ESTIMATION AND TESTING OF HYPOTHESIS, LARGE SAMPLE TEST

PART-A : ESTIMATION

Population:

Population are (or) universe is the aggregate or totality of statistical data forming a subject of investigation.

For example,

1. The population of heights of Indians.
2. The population of rationlized banks in India.

Sample:

Most of the times study of the entire population is not possible to carry out and hence and a part alone is selected from the given population.

A portion of the population which is examined with a view of determining the population characteristics is called a sample. size of the sample is denoted by 'n'.

Classification of samples:

1. Large sample:

If the size of the sample n greater than or equal to 30 then the sample is said to be large sample.

2. Small sample:

If the size of the sample n less than 30 then the sample is said to be small sample or exact sample.

Parameters and statistics:

Parameter is a statistical measure based on observations of a population.

Statistic is a statistical measure based only on all the units selected in a sample.

For example, the constants of the population namely mean(μ), variance (σ^2), standard deviation (σ) are called parameters.

The constants of the sample namely mean($\bar{x}$), variance (s^2), standard deviation(s) are called statistics.

Imp Estimate:

An estimate is a statement made to find an unknown population parameter.

Imp Estimator:

The procedure (or) rule to determine an unknown population parameter is called an estimator.

Imp Estimation:

The process of determining estimator in such a way that they are due to parameter value, this is known as estimation.

Types of estimations:

Basically there are two kinds of estimates to determine the statistic of population parameter namely

1. Point estimation
2. Interval estimation.

1. Point estimation:

If an estimate of the population parameter is given by a single value, then the estimate is called a point estimation of parameter.

EX: If the height of the student is measured as 162cm then the measurement gives point estimation.

2. Interval estimation:

If an estimate of a population parameter is given by two different values between which the parameter may be consider to lie, then the estimate is called an interval estimation of the parameter.

EX: If the height is given as 160 ($\pm$3.5) cm then the height lies between 159.5cm and 166.5cm is an interval estimation.

properties of Estimation:

An estimator is not expected to estimate the population parameter without error. An estimator should be close to the true value of the unknown parameter.

unbiased estimation:

In an interval estimation of the Population parameter θ, if we can find two two quantities t_1 and t_2 based on sample observations drawn from the population such that the unknow parameter θ is included in the interval $[t_1, t_2]$ in the specified percentage of cases then this interval is called Confidence interval for the parameter θ.

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$ has a standard normal distribution with mean $=0$ and $\sigma = 1$.

I. Confidence limits for population mean(μ):

1. 95% Confidence limits are $\bar{x} \pm 1.96$ (standard ~~erro~~ errors of $\bar{x}$).
2. 99% Confidence limits are $\bar{x} \pm 2.58$
3. 99.73% Confidence limits are $\bar{x} \pm 3$
4. 90% Confidence limits are $\bar{x} \pm 1.64$

II. Confidence limits for population proportion(P):

1. 95% Confidence limits are $P \pm 1.96$.
2. 99% Confidence limits are $P \pm 2.58$
3. 99.73% Confidence limits are $P \pm 3$
4. 90% Confidence limits are $P \pm 1.64$

Determination of proper sample size (n):

Determination of proper sample size is important for testing of Hypothesis in business problems. The size of sample should neither be too small nor too large. If the size of the sample is too small then it may not give a wild conclusion. On the other hand if the size of the sample is too large then there may be loss of time and money without getting the required result.

Sample size for estimating population mean:

Let $\bar{x}$ be the mean of a random sample drawn from a population having mean μ and standard deviation σ. Let the sample distributio of the sample mean $\bar{x}$ be approximately a normal distribution with mean μ and standard deviation σ.

Let E be the sampling error then $\boxed{E = \bar{x} - \mu}$

The Confidence interval for the population mean(μ) is equal to $\bar{x} \pm E$.

where $E = z \times \frac{\sigma}{\sqrt{n}}$, n being the sample size and σ be the standard deviation.

$$\boxed{E = \frac{z\sigma}{\sqrt{n}} \text{ (or) } \sqrt{n} = \frac{z\sigma}{E}}$$

$$\Rightarrow \boxed{n = \left(\frac{z\sigma}{E}\right)^2}$$

where z is standard error of $\bar{x}$ (or) Confidence coefficient.

Sample size for estimating population proportion:

Let p be the sample proportion and P be the population proportion.

$$\boxed{E = p - P}$$

The confidence interval for the population proportion (P) is $p \pm E = p \pm z$

where z = standard error of $\bar{x}$.

$$\boxed{E^2 = \frac{z^2PQ}{n}} \quad \text{where } Q = 1 - P$$

$$\boxed{n = \frac{z^2PQ}{E^2}}$$

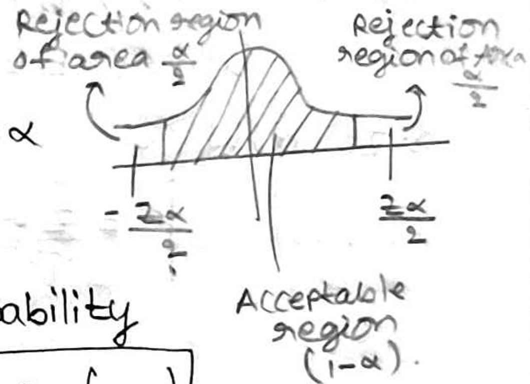
Maximum error for large sample:

Since, the sample mean estimate very rarely equal to the mean of the population μ, A point estimate is generally accompained with a statement of error which gives the difference between the estimate and the quantity to be estimated, thus error is

$|\overline{x} - \mu|$.

For large n, the random variable $\frac{\overline{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ is normally variate approximately.

Here, probability of $P\left(-\frac{Z\alpha}{2} < Z < \frac{Z\alpha}{2}\right) = 1-\alpha$



Hence,

$$P\left(-\frac{Z\alpha}{2} < \frac{\overline{x}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{Z\alpha}{2}\right) = 1-\alpha$$

The maximum error of estimate E with $1-\alpha$ probability is given by $\cancel{E = \frac{\frac{Z\alpha}{2}}{\frac{\sigma}{\sqrt{n}}}}$ $\boxed{E = \frac{Z\alpha}{2}\left(\frac{\sigma}{\sqrt{n}}\right)}$

Confidence interval for μ, σ known:

If $\overline{x}$ is the mean of sample of size n from the population with variance σ^2, $(1-\alpha)100\%$ confidence interval for μ is given by

$$\boxed{\overline{x} - \frac{Z\alpha}{2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \overline{x} + \frac{Z\alpha}{2}\left(\frac{\sigma}{\sqrt{n}}\right)}$$

* Maximum error of estimate E for small sample:

when $n < 30$, small sample we use s for the standard deviation of sample of determine E.

$$E = \frac{t\alpha}{2}\frac{S}{\sqrt{n}}.$$

Confidence interval for sample space is

$$\boxed{\bar{x} - t_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{x} - t_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right).}$$

Note:
At confidence limit 95%. the error is

$$1-\alpha = 95\% = 0.95$$

$$\alpha = 0.05$$

$$\frac{\alpha}{2} = 0.025$$

Now, $0.5 - 0.025 = 0.475$

At area 0.475, $z = 1.96$ (From table).

What is the maximum error one can expect to make with probability 0.90 when using the mean of a random sample of size $n=64$ to estimate the mean of population with $\sigma^2 = 2.56$.

Given that,

the probability $= 0.90$

Confidence limit $= 90\%$.

$$\frac{z_\alpha}{2} = 1.64$$

$$\sigma^2 = 2.56 \Rightarrow \boxed{\sigma = 1.6}$$

$$n = 64.$$

Maximum error $E = \frac{z_\alpha}{2}\left(\frac{\sigma}{\sqrt{n}}\right)$

$$= 1.64\left(\frac{1.6}{\sqrt{64}}\right)$$

$$= 1.64\frac{(1.6)}{8}$$

$$\cancel{= 1.312} = 0.328.$$

$$\therefore \boxed{\cancel{E = 1.312}} \quad \boxed{E = 0.328}$$

A random variable sample size 100 has a SD of 5 when can you say that maximum error with 95% confidence.

