98, 40 + 0.98)$

$= 39.02, 40.98$

Confidence interval $= (39.02, 40.98)$

Testing of Hypothesis For large Sample difference of means

Test statics for large sample of difference of means $z_{cal} = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

In case if 2 samples are drawn from two populations with sample S.D as s_1 & s_2 in that case the test statistic is $\dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

If the difference b/w the means is given then test statistic is $\dfrac{\bar{x}_1 - \bar{x}_2 - \Delta}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

1) The means of 2 samples of sizes 1000 & 2000 members are 67.5 inches and 68 inches respectively. Can this be regarded as drawn from the same population of S.D 2.5 inches.

$n_1 = 1000$ $\quad n_2 = 2000$ $\quad \sigma_1 = 2.5$

$\bar{x}_1 = 67.5$ $\quad \bar{x}_2 = 68$ $\quad \sigma_2 = 2.5$

The given problem is large sample difference of means

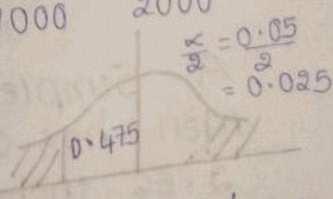
Null hypothesis $H_0 : \mu_1 = \mu_2$

Alternate hypothesis $H_1 : \mu_1 \neq \mu_2$ (Two tail)

level of significance $\alpha = 0.05$

$$z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{67.5 - 68}{\sqrt{\dfrac{(2.5)^2}{1000} + \dfrac{(2.5)^2}{2000}}}$$

$|z_{cal}| = |-5.16| = 5.16$



$z_{tab} = z_{\frac{\alpha}{2}} = 1.96$

$z_{cal} > z_{tab}$

reject null hypothesis, accept alternate hypothesis, $\mu_1 \neq \mu_2$

∴ The 2 samples are not drawn from same population

The research investigator is interested in studying whether there is a significant difference in the salaries of MBA Grades in 2 metropolitan cities. A random sample of size 100 from mumbai yields an avg income of ₹20150 another random sample of 60 from chennai results in an avg income of ₹ 20250. If the variance of both the populations are given as $\sigma_1^2 = 40000$ & $\sigma_2^2 = 32400$ respectively.

$n_1 = 100 \quad \bar{x}_1 = 20150 \quad \sigma_1^2 = 40000$

$n_2 = 60 \quad \bar{x}_2 = 20250 \quad \sigma_2^2 = 32400$

Large sample difference of means

Null hypothesis $H_0 : \mu_1 = \mu_2$

Alternate hypothesis $H_1 : \mu_1 \neq \mu_2$, $\alpha = 0.05$

Test Statistics $Z_{cal} = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

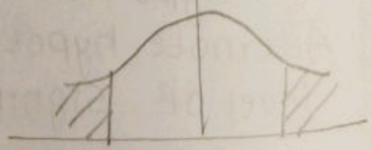
$= \dfrac{\cancel{100 - 60}\ 20150 - 20250}{\sqrt{\dfrac{40000}{100} + \dfrac{32400}{60}}}$

$= -3.26$

$|Z_{cal}| = |-3.26| = 3.26$

$Z_{tab} = Z_{\frac{\alpha}{2}} = 1.96$

$Z_{cal} > Z_{tab}$



reject null hypothesis

accept alternate hypothesis

$\mu_1 \neq \mu_2$

A simple sample of height of ~~60~~400 English men has a mean of 67.85 inches and a S.D 2.56 inches while a simple sample of heights of 1600 Australians has a mean of 68.55 inches

& S.D of 2.52 inches. Do the data indicate that Australians are on the avg taller than the English men. Use $\alpha = 0.01$