Given that,

$$\sigma = 5$$

$$n = 100$$

$Z_{\frac{\alpha}{2}}$ value at 95% confidence $= 1.96$.

$$\text{Maximum error} = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$= 1.96 \times \frac{5}{\sqrt{100}}$$

$$= 1.96 \times \frac{5}{10}$$

$$= 0.98$$

$$\therefore \boxed{E = 0.98}$$

If we assert with 95% that the maximum error is 0.05 and $P = 0.2$ find the size of the sample.

Given that,

$$E = 0.05$$

$Z_{\frac{\alpha}{2}}$ value at 95% confidence $= 1.96$

$$P = 0.2$$

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

we know that,

$$n = \frac{Z^2 PQ}{E^2}$$

$$n = \frac{(1.96)^2 (0.2)(0.8)}{(0.05)^2}$$

$$n = 245.86$$

$$\boxed{n = 246}$$

What is the size of the smallest sample required to estimate an unknown proportion to within a maximum error of 0.06 with atleast 95% confidence.

Given that,

$E = 0.06$

$Z_{\frac{\alpha}{2}}$ value at 95% confidence $= Z = 1.96$.

consider, $P = \frac{1}{2}$ $\quad$ ($\because$ P is not given).

$Q = 1 - P = 1 - \frac{1}{2} = \frac{1}{2}$.

we know that,

$$n = \frac{Z^2 PQ}{E^2}$$

$$= \frac{(1.96)^2 (\frac{1}{2})(\frac{1}{2})}{(0.06)^2}$$

$$= 266.78$$

$$\boxed{n = 267}$$

In study of an automobile insurance of a random sample of 80 body repair costs had a mean of 472.36/- and S.D of ₹62.35. If $\bar{x}$ is used as a point estimate to the true average repair costs with what confidence we can assert that the maximum error does not exceed ₹10.

Given that,

size of random sample $(n) = 80$.

Mean $(\bar{x}) = ₹472.36$

S.D $(\sigma) = ₹62.35$

Maximum error $(E) = 10$.

we know that,

$$E = Z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$

$$Z_{\frac{\alpha}{2}} = \frac{E \times \sqrt{n}}{\sigma}$$

$$= \frac{10 \times \sqrt{80}}{62.35}$$

$$= 1.4345$$

$$\boxed{Z_{\frac{\alpha}{2}} = 1.4345}$$

The area when $z = 1.43$ is "$\underline{0.4236}$." (From table)

Area between $-z_{\frac{\alpha}{2}}$ and $z_{\frac{\alpha}{2}}$ $= 2(0.4236)$

$1-\alpha = 0.8472$

The confidence limit is $(1-\alpha) \times 100$ %.

$= (0.8472) \times 100$

$= 84.72\%$.

A random sample of size 50, Mean and S.D of population are 11,795 and 14054 respectively. Find 95% confidence interval for the mean.

Given that,

size of sample $= 50 = n$

$\bar{x} = 11795$

$\sigma = 14054$.

$z_{\frac{\alpha}{2}}$ value at 95% confidence $= 1.96$.

confidence interval is $\bar{x} \pm E$

where $E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

$= (1.96) \frac{(14054)}{\sqrt{50}}$

$E = 3895.57$

~~Confider~~ Now, $\bar{x} + E = 11795 + 3895.57$

$\bar{x} + E = 15690.57$

$\bar{x} - E = 11795 - 3895.57$

$\bar{x} - E = 7899.43$.

Confidence interval is

$$\left[\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

$= [\bar{x} - E < \mu < \bar{x} + E]$

$= 7899.43 < \mu < 15690.57$ $\quad \therefore [7899.43, 15690.57]$

A sample is collected from the items produced by a factory. The sample size is 81. The SD of the population is 0.3. Find the standard error of the mean of sampling distribution.

Given that,

$\sigma = 0.31$

$n = 81$

$$\text{standard error of mean} = \frac{\sigma}{\sqrt{n}}$$

$$= \frac{0.31}{\sqrt{81}}$$

$$= 0.034$$

$\therefore \boxed{S.E \text{ of } \bar{x} \ (z) = 0.033}$

Among 900 people in a state 90 are found to be chapati enters. Construct 99% confidence interval for the true population.

Given that,

$n = 900$

The proportion is $P = \frac{90}{900} = \frac{1}{10} = 0.1$

$\boxed{P = 0.1}$

$Q = 1 - P = 1 - 0.1 = 0.9$

$\boxed{Q = 0.9}$

$Z_{\alpha/2}$ value at 99% confidence $= 2.58$.

Confidence interval is $\boxed{P \pm E}$

$$E = Z_{\alpha/2} \frac{\sqrt{PQ}}{\sqrt{n}}$$

$$E = 2.58 \frac{\sqrt{(0.1)(0.9)}}{\sqrt{900}}$$

$$E = 2.58 \frac{\sqrt{(0.1)(0.9)}}{30}$$

$\boxed{E = 0.0258}$

Confidence interval is $P \pm E$

i.e, $[P-E, P+E]$

$= [0.1 - 0.0258, 0.1 + 0.0258]$

$= [0.0742, 0.1258]$

$\therefore$ Confidence interval is $[0.0742, 0.1258]$.

If 80 patients are treated with an antibiotic, 59 got cured. Find a 99% confidence limits to the true proportion of cure.

Given that,

$n = 80$

Proportion is $P = \frac{59}{80} = 0.7375$

$Q = 1 - P = 1 - 0.7375 = 0.2625$

$Z_{\alpha/2}$ value at 99% confidence $= 2.58$

$$E = Z_{\alpha/2}\sqrt{\frac{PQ}{n}}$$

$$= 2.58\sqrt{\frac{0.7375(0.2625)}{80}}$$

$$\boxed{E = 0.1269}$$

Confidence ~~limits~~ interval is $P \pm E$.

i.e, $[P - E, P + E]$

$= [0.7375 - 0.1269, 0.7375 + 0.1269]$

$= [0.6106, 0.8644]$

$\therefore$ Confidence interval is $[0.6106, 0.8644]$.

A sample size 10 and standard deviation 0.03 is taken from a population. Find the maximum error with 99% Confidence.

Given that,

$n = 10$

$\sigma = 0.03$

$Z_{\alpha/2}$ value at 99% confidence $= 2.58$

Maximum error $E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

$E = 2.58 \times \frac{0.03}{\sqrt{10}}$

$E = 0.02447$

$\therefore$ $\boxed{\text{Maximum error} = 0.02447}$

Assuming that $\sigma = 20$, how large a random sample be taken to assert with probability 0.95 that the sample mean will not differ from true mean by more than 3 points.

Given that,

$\sigma = 20$

Maximum error$(E) = 3$.

$Z_{\alpha/2}$ value at 0.95 (or) 95% confidence $= 1.96$.

we know that,

$E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$3 = 1.96 \times \frac{20}{\sqrt{n}}$

$\sqrt{n} = 1.96 \times \frac{20}{3}$

$\sqrt{n} = 13.0667$

$n = 170.7386$

$\boxed{n = 171}$ (approximately)

The guaranteed average life of a certain type of electric bulbs is 1500hrs with a SD of 120hrs. It is decided to sample the output so as to ensure that 95% of bulbs do not fall shrt of the generated average by more than 2%. what is will be the maximum sample size.

Given that,

$$SD(\sigma) = 120$$

$$\text{Max. sample size } (n) = \left(\frac{Z\sigma}{E}\right)^2$$

$Z_{\alpha/2}$ value at 95% confidence $= 1.96 = Z$

Maximum error $(E) = \underset{\text{of } 1500}{2\%} = 0.02 \times 1500 = 30$

$$n = \left(\frac{Z\sigma}{E}\right)^2$$

$$= \left(\frac{1.96 \times 120}{\cancel{0.02}\ 30}\right)^2$$

$$= 61.4656$$

$$= 62 \text{ (approximately)}.$$

The mean of random sample is an unbiased. estimate of the mean of the population

3, 6, 9, 15, 27.

a) List all possible samples of size 3 can be taken without replacement from the finite population.

b) calculate the mean of each of the sample listed in (a) and assigning each sample a Probability of 1/10. verify that the mean of this $\overline{x} = 12$. which equal to the population mean of the θ i.e, $E(\overline{x}) = \theta$.

Given that,

3, 6, 9, 15, 27.

The possible samples of size 3 from 3, 6, 9, 15, 27 without replacement are $^5C_3 = {^5C_2} = 10$.

(3,6,9), (3,6,15), (3,6,27), (3,9,15), (3,9,27), (3,15,27)
(6,9,15), (6,9,27), (9,15,27), (6,15,27).

b) Mean of each of the samples are

Mean of (3,6,9) = 6
Mean of (3,6,15) = 8
Mean of (3,6,27) = 12
Mean of (3,9,15) = 9
Mean of (3,9,27) = 13
Mean of (3,15,27) = 15
Mean of (6,9,15) = 10
Mean of (6,9,27) = 14
Mean of (9,15,27) = 17
Mean of (6,15,27) = 16

probability assigned to each one is 1/10

x	6	8	9	12	13	14	15	16	17	10
$P(x)$	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10

Mean of the population $= \frac{3+6+9+15+27}{5} = \frac{60}{5} = 12$
$= \theta$

$$E(x) = \Sigma x\,P(x) = \frac{6}{10} + \frac{8}{10} + \frac{9}{10} + \frac{12}{10} + \frac{13}{10} + \frac{14}{10} + \frac{15}{10} + \frac{16}{10} + \frac{17}{10} + \frac{10}{10}$$

$$= \frac{120}{10}$$

$$= 12$$

$\therefore E(\overline{x}) = 12 = \theta$

$\therefore \boxed{E(\overline{x}) = \theta}$

Hence, it is verified.

A random sample of size 100 is taken from an infinite population having the mean $\mu = 76$, $\sigma^2 = 256$. What is the probability that x will be between 75 and 78.

Given that,

$\sigma^2 = 256$, $n = 100$

$\sigma = 16$

$\mu = 76$.

we know that $\boxed{z = \frac{x - \mu}{\sigma}}$

when $x = 75$ then $z_1 = \frac{75 - 76}{16} = -0.0625$

when $x = 78$ then $z_2 = \frac{78 - 76}{16} = 0.125$.

$P(75 < x < 78) = P(-0.0625 < z < 0.125)$

$z_1 = -0.0625 < 0$

$z_2 = 0.125 > 0$

$= A(-0.0625) + A(0.125)$

$= A(0.06) + A(0.13)$

$= 0.0239 + 0.0517$

$= 0.0756$.

$\therefore \boxed{P(75 < x < 78) = 0.0756}$

(ii)

We know that, $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

$z_1 = \frac{75 - 76}{\frac{16}{\sqrt{100}}} = \frac{-1}{16/10} = -1/1.6 = 0.625$

$z_2 = \frac{78 - 76}{\frac{16}{\sqrt{100}}} = \frac{2}{16/10} = 1.25$.

$P(75 < x < 78) = P(-0.625 < z < 1.25)$.

$= A(-0.625) + A(1.25)$

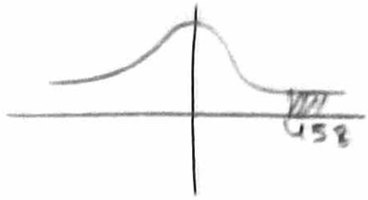
$= A(0.625) + A(1.25)$

$= 0.2324 + 0.3944$.

$= 0.6268$

The mean voltage of a battery is 150 and SD is 0.2. Find the probability that 4 such batteries connected in a series will have a combined voltage of 60.8 or more volts.

Given that,



$\mu = 150$

$\sigma = 0.2$

$n = 4$.

we know that, $\boxed{z = \dfrac{x - \mu}{\frac{\sigma}{\sqrt{n}}}}$

when $x = 60.8$ then $z_1 = \dfrac{60.8 - 150}{\frac{0.2}{\sqrt{4}}}$

$z_1 = 458$

$P(x \geq 60.8) = P(z \geq 458)$

$= P(-\infty < z < 0) + P(0 < z < 458)$

$= 0.5 + 0.5$

$= 1$.

A firm wishes to estimate with a maximum allowable error of 0.05 and 95% level of confidence, the proportion of consumers who prefer its Product. How large a sample will be required in order to make such an estimate, if the Preliminary sales report indicates that 25% of all consumers perform the firm's product.

Given that, Maximum error (E) $= 0.05$

$P = 25\% = 0.25$

$Q = 1 - P = 1 - 0.25 = 0.75$

$Z_{\alpha/2}$ value with confidence 95% $= 1.96 = Z$

We know that, $n = \frac{Z^2 \cdot PQ}{E^2}$

$$= (1.96)^2 \frac{(0.25)(0.75)}{(0.05)^2}$$

$$= 288.12$$

$\boxed{n = 288}$ (approximately)

Testing of Hypothesis:

Standard error of a statistic: (S.E).

The standard error of a statistic i.e, (S.E of sample mean or sample S.D) is the standard deviation of the sampling distribution of statistic. Thus S.E of sample mean($\bar{x}$) is the S.D is the sampling distribution of the sample mean.

It is used for assert the difference between expected value and observed value.

Formulae for S.E:

1. S.E of sample mean $\bar{x} = \frac{\sigma}{\sqrt{n}}$

2. S.E of sample proportion $P = \sqrt{\frac{PQ}{n}}$.

3. S.E of sample S.D $= \frac{\sigma}{\sqrt{2n}}$.

4. S.E of the difference of 2 sample means $\bar{x}_1$ and $\bar{x}_2$

$$S.E \text{ of } (\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

5. S.E of difference of 2 sample proportions P_1 and P_2.

$$S.E \text{ of } (P_1 - P_2) = \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}.$$

6. S.E of S.D $(S_1 - S_2) = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$

Infinite population:

Suppose the samples are drawn from an Infinite population N i.e, $N \to \infty$ then

(i) mean of the sampling distribution

$$\mu_{\bar{x}} = \mu$$

$$\text{i.e, } E(\bar{x}) = \mu.$$

(ii) Variance: $\boxed{\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}}$

Finite population:

Consider a finite population of size N and with mean μ, standard deviation σ. Draw all possible samples of size n without replacement. From this population

i) Mean $\mu_{\bar{x}} = \mu$

i.e, $\boxed{E(\bar{x}) = \mu}$

ii) variance $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$

The factor $\frac{N-n}{N-1}$ is called the finite population correction factor.

Central limit theorem:

If $\bar{x}$ be the mean of a sample size n drawn from a population with mean μ and standard deviation (s.D) σ then the sample mean

$$\boxed{z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}}$$

If $\bar{x} = 31.45$, $\mu = 32.3$, $\sigma = 1.6$, $n = 40$. Find z-test statistic

Given that, $\bar{x} = 31.45$

$\mu = 32.3$

$\sigma = 1.6$

$n = 40$

we know that, $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

$$z = \frac{31.45 - 32.3}{\frac{1.6}{\sqrt{40}}}$$

$$z = -3.359$$

$\therefore \boxed{z = -3.359}$

What is the value of correction factor, if $n = 5$, $N = 200$

Given that, $n = 5$

$N = 200$

$$\text{correction factor} = \frac{N-n}{N-1}$$

$$= \frac{200-5}{200-1}$$

$$= \frac{195}{199}$$

$$= 0.9798$$

$\therefore$ correction factor $= 0.9798$.