$n_1 = 6400$ $\quad n_2 = 1600$

$\bar{x}_1 = 67.85$ $\quad \bar{x}_2 = 68.55$

$s_1 = 2.56$ $\quad s_2 = 2.52$ $\quad \alpha = 0.01$

large sample difference of means

Test Statistics $z_{cal} = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

Null hypothesis $H_0 : \mu_1 = \mu_2$

Alternate hypothesis $H_1 : \mu_1 < \mu_2$ (Australians are taller than English) (Left tail)

$$z_{cal} = \frac{67.85 - 68.55}{\sqrt{\dfrac{(2.56)^2}{6400} + \dfrac{(2.52)^2}{1600}}}$$

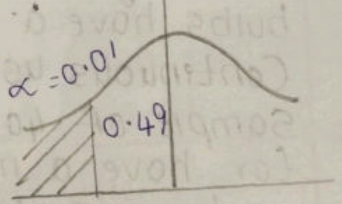
$|z_{cal}| = |-9.906| = 9.90$

$z_{tab} = z_\alpha = 2.32$

$z_{tab} < z_{cal}$

reject null hypothesis

accept alternate hypothesis



$\therefore$ Australians are taller than English men.

The avg marks scored by 32 boys is 72 with a S.D of 8 while that for 36 girls is 70 with a S.D of 6. Does this indicate that the boys perform better than girls at level of significance 0.05

$n_1 = 32$ $\quad n_2 = 36$

$\bar{x}_1 = 72$ $\quad \bar{x}_2 = 70$

$s_1 = 8$ $\quad s_2 = 6$ $\quad \alpha = 0.05$

large sample difference of means

Null hypothesis $H_0 : \mu_1 = \mu_2$

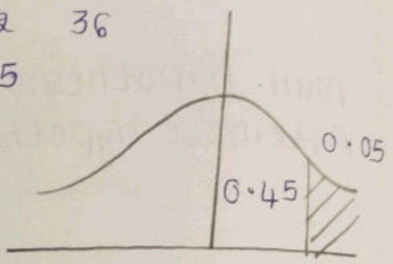
alternate hypothesis $H_1 : \mu_1 > \mu_2$ (Boys perform better than girls)

Test statistics $$Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{72-70}{\sqrt{\frac{(8)^2}{32} + \frac{(6)^2}{36}}}$$

$$Z_{cal} = 1.15$$

$Z_{tab} = Z_\alpha = 0.05 = 1.64$



$Z_{tab} > Z_{cal}$

accept null hypothesis

$\mu_1 = \mu_2$

Both boys & girls perform equally.

A Company Claims that its bulbs are superior to those of its main Competetor. If a study showed that a sample of 40 of its bulbs have a mean life of 647 hours of Continuous use with a S.D of 27 hrs. While a sample of 40 bulbs may buy its main Competetor have a mean life of 638 with a S.D of 31 hrs. Test the Significance b/w the difference of 5% level

$n_1 = 40$ $\quad n_2 = 40$

$\bar{x}_1 = 647$ $\quad \bar{x}_2 = 638$

$S_1 = 27$ $\quad S_2 = 31$ $\quad \alpha = 0.05$

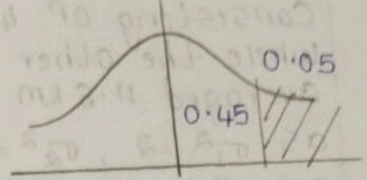
Null hypothesis $H_0 : \mu_1 = \mu_2$

Alternate hypothesis $H_1 : \mu_1 > \mu_2$ (Right tail)

Test Statistics $$Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$Z_{cal} = \frac{647-638}{\sqrt{\frac{(27)^2}{40}+\frac{(31)^2}{40}}}$$

$$= 1.38$$



$z_{tab} = z_\alpha = 0.05 = 1.65$

$z_{tab} > z_{cal}$

accept null hypothesis

$\mu_1 = \mu_2$

Samples of students were drawn from 2 universities and from their weights in kg. mean & S.D are calculated as shown below. make a large sample test to test the significance of difference b/w the means

	mean	SD	size of sample
University A	55	10	400
University B	57	15	100

$n_1 = 400$, $n_2 = 100$, $\bar{x}_1 = 55$, $\bar{x}_2 = 57$, $\sigma_1 = 10$, $\sigma_2 = 15$