What is the value of a finite population correction factor $n = 10$, $N = 1000$.

$$\text{Correction factor} = \frac{N-n}{N-1}$$

$$= \frac{1000-10}{1000-1}$$

$$= 0.9909$$

$\therefore$ correction factor $= 0.9909$.

The size of the population is 2000 and size of the sample is 200. Find the correction factor of population.

Given that, $N = 2000$

$n = 200$

$$\text{correction factor} = \frac{N-n}{N-1}$$

$$= \frac{2000-200}{2000-1}$$

$$= \frac{1800}{1999}$$

$$= 0.90045$$

$\therefore$ correction factor $= 0.9005$

How many different samples of size 2 can be choosen from a finite population of 25

No. of different samples of size 2 from $25 = {}^{25}C_2 = 300$.

Test of Hypothesis:

We need to decide whether to accept or reject the statement about the parameter. This statement is called a Hypothesis. And the decision making procedure about Hypothesis is called Test of Hypothesis.

Statistical Hypothesis.

In many circumstances to arrive at decisions about the population on the basis of sample information, we make assumptions (or) guesses about the population parameters involved. such an assumption or statement is called a statistical hypothesis which may or maynot be true. The procedure to decide based on sample results whether a hypothesis is true or not is called test of Hypothesis (or) test of significance.

There are two types of hypothesis:

1. NULL hypothesis
2. Alternative hypothesis

Imp

1. Null hypothesis:

A Null hypothesis is the hypothesis which asserts that there is no significant difference between the statistic and the population Parameter and whatever observed difference is there it is due to fluctuations in sampling from the same population. It is denoded by H_0

Imp 2. Alternative hypothesis:

Any hypothesis which contradicts the null hypothesis is called an Alternative hypothesis. It is usually denoted by H_1.

→ The two hypothesis H_0 and H_1 are such that if one is true, the other one is false and vice versa.

For example,

$$H_0 = \mu = \mu_0$$

The alternative hypothesis would be

i. H_1 ~~≠ μ~~ $\mu \neq \mu_0$

ii. $\mu > \mu_0$

iii. $\mu < \mu_0$

The alternative hypothesis (i) is known as a two tailed alternative.

The alternative hypothesis (ii) is known as Right tailed alternative.

The alternative hypothesis (iii) is known as left tailed alternative.

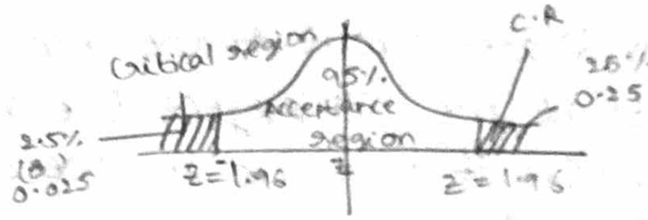
Imp Critical region:

Under a given hypothesis, let the sampling distribution of a statistic is approximately a normal distribution and standard deviation σ_x = standard error (S.E) of t.

then $$z = \frac{t - E(t)}{S.E \text{ of } t}$$

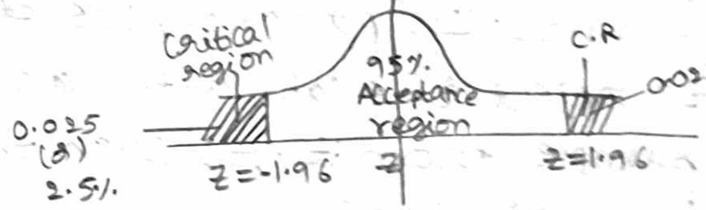
$$t = \frac{\text{observed value} - \text{Expected value}}{S.E}$$ is

called the standarised normal variate or z-test statistic and its distribution is the standard normal distribution with mean 0 and S.D is 1 when sample is large.



The set of Z-scores outside the range -1.96 and 1.96 i.e, $|Z|>1.96$ consists the critical region or the critical region or region of rejection of the hypothesis (or) region of significance.

Acceptable region:



The set of Z-scores inside the range -1.96 and 1.96 is called the region of acceptance of hypothesis.

→ the values -1.96 and 1.96 are called critical values at 5% level of significance.

Note:

A region corresponding to statistic t in the sample s which leads to the rejection of H_0 is called critical region or rejection region. Those region which leads to the acceptance of H_0 give us a region called acceptance region.

Level of significance:

The level of significance is denoted by α is the confidence with with we rejects or accepts the null hypothesis H_0

i.e, it is maximum possibile probability which we are willing to risk an error in rejectiong H_0 when it is true.

In practise, 5% level of significance in a test procedure indicates about there are about ~~to~~ 5 cases in 100 that we will reject the null hypothesis H_0 when it is true i.e, we are about 95% confidence that we have made a right decision.

Errors of sampling:

The main objective in sampling theory is to draw valid inferences about the population parameters on the basis of sample results. In practice, we decide to accept or to reject the lot after the examing the sample from it. As such that we have two types of errors.

Type I error

Reject H_0 when it is true.

If the null hypothesis

Here α, β are called sizes of Type 1 and Type 2 errors respectively.

The sizes of type 1 and type 2 errors are also known as producers risk and ~~customers~~ ~~(or)~~ consumers risk ~~risk~~ respectively.

IMP one tailed and two tailed test:

If we have to test whether the population mean μ has specified value μ_0, then the null hypothesis is $H_0: \mu = \mu_0$ and the alternate hypothesis may be

(i) $H_1: \mu \neq \mu_0$ (i.e, $\mu > \mu_0$ (or) $\mu < \mu_0$)

(ii) $H_1: \mu > \mu_0$ ~~(or)~~

(iii) $H_1: \mu < \mu_0$

→ The alternative hypothesis in (i) is known as a two tailed alternatives ~~and~~ ~~the~~ (i.e both left and right tail)

→ The alternative hypothesis in (ii) & (iii) are known as (one tailed alternative) right tailed and left tailed alternatives respectively.

One tailed test:

→ If the alternative hypothesis (H_1) in a test of stastical hypothesis be one tailed then the test is called a one tailed test.

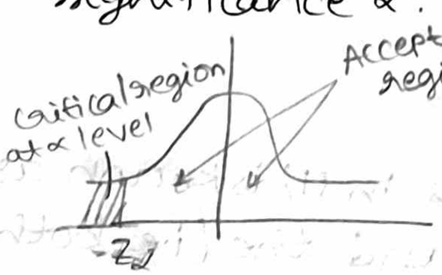
For example, to test whether the population mean $\mu = \mu_0$, we have H_0 is $\mu = \mu_0$ against the alternative hypothesis H_1 given by

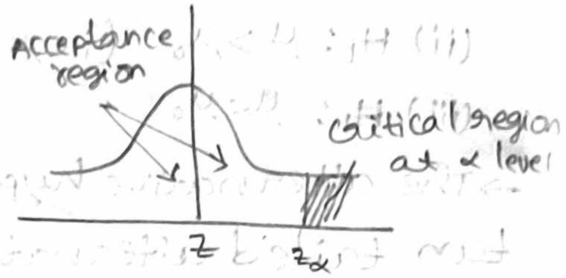
(i) $H_1: \mu > \mu_0$ (Right tailed)

(ii) $H_1: \mu < \mu_0$ (left tailed) and the corresponding test is a single tailed (or) one tailed (or) one sided.

→ In the Right tailed test H_1 is $\mu > \mu_0$, the critical region $z > z_\alpha$ lies entirely in the right tile of the sampling distribution of sample mean $\bar{x}$ with area equal to the level of significance α.

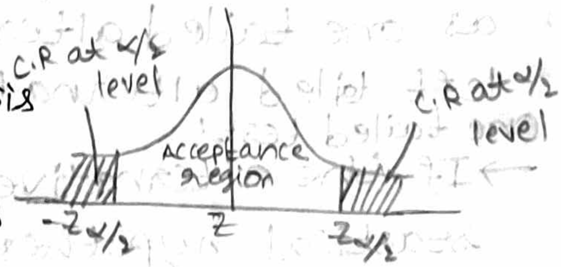
Similarly, In the left tailed test H_1 is $\mu < \mu_0$. the critical region $z < z_\alpha$ lies entirely in the left tail of the distribution of sample mean $\bar{x}$ with area equal to the level of significance α.





Two tailed test:

Suppose, we want to test the null hypothesis $\mu = \mu_0$ against the alternative hypothesis H_1 is $\mu \neq \mu_0$.



Since, H_1 is two tailed alternative hypothesis, the critical region under the curve is equally distributed on both sides of the mean.

Thus the critical area under the right tail is equal to critical area under the left trail = half of the total area

$= \frac{1}{2}$(probability of rejection)

$= \alpha/2$ with critical statistics $z_{\alpha/2}$

where α is the level of significance.

the critical region is $z \le -z_{\alpha/2}$, $z_{\alpha/2} \le z$

Critical values of z:

Level of significance α	1%	5%	10%
Critical values for two tailed test	$\lvert z_{\alpha}\rvert = 2.58$	$\lvert z_{\alpha}\rvert = 1.96$	$\lvert z_{\alpha}\rvert = 1.64$
Critical values for right tailed test	$z_{\alpha} = 2.33$	$z_{\alpha} = 1.645$	$z_{\alpha} = 1.28$
Critical values for the left tailed test	$z_{\alpha} = -2.33$	$z_{\alpha} = -1.645$	$z_{\alpha} = -1.28$

Procedure for testing of hypothesis:

Various steps involved in testing of hypothesis are given below.

Step-1: Null hypothesis

Define or set up a null hypothesis H_0 taking into consideration the nature of problem and data involved.

Step-2: Alternative hypothesis.

set up the alternative hypothesis H_1 so that we could decide whether we should use one tailed (or) two tailed test.

Step-3: Level of significance

Select the appropriate level of significance α depending on the reliability of the estimates i.e. a suitable α is selected in advance. If it is not given in the problem then we take 5% of level of significance.

Step-4: Test statistic

Compute the test statistic

$$z = \frac{t - E(t)}{S.E \text{ of } t}$$ under the null hypothesis.

(or) $$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

Step-5: Conclusion

We compare the computed value of the test statistic z with the critical z_α at given level of significance α.

i. If $|z| < z_\alpha$ then we conclude that it is not significant. We accept the null hypothesis.

ii. If $|z| > z_\alpha$ then the difference is significant. We reject the null hypothesis

Test of significance for large sample:

If the sample size $n \geq 30$ then we consider such samples as large samples.

If n is very large then we take z as the standard normal variate

i.e, $z = \frac{x - \mu}{\sigma}$

Large Sample test:

For large sample we will use z-test.

A coin was tossed 960 times and returned heads 183 times. Test the hypothesis that the coin is unbiased. Use 0.05 level of significance.

Here $n = 960$

p = probability of getting head in one trail $= 1/2$

$q = 1 - p = 1 - 1/2 = 1/2$

We know that,

$\mu = np$,

$= 960(1/2)$

$\boxed{\mu = 480}$

$\sigma^2 = npq = 960\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = 240$

$\therefore \sigma = \sqrt{240}$

$\boxed{\sigma = 15.49}$

x = No. of heads returned $= 183$.

level of significance $(\alpha) = 0.05 = 5\%$

1. Null hypothesis:

H_0: The coin is unbiased.

2. Alternative hypothesis:

H_1: The coin is biased.

3. level of significance

$\alpha = 0.05 = 5\%$

4. Test statistic

$$z = \frac{x - \mu}{\sigma}$$

Now, $z = \dfrac{183 - 480}{15.49}$

$\boxed{z = -19.17}$

$|z| = |-19.17| = 19.17$

5. Conclusion:

$Z_\alpha = 1.96$ $\quad$ ($\because$ two tailed test at $\alpha = 5\%$.)

$|Z| = 19.17$

Here, $|Z| > Z_\alpha$

$\therefore$ The null hypothesis H_0 is rejected.

i.e, The coin is biased.

A coin was tossed 400 times and returned heads 216 times. Test the hypothesis that the coin is unbiased use 0.05 level of significance.

Given that,

$$n = 400$$

P = Probability of getting a head in single trail $= 1/2$.

$q = 1 - P = 1 - 1/2 = 1/2$.

We know that,

$$\mu = nP = 400\left(\frac{1}{2}\right) = 200$$

$$\sigma^2 = npq = 400\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = 100$$

$$\boxed{\sigma = 10}$$

X = No. of heads returned $= 216$.

1. Null hypothesis

H_0: The coin is unbiased.

2. Alternative hypothesis

H_1: The coin is biased.

3. level of significance

$\alpha = 0.05 = 5\%$.

4. Test statistics

$$z = \frac{X - \mu}{\sigma}$$

$$= \frac{216 - 200}{10}$$

$\boxed{Z = 1.6}$

$|Z| = 1.6.$

5. Conclusion:

$Z_\alpha = 1.96$ $\quad$ ($\because$ two tailed test at $\alpha = 5\%$)

$|Z| = 1.6$

Here $|Z| < Z_\alpha$.

$\therefore$ The null hypothesis H_0 is accepted.

$\therefore$ The coin is unbiased.

A dice is tossed 960 times and falls with 5 upwards 184 times. Is the dice is unbiased at a level of significance 0.05.

Given that,

$n = 960$

P = Probability of getting 5 in single trail = $1/6$

$q = 1 - P = 1 - 1/6$

$\boxed{q = 5/6}$

We know that,

$\mu = nP = 960(1/6) = 160$

$\sigma^2 = npq = 960\left(\frac{1}{6}\right)\left(\frac{5}{6}\right) = \frac{400}{3} = 133.34$

$\sigma = \sqrt{npq} = 11.55$

x = No. of times 5 returned = 184.

1. Null hypothesis:

The dice is unbiased.

2. Alternative hypothesis

The dice is biased

3. level of significance

$\alpha = 0.05 = 5\%.$

4. Test statistics

$$Z = \frac{x - \mu}{\sigma} = \frac{184 - 160}{11.55} = 2.078$$

5. Conclusion

$Z_\alpha = 1.96$

$|Z| = 2.078$

Here $|Z| \not< Z_\alpha$

$\therefore$ The null hypothesis is ~~accepted~~ rejected

i.e, The ~~coin is~~ dice is ~~un~~biased.

A dice is tossed 256 times and it turns up with an even digit 150 times. Is the die biased?

Given that,

$n = 256$

P = probability of getting an even number $= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$

$q = 1 - P = 1 - \frac{1}{2} = \frac{1}{2}$.

Here, $\mu = nP = 256\left(\frac{1}{2}\right) = 128$

$\sigma^2 = npq = 256\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = 64$

$\sigma = 64$.

Given, $\bar{x} = 150$

1. Null hypothesis

The dice is biased.

2. Alternative hypothesis

The dice is unbiased.

3. Level of significance

$\alpha = 5\%$.

4. Test statistic

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

$= \frac{150 - 128}{8}$

$= 2.75$

$|Z| = 2.75$

5. Conclusion

At $\alpha = 5\%$, $Z_\alpha = 1.96$

$|Z| = 2.75$

$|Z| > Z_\alpha$

$\therefore$ The null hypothesis is rejected.

i.e. the dice is unbiased.