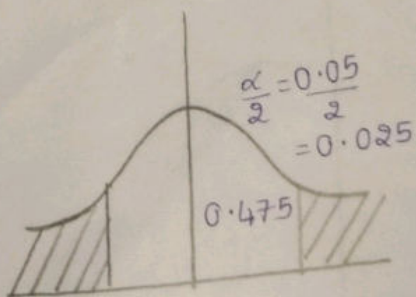
Null hypothesis $H_0 : \mu_1 = \mu_2$

Alternate hypothesis $H_1 : \mu_1 \neq \mu_2$ (Two tail)

Test Statistics $Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} = \frac{55-57}{\sqrt{\frac{(10)^2}{400}+\frac{(15)^2}{100}}} = -1.26$

$Z_{cal} = |-1.26| = 1.26$

$Z_{tab} = Z_{\frac{\alpha}{2}} = 1.65$



$Z_{tab} > Z_{cal}$

accept null hypothesis

$\mu_1 = \mu_2$

Two types of new cars produced in USA are tested for petrol milage. One sample is consisting of 42 cars avg 15 kms per litre while the other sample consisting of 80 cars averaged 11.8 km/litre with population variances as $\sigma_1^2 = 2$, $\sigma_2^2 = 1.5$. Test whether there is any significance difference in the petrol consumption of this 2 types of cars - Use $\alpha = 0.01$

Given $n_1 = 42 \quad \bar{x}_1 = 15 \quad \sigma_1^2 = 2$

$n_2 = 80 \quad \bar{x}_2 = 11.8 \quad \sigma_2^2 = 1.5$

Null hypothesis $H_0 : \mu_1 = \mu_2$

Alternate hypothesis $H_1 : \mu_1 \neq \mu_2$

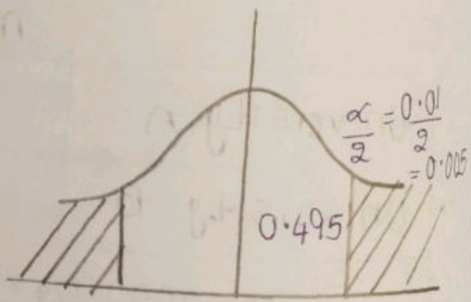
level of significance $\alpha = 0.01$

Test Statistic $$Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{15 - 11.8}{\sqrt{\frac{2}{42} + \frac{1.5}{80}}} = 12.42$$

$Z_{cal} = |12.42| = 12.42$

$Z_{tab} = Z_{\frac{\alpha}{2}} = 2.58$

$Z_{cal} > Z_{tab}$



reject null hypothesis

accept alternate hypothesis

Test of Significance for large Sample

Single proportion.

Test Statistic $= \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}}$

Where P = population proportion

p = Sample proportion

$Q = 1 - P$

Standard Error $= \sqrt{\dfrac{PQ}{n}}$

maximum Error $= Z_{\frac{\alpha}{2}} \times \sqrt{\dfrac{PQ}{n}}$

Confidence limits = $P - M.E \quad P + M.E$

Confidence intervals = $(P - M.E,\ P + M.E)$

1. A manufacturer Claimed that atleast 95% of the equipment which he Supplied to a Factory confirmed to Specification - An examination of Sample of 200 pieces revealed that 18 were faulty. Test his claim at 5% level of significance.

$P = 95\% = \dfrac{95}{100} = 0.95$

$n = 200$

$p = \dfrac{182}{200}$ (if 18 were faulty out of 200 ∴ 182 were good items)

$p = 0.91$

$Q = 1 - 0.95 = 0.05$

large Sample Single proportion

Null hypothesis $H_0 : 0.95 = P$

Alternate hypothesis $H_1 : P < 0.95$ (atleast)

$\alpha = 5\% = 0.05$

Test Statistic $= \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}} = \dfrac{0.91 - 0.95}{\sqrt{\dfrac{0.95 \times 0.05}{200}}} = 2.59$