Under large sample test we will see 4 important tests to test the significance

1. Test of significance for single mean
2. Test of significance for difference of means
3. Test of significance for single proportion
4. Test of significance for difference of means.

1. Test of significance for single mean (large samples $n > 30$)

Let a random sample of size $n > 30$ has sample mean $\bar{x}$, μ be the population mean. Also, the population mean μ has a specified value μ_0.

Working rule:

1. Null hypothesis
2. Alternative hypothesis
3. level of significance (α)
4. Test of statistic

case (i): When the S.D (σ) of population is known.

In this case S.E of $\bar{x} = \frac{\sigma}{\sqrt{n}}$

$$\boxed{Z = \frac{\bar{x} - \mu}{S.E \text{ of } \bar{x}} = \frac{x - \mu}{\frac{\sigma}{\sqrt{n}}}}$$

case (ii): When the S.D (σ) of population is not known then we can take S.D of sample s.

$$\boxed{Z = \frac{\bar{x} - \mu}{S.E \text{ of } \bar{x}} = \frac{x - \mu}{\frac{s}{\sqrt{n}}}}$$

5. Conclusion

If $|Z| < Z_\alpha$, we accept the null hypothesis

If $|Z| > Z_\alpha$, we reject the null hypothesis

It is claimed that a random sample of 49 tyres has a mean life of 15200 km. This sample was dracon from a population whose mean is 15150 km and a standard deviation is 1200 km. Test the ~~hagp~~ significance at 0.05 level.

Given that,

$$n = 49$$

Sample mean $(\bar{x}) = 15200$

Population mean $(\mu) = 15150$.

standard deviation $(\sigma) = 1200$.

1. Null hypothesis

$$\mu = 15150$$

2. Alternative hypothesis

$$\mu \neq 15150$$

3. level of significance

$$\alpha = 0.05 = 5\%.$$

4. Test of statistic

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

$$z = \frac{15200 - 15150}{\frac{1200}{\sqrt{49}}}$$

$$z = \frac{50 \times 7}{1200}$$

$$z = 0.292$$

$$|z| = 0.292$$

5. Conclusion

At $\alpha = 5\%$, $z_\alpha = 1.96$.

$$|z| = 0.292.$$

$|z| < z_\alpha$ i.e, $0.292 < 1.96$.

$\therefore$ The null hypothesis is accepted.

$\therefore \mu = 15150$.

An oceanographer wants to check whether the depth of the ocean in a certain region is 57.4 fathoms, has had previously been recorded. what can he conclude at the 0.05 level of significance if reading taken at 40 random locations in the given region yielded a mean of 59.1 fathoms with a S.D of 5.2 fathoms.

Given that,

$n=40$

$\bar{x} = 59.1$ fathoms

$\sigma = 5.2$

$\mu = 57.4$.

1. Null hypothesis

$\mu = 57.4$

2. Alternative hypothesis

$\mu \neq 57.4$

3. level of significance

$\alpha = 0.05 = 5\%$.

4. Test of statistic

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

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

$$= \frac{1.7}{0.8222}$$

$z = 2.067$

$|z| = 2.067$

5. Conclusion

At $\alpha = 95\%$, $z_\alpha = 1.96$

$|z| = 2.067$

$|z| > z_\alpha$

i.e, $2.067 > 1.96$.

$\therefore$ the null hypothesis is rejected.

i.e, $\mu \neq 57.4$

An ambulance service claims that it takes on the average less than 10 min to reach its destinati in emergency calls. A sample of 36 calls has a mean of 11 min and variance of 16 min. Test the claim at 0.05 level of significance.

Given that,

$$n = 36$$
$$\bar{X} = 11$$
$$\mu = 10.$$
$$\sigma^2 = 16$$
$$\sigma = 4$$

1. Null hypothesis

$$\mu = 10.$$

2. Alternative hypothesis

$$\mu \neq 10$$

3. level of significance.

$$\alpha = 0.05 = 5\%.$$

4. Test of statistic

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

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

$$= 1.5$$

$$|Z| = 1.5$$

5. Conclusion

At $\alpha = 5\%$, $Z_\alpha = 1.96$

$$|Z| = 1.5$$

Here, $|Z| < Z_\alpha$

i.e, $1.5 < 1.96$.

$\therefore$ The null hypothesis is accepted.

i.e, $\mu = 10$.

A sample of 400 items is taken from a population whose S.D is 10. The mean of the sample ~~of~~ is 40. Test whether the sample has population with mean 38. Also calculate 95% confidence interval for population.

Given that,

$n = 400$

$\bar{x} = 40$

$\sigma = 10.$

$\mu = 38$

i. Null hypothesis

$H_0 : \mu = 38$

ii. Alternative hypothesis

$H_1 : \mu \neq 38$

iii. level of significance

$\alpha = 0.05 = 5\%.$

iv. Test of statistic.

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

$$= \frac{40 - 38}{\frac{\cancel{38}10}{\sqrt{400}}}$$

$$= \frac{2(20)}{10}$$

$Z = 4$

v. Conclusion

At $\alpha = 5\%$, $Z_\alpha = \cancel{\pm}\, 1.96$

$|Z| = 4$

$|Z| > Z_\alpha$ i.e, $4 > 1.96.$

$\therefore$ The null hypothesis is rejected.

i.e, $\mu \neq 38$

Confidence interval:

At 95%,

Confidence interval is $[\bar{x} - E, \bar{x} + E]$

$$E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1.96 \frac{(10)}{\sqrt{400}} = 0.98.$$

confidence interval $= [40 - 0.98, 40 + 0.98]$

$= [39.02, 40.98]$

According to the norms for a mechanical aptitude test, persons who are 18 years old have an average height of 73.2 with a standard deviation of 8.6. If 4 randomly selected persons of that age averaged 76.7. test the hypothesis $\mu = 73.2$ against the alternative hypothesis $\mu > 73.2$ at 0.01 level of significance

Given that,

$(0.5 - 0.01 = 0.49)$

$n = 4$

$\sigma = 8.6$

$\bar{x} = 76.7$

$\mu = 73.2$

1. Null hypothesis

$H_0 : \mu = 73.2$

2. Alternative hypothesis

$H_1 : \mu > 73.2$ (one tailed test)

3. level of significance

$\alpha = 0.01 = 1\%$

4. Test of statistic

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{76.7 - 73.2}{\frac{8.6}{\sqrt{4}}}$$

$z = 0.813$

5. Conclusion

At $\alpha = 1\%$, $z_\alpha = 2.33$

$|z| = 0.813$

$|z| < z_\alpha$

i.e, $0.813 < 2.33$

$\therefore$ The null hypothesis is accepted

i.e, $\mu = 73.2$

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

Given that,

0.5 - 0.025

1.89

$$n = 60$$
$$\bar{x} = 33.8$$
$$\sigma = 6.1$$
$$\mu = 32.6.$$

1. Null hypothesis.

$H_0: \mu = 32.6$

2. Alternative hypothesis

$H_1: \mu > 32.6$

3. level of significance

$\alpha = 0.025 = 2.5\%$

4. Test of statistic

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

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

$$z = 1.523$$

5. Conclusion

At $\alpha = 0.025$, $z_\alpha = 1.96$

$|z| = 1.523$.

$|z| < z_\alpha$

i.e, $1.523 < 1.96$

∴ The null hyopthesis is accepted.

i.e, $\mu = 32.6$

∴ we can't rejecte null hypothesis.

The length of life X of certain Computers is approximately normally distributed with mean 800 hrs and S.D 40 hrs. If a random sample of 30 Computes has an average life of 788 hrs. Test the null hypothesis that $\mu = 800$ hrs against the alternative that $\mu \neq 800$ at ① 0.5 % ② 1% ③ 4% ④ 5%. ⑤ 10% ⑥ 15%

Given that,

$$n = 30$$
$$\bar{x} = 788$$
$$\mu = 800$$
$$S.D = 40 = \sigma$$

1. Null hypothesis

$$H_0 : \mu = 800$$

2. Alternative hypothesis

$$H_1 : \mu \neq 800 \quad \text{(Two tailed)}$$

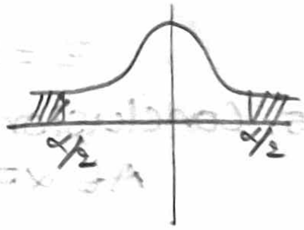
3. level of significance

$$\alpha$$

4. Test of statistic

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

$$= \frac{788 - 800}{\frac{40}{\sqrt{30}}}$$



$$z = -1.643$$

$$|z| = 1.643.$$

5. Conclusion:

case ①: At $\alpha = 0.5\% = 0.005$

$$\alpha/2 = \frac{0.005}{2} = 0.0025. \qquad 0.5 - 0.0025 = 0.4975$$

$$z_{\alpha/2} = 2.81$$

$$|z| = 1.643$$

$$|z| < z_\alpha$$

∴ Null hypothesis is accepted.

i-e, $\mu = 800$.

case ②: $\alpha = 1\% = 0.01$

$\alpha/2 = \frac{0.01}{2} = 0.005$ $\qquad$ $0.5 - 0.005 = 0.495$

$Z_\alpha = 2.58$

$|Z| = 1.643$

$|Z| < Z_\alpha$

$\therefore$ Null hypothesis is accepted.

$\therefore \mu = 800$

case ③: $\alpha = 4\% = 0.04$

$\frac{\alpha}{2} = \frac{0.04}{2} = 0.02$ $\qquad$ $0.5 - 0.02 = 0.48$

$Z_\alpha = 2.06$

$|Z| = 1.643$

$|Z| < Z_\alpha$

i.e, $1.643 < 2.06$

$\therefore$ Null hypothesis is accepted

$\therefore \mu = 800$

case ④: $\alpha = 5\% = 0.05$

$Z_\alpha = 1.96$

$|Z| = 1.643$

$|Z| < Z_\alpha$

i.e, $1.643 < 1.96$

$\therefore$ Null hypothesis is accepted.

i.e, $\mu = 800$

case ⑤: $10\% = \alpha = 0.1$

$\alpha/2 = \frac{0.1}{2} = 0.05$ $\qquad$ $0.5 - 0.05 = 0.45$

$Z_\alpha = 1.65$

$|Z| = 1.643$

$|Z| < Z_\alpha$

i.e, $1.65 > 1.643$

Null hypothesis is accepted

i.e, $\mu = 800$

Case ③: $\alpha = 15\% = 0.15$

$\frac{\alpha}{2} = \frac{0.15}{2} = 0.075$ $\quad$ $0.5 - 0.075 = 0.425$

$Z_\alpha = 1.44$

$|Z| > Z_\alpha$

i.e, $1.643 > 1.44$

$\therefore$ Null hypothesis is rejected

i.e, $\mu \neq 800$.

2. Test of significance for difference of means of two large samples:

Let $\bar{x}_1$ and $\bar{x}_2$ be the sample means of two independent large samples sizes n_1 and n_2 drawn from ~~po~~ two populations having mean μ_1 and μ_2 and S.D's σ_1 and σ_2. To test whether two population means are equal

1. Null hypothesis : $H_0 : \mu_1 = \mu_2$
2. Alternative hypothesis : $H_1 : \mu_1 \neq \mu_2$
3. level of significance α
4. Test of statistics

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{S.E \text{ of } (\bar{x}_1 - \bar{x}_2)}$$

when $\delta = \mu_1 - \mu_2$

if $\delta = 0$ then thes population means are equal

If $\delta \neq 0$ then two population means are different.

if $\mu_1 = \mu_2$, $Z = \dfrac{\bar{x}_1 - x_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

5. Conclusion

if $|Z| < Z_\alpha$ then null hypothesis is accepted.

if $|Z| > Z_\alpha$ then null hypothesis is rejected.

The mean yeild of wheat from a district A was 210 lbs with S.D 10 lbs per acre from a sample of 100 Plots. In another district the mean yeild was 220 lbs with S.D 12 lbs from a sample of 150 plots. Assuming that the S.D of yeild in the entire state was 11 lbs, test whether there is any significant difference b/w the mean yield of Crops in the two districts.

Given that,

$$\bar{x}_1 = 210\,lbs, \quad \bar{x}_2 = 220\,lbs$$
$$n_1 = 100, \quad n_2 = 150$$
$$S_1 = 10\,lbs, \quad S_2 = 12\,lbs$$

Population standard deviation $(\sigma) = 11$.

1. Null hypothesis

$H_0: \mu_1 = \mu_2$ i.e, there is no difference b/w mean yield of Crops in two districts

2. Alternate hypothesis

$H_1: \mu_1 \neq \mu_2$

3. level of significance

$\alpha = 5\% = 0.05$

4. Test of statistic

$$Z = \frac{x_1 - x_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{210-220}{\sqrt{\frac{100}{100} + \frac{144}{150}}} \qquad = \frac{-10}{1.42}$$

$$= -7.042.$$

$$= -7.1428 \qquad |Z| = 7.042$$

$$|Z| = 7.1428$$

5. Conclusion

At $\alpha = 5\%$, $Z_\alpha = 1.96$

$|Z| = 7.1428$ $\quad |Z| = 7.042$

$|Z| > Z_\alpha$

∴ the null hypothesis is rejected.

$\mu_1 \neq \mu_2$.

Samples of students were drawn from two universities and from their weights in kg, mean and S.D are calculated and as shown below. Make a large sample test to test the significance of the difference between the means.

	Mean	S.D	Size of sample
University A	55	10	400
University B	57	15	100

Given that,

$n_1 = 400 \qquad n_2 = 100$

$S_1 = 10 \qquad S_2 = 15$

$\bar{x}_1 = 55 \qquad \bar{x}_2 = 57$

1. Null hypothesis

$H_0 : \mu_1 = \mu_2 \quad \bar{x}_1 = \bar{x}_2$

2. Alternate hypothesis

$H_1 : \mu_1 \neq \mu_2 \quad \bar{x}_1 \neq \bar{x}_2$ (Two tailed)

3. level of significance

$\alpha = 0.05 = 5\%$.

4. Test of statistic

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

$$z = \frac{55 - 57}{\sqrt{\frac{100}{400} + \frac{225}{100}}}$$

$z = -1.264$

$|z| = 1.264$

5. conclusion

At $\alpha = 5\%$, $z_\alpha = 1.96$

$|z| = 1.264$

$|z| < z_\alpha$ i.e, $1.264 < 1.96$

$\therefore$ The null hypothesis is accepted

i.e, $x_1 = x_2$

The means of two large samples of size 1000 & 2000 members are 67.5 inches and 68 inches respectively. Can the samples be regarded as drawn from the same population of S.D 2.5 inches.

Given that,

$n_1 = 1000 \qquad n_2 = 2000$

$\bar{x}_1 = 67.5 \qquad \bar{x}_2 = 68$

S.D of population $(\sigma) = 2.5$

1. Null hypothesis

$H_0 : \mu_1 = \mu_2$

2. Alternate hypothesis

$H_1 : \mu_1 \neq \mu_2$ (Two tailed test)

3. level of significance

$\alpha = 0.05 = 5\%$

4. Test of statistic

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{67.5 - 68}{\sqrt{\frac{(2.5)^2}{1000} + \frac{(2.5)^2}{2000}}}$$

$$= -5.163$$

$|z| = 5.163$

5. Conclusion;

At $\alpha = 5\%$, $z_\alpha = 1.96$

$|z| = 5.163$

$|z| > z_\alpha$

i.e, $5.163 > 1.96$

$\therefore$ The null hypothesis H_0 is rejected.

i.e, $\mu_1 \neq \mu_2$.