$Z_{tab} = Z_\alpha$

$Z_{cal} > Z_{ta}$

reject null

accept alt

The manu

2. In a sar

540 are r

Can we as

equally p

significar

$n = 1$

$p = \dfrac{540}{100}$

$P = 0.5$

large Sa

Null hy

Alternat

$\alpha =$

Test St

Z_{ca}

$Z_{tab} =$

Z_{tab}

accep

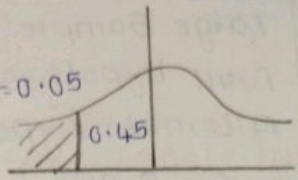
3. In a

found

Suppor

men i

$Z_{tab} = Z_{\alpha} = 1.64$



$Z_{cal} > Z_{tab}$

reject null hypothesis

accept alternate hypothesis

The manufacturers Claim is false

2. In a Sample of 1000 people in Karnataka 540 are rice eaters and rest are wheat eaters. Can we assume that both rice and wheat are equally popular in this state at 1% level of significance

$n = 1000$

$p = \frac{540}{1000} = 0.54$

$P = 0.5 \quad Q = 0.5$

large Sample Single proportion

Null hypothesis $H_0 : \underline{P} = 0.5$

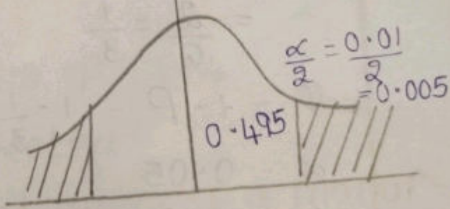
Alternate hypothesis $H_1 : \underline{P} \neq 0.5$

$\alpha = 0.01$

Test Statistic $= \frac{p - \underline{P}}{\sqrt{\frac{PQ}{n}}} = \frac{0.54 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{1000}}} = 2.52$

$Z_{cal} = 2.52$

$Z_{tab} = Z_{\frac{\alpha}{2}} = 2.58$



$Z_{tab} > Z_{cal}$

accept null hypothesis

3. In a big City 325 men out of 600 men were found to be Smokers. Does this information Support to Conclusion that the majority of men in City are Smokers?

$n = 600$

$p = \frac{325}{600} = 0.54 \qquad \underline{P} = 0.5 \quad Q = 0.5$

large Sample Single proportion

Null hypothesis $H_0 : P = 0.5$

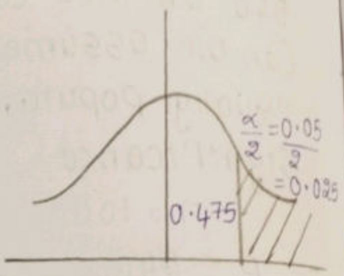
Alternate hypothesis $H_1 : P > 0.5$

$\alpha = 0.05$

Test Statistic $Z_{cal} = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}} = \dfrac{0.54 - 0.5}{\sqrt{\dfrac{0.5 \times 0.5}{600}}}$

$= 1.95$

$z_{tab} = z_{\frac{\alpha}{2}} = 1.65$



$z_{cal} > z_{tab}$

reject null hypothesis

accept alternate hypothesis

15/11

4. A die was thrown 9000 times and of these 3220 yielded a 3 or 4. Is this consistent with the hypothesis that the die was unbiased.

$n = 9000 > 30$ (large Sample)

$p = \dfrac{\text{favourable}}{\text{total}} = \dfrac{3220}{9000} = 0.35$

P = population proportion / probability of success
probability of getting 3 or 4

$= \dfrac{2}{6} = \dfrac{1}{3}$

$Q = 1 - P = 1 - \dfrac{1}{3} = \dfrac{2}{3}$

$\alpha = 0.05$

large Sample Single proportion

Null hypothesis $H_0 : P = \frac{1}{3}$

Alternat hypothesis $H_1 : P \neq \frac{1}{3}$ (Two tail)

Test Statistic $Z_{cal} = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}} = \dfrac{0.35 - 0.11}{\sqrt{\dfrac{0.11 \times 0.66}{9000}}}$

$= 3.35$

$Z_{tab} = z_{\frac{\alpha}{2}} = 1.$

$z_{cal} > z_{tab}$

reject null hy

∴ The die is

5. (A Social work
that only 4%.
A random sa
100 are defec

$P = 4\% =$

$Q = 1 - P =$

$p = \dfrac{100}{500} = 0$

large Sampl
Null hypothe
Alternate hy
Test Statis

$Z_{tab} = z_{\alpha}$

$z_{cal} > z_{ta}$

accept alt

∴ More th

6. Experience
actured pro
production
quality. Te

$P = 20$

$Q = 1$

$n = 400$

$Z_{tab} = Z_{\frac{\alpha}{2}} = 1.96$

$Z_{cal} > Z_{tab}$

reject null hypothesis, accept alternate hypothesis

$\therefore$ The die is biased $\quad P \neq \frac{1}{3}$

5. ~~(A Social worker beli)~~ A manufacturer Claims that only 4% of his products are defective. A random sample of 500 were taken in which 100 are defective. Test hypothesis at 0.05 level

$P = 4\% = \frac{4}{100} = 0.04$

$Q = 1 - P = 1 - 0.04 = 0.96 \quad , n = 500$

$p = \frac{100}{500} = 0.2 \quad , \alpha = 0.05$

large sample single proportion

Null hypothesis $H_0 : P = 0.04$

Alternate hypothesis $H_1 : P > 0.04$ (Right tail)

Test Statistic $Z_{cal} = \dfrac{p - P}{\sqrt{\frac{PQ}{n}}} = \dfrac{0.2 - 0.04}{\sqrt{\frac{0.04 \times 0.96}{500}}}$

$= 18.25$

$Z_{tab} = Z_{\alpha} = 1.65$

$Z_{cal} > Z_{tab}$

accept alternate hypothesis $P > 0.04$

$\therefore$ More than 4% are defective

6. Experience had shown that 20% of a manufactured product is of top quality. In 1 days production 400 articles only 50 are of top quality. Test the hypothesis at 0.05 level

$P = 20\% = \frac{20}{100} = 0.2$

$Q = 1 - P = 1 - 0.2 = 0.8$

$n = 400 \quad , p = \frac{50}{400} = 0.125 \quad \alpha = 0.05$

Null hypothesis $H_0 : P = 0.2$

Alternate hypothesis $H_1 : P < 0.2$

Test statistic $z_{cal} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.125 - 0.2}{\sqrt{\frac{0.2 \times 0.8}{400}}}$

$= -3.75$

$z_{cal} = |-3.75| = 3.75$

$z_{tab} = 1.965$

$z_{tab} < z_{cal}$

reject null hypothesis

accept alternate hypothesis

i.e $P < 0.2$

A manufacture Claimed atleast 95% of the Equipment which he supply

7. Among 900 people in a State 90 are found to be Chapati eaters. Construct 99% Confidence interval for the true proportion.

$P = \cancel{\{95\% = \frac{95}{100} = 0.95\}}\ \frac{90}{900} = 0.1 \quad Q = 1 - 0.1 = 0.9$

Standard error $= \sqrt{\frac{PQ}{n}} = \sqrt{\frac{0.9 \times 0.1}{900}}$

$= 0.01$

Maximum error $= (2.57)(0.01) \quad 1 - \alpha = 99\% = \frac{99}{100} = 0.99$

$= 0.0257 \quad\quad 1 - \alpha = 0.99$

Confidence limits $\quad 1 - 0.99 = \alpha$

$= P - ME, P + ME \quad\quad \alpha = 0.01$

$= 0.1 - 0.0257, 0.1 + 0.0257 \quad z_{\frac{\alpha}{2}} = z_{\frac{0.01}{2}} = z_{0.005}$

$= 0.0743, 0.1257 \quad\quad = 2.57$

Confidence interval $= (0.0743, 0.1257)$

8. In a random Sample of 160 workers exposed to a Certain amount of radiation, 24 experienced Some ill effects. Construct a 99% Confidence interval for the true proportion.

$$p = \frac{24}{160} = 0.15$$

$$q = 1 - p = 1 - 0.15 = 0.85$$

$$S.E = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.15 \times 0.85}{160}}$$

$$= 0.028$$

$$M.E = (2.57)(0.028)$$

$$= 0.0725$$

Confidence limits $= P - ME,\ P + ME$

$$= 0.15 - 0.0725,\ 0.15 + 0.0725$$

$$= 0.0775,\ 0.2225$$

Confidence interval $(0.0775, 0.2225)$

Test of Significance for large Sample difference of proportions

Let p_1 & p_2 be the Sample proportions in two large random samples of sizes n_1 & n_2 drawn from 2 populations having proportions P_1, P_2

$$\text{Test Statistic} = z_{cal} = \frac{p_1 - p_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

Note:

1) When population proportions P_1, P_2 are not known but Sample proportions p_1 & p_2 are know then test Statistic =

$$\frac{p_1 - p_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

2) Method of pooling:

In this method, the estimated value for 2 population proportions is obtained by Cooling the two sample proportions p_1, p_2 into a single proportion p given by

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$\therefore \text{Test Statistic} = z_{cal} = \frac{p_1 - p_2}{\sqrt{\frac{pq}{n_1} + \frac{pq}{n_2}}}$$

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

1) Random Sample of 400 men & 600 women were asked whether they would like to have flyover near their residence. 200 men, 325 women were in favour of the proposal. Test the hypothesis the proportions of men & women in favour of the proposal are same at 5% level.

$n_1 = 400 \quad n_2 = 600$

$P_1 = \frac{200}{400} = \frac{1}{2} \quad P_2 = \frac{325}{600} = 0.541$

$= 0.5$

$\alpha = 0.05$

largle sample difference of proportions

my method of pooling

$$p = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

$$= \frac{400(0.5) + 600(0.541)}{400 + 600}$$

$$= 0.525$$

$$q = 1 - p = 1 - 0.525 = 0.475$$

Null hypothesis $H_0 : \underline{P_1} = \underline{P_2}$

Alternate hypothesis $H_1 : P_1 \neq P_2$ (Two tail)

Test Statistics $= \dfrac{P_1 - P_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$$= \frac{0.5 - 0.541}{\sqrt{(0.525)(0.475)\left[\frac{1}{400} + \frac{1}{600}\right]}}$$

$$= 1.274$$

$z_{tab} = z_{\frac{\alpha}{2}} = 1.96$

$z_{tab} > z_{cal}$

accept null hypothesis

$\therefore$ men & women are in favour of proposal

2) Random sample of 400 men & 200

A

17/11

3) A manufacturer of electronic equipment subjects samples of 2 competing brands of transistors to an accelerated performance test. If 45 of 180 transistors of 1^{st} kind & 34 of 120 of 2^{nd} kind fail the test - what can we conclude at $\alpha = 0.05$ about the difference b/w corresponding sample proportions.

$n_1 = 180 \quad n_2 = 120$

$P_1 = \frac{45}{180} = 0.25 \quad P_2 = \frac{34}{120} = 0.28$

$\alpha = 0.05$

large sample difference of proportions

Null hypothesis $H_0 : P_1 = P_2$

Alternat hypothesis $H_1 : P_1 \neq P_2$

Test Statistic $= P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$

$$= \frac{(180)(0.25) + 120(0.28)}{180+120}$$

$$= 0.262$$

$$Q = 1 - 0.262 = 0.738$$

$$Z_{cal} = \frac{0.25 - 0.28}{\sqrt{(0.262)(0.738)\left[\frac{1}{180} + \frac{1}{120}\right]}}$$

$$= -0.578$$

$|Z_{cal}| = |-0.578| = 0.578$

$Z_{\frac{\alpha}{2}} = Z_{tab} = 1.96$

$Z_{tab} > Z_{cal}$

accept null hypothesis

17/11