Note: 1. If the samples have been drawn from the population with common S.D (σ) then

$\sigma_1^2 = \sigma_2^2 = \sigma$.

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

2. If the two samples are drawn from the 2 populations with unknown S.D's standard variances σ_1^2 and σ_2^2 then σ_1^2 and σ_2^2 can be replaced by sample variance s_1^2 and s_2^2 provided both the samples n_1 and n_2 are large.

$$\therefore z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_1}}}$$

3. Test of significance for single proportion:

To the test the hypothesis that the population P in the population has specified value P_0.

1. Null hypothesis: $H_0 : P = P_0$
2. Alternative hypothesis: $H_1 : P \neq P_0$ (or) $P > P_0$ (or) $P < P_0$
3. level of significance α.
4. Test of statistic

$$z = \frac{p - P_0}{S.E \text{ of } p} = \frac{p - P_0}{\sqrt{\frac{PQ}{n}}}$$

where p is the sample proportion

P is population proportion.

1. A manufacturer claimed that atleast 95% of the equipment which he supplied to a factor confirmed to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 were faulty. Test his claim at 5% ~~of~~ level of significance.

Given that,

Population proportion $P = 95\% = \frac{95}{100} = 0.95$

Sample size $(n) = 200$.

~~Sample proportion $p = 200 - 18 = 182$.~~

i.e, No. of pieces confirmed to specifications $= 200 - 18 = 182$

proportion of pies confirming to specifications $p = \frac{x}{n} = \frac{182}{200} = 0.91$

$Q = 1 - P = 1 - 0.95 = 0.05$.

1. Null hypothesis: $H_0: p = P_0$
2. ~~Alt~~ Alternative hypothesis: $H_1: p \leq P_0$ (~~two~~ one tailed)
3. level of significance: $\alpha = 0.05 = 5\%$.
4. Test of statistic

$$z = \frac{p - P_0}{\sqrt{\frac{PQ}{n}}}$$

$$= \frac{0.91 - 0.95}{\sqrt{\frac{(0.95)(0.05)}{200}}}$$

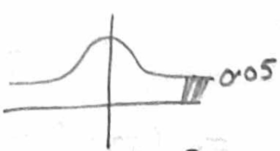
$$= -2.59$$

$$|z| = 2.59.$$

5. Conclusion

At $\alpha = 5\%$, $z_\alpha = 1.65$

$|z| = 2.59$



$0.5 - 0.05 = 0.45$

$|z| > z_\alpha$

$\therefore$ The Null hypothesis rejected.

i.e, $p < P$

$p < 0.95$.

2. In a big ~~b~~ city, 325 men out of 600 men were found to be smokers. Thus this information supports the conclusion that the majority of men in this city are smokers.

Given that,

$n = 600$

$x = 325$

p = proportion of sample of smokers

$= x/n = \frac{325}{600} = 0.5417.$

population proportion $(P) = 50\% = \frac{50}{100} = \frac{1}{2}$ of smokers

$Q = 1 - P = 1 - \frac{1}{2} = \frac{1}{2} = 0.5.$

1. Null hypothesis

$H_0 : p = P \Rightarrow P = 0.5$

2. Alternative hypothesis

$H_1 : P > 0.5$ (one tailed)

3. Level of significance

$\alpha = 5\% = 0.05$

4. Test of statistic

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$$= \frac{0.5417 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{600}}}$$

$= 2.042$

$|z| = 2.042.$

5. Conclusion

At $\alpha = 5\%$, $z_\alpha = 1.65$

$|z| = 2.042$

$|z| > z_\alpha$

$\therefore$ The null hypothesis is rejected.

$\therefore$ $P > 0.5.$

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

Given that, $n = 9000$

$x = 3220$

p = Proportion of success of getting 3 or 4 in 9000 throws

$$= \frac{x}{n} = \frac{3220}{9000} = 0.3578.$$

P = Population proportion of getting 3 or 4

$$= 1/6 + 1/6 = 2/6 = 1/3 = 0.334.$$

$$Q = 1 - P = 1 - 1/3 = 2/3 = 0.667$$

1. Null biased hypothesis

H_0: The die is unbiased

2. Alternative hypothesis

H_1: The die is biased (Two-tailed)

3. Level of significance

$\alpha = 5\% = 0.05$

4. Test of hypothesis

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$$= \frac{0.3578 - 0.334}{\sqrt{\frac{(0.334)(0.667)}{9000}}}$$

$$= 4.78$$

5. Conclusion

At $\alpha = 5\%$, $Z_\alpha = 1.96$

$|Z| = 4.78$

$|Z| > Z_\alpha$

$\therefore$ The null hypothesis is rejected.

i.e, The dice is biased.

4. In a sample of 1000 people in karnataka, 540 are rice eaters and rest of are wheat eaters. Can we assume that both rice and wheat are equally popular in the state at 1% level of significance.

Given that,

$$n = 1000$$

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

P = Population proportion of rice eaters $= 50\% = \frac{50}{100} = 0.5$

$$Q = 1 - P = 1 - 0.5 = 0.5$$

1. Null hypothesis

$H_0 : P = 0.5$

2. Alternative hypothesis

$H_1 : P \neq 0.5$ (Two tailed)

3. Level of significance

$\alpha = 1\% = 0.01$

4. Test of statistic

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$$= \frac{0.54 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{1000}}}$$

$$= 2.529$$

5. Conclusion

At $\alpha = 1\%$, $z_\alpha = 2.58$

$|z| = 2.529$

$0.5 - \frac{0.01}{2}$

$= 0.5 - 0.005$

$= 0.495$

$|z| < z_\alpha$

$\therefore$ The Null hypothesis is accepted.

i.e, $P = 0.5$.

5. In a sample of 500 from a village in Rajasthan 280 are found to be wheat eaters and rest are rice eaters. can we assume that the both are equally popular.

Given that,

$n = 500$

$x = 280$

$p = \frac{x}{n} = \frac{280}{500} = 0.56$ = sample proportion

population proportion $P = 50\% = \frac{50}{100} = 0.5$

$Q = 1 - P = 1 - 0.5 = 0.5$

1. Null hypothesis

$H_0: p = P$ i.e, $P = 0.5$

2. Alternative hypothesis

$H_1: P \neq 0.5$ (two tailed)

3. level of significance

$\alpha = 5\% = 0.05$

4. Test of statistic

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$$= \frac{0.56 - 0.5}{\sqrt{\frac{0.5(0.5)}{500}}}$$

$$= 2.68$$

5. Conclusion

At $\alpha = 5\%$, $z_\alpha = 1.96$

$|z| = 2.68$

$|z| > z_\alpha$

∴ The null hypothesis is ~~accepted~~ rejected.

i.e, $P \neq 0.5$.

6. A manufacturer claims that only 4% of his products are defective. A random sample of 500 were taken among which 100 were defective. Test the hypothesis at 0.05 level.

Given that,

$$n = 500$$

$$x = 100$$

$$p = \frac{x}{n} = \frac{100}{500} = 1/5 = 0.2$$

$$P = 4\% = \frac{4}{100} = \frac{1}{25} = 0.04$$

$$Q = 1 - P = 1 - 0.04 = 0.96$$

1. Null hypothesis

$H_0 : p = P$ i.e, $P = 0.04$

2. Alternative hypothesis

$H_1 : P \neq 0.04$ (Two tailed).

3. level of significance

$\alpha = 0.05 = 5\%$.

4. Test statistic

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$$= \frac{0.2 - 0.04}{\sqrt{\frac{0.04(0.96)}{500}}}$$

$$= 18.257$$

5. Conclusion

At $\alpha = 5\%$, $Z_\alpha = 1.96$

$|Z| = 18.257$

$|Z| > Z_\alpha$

i.e, $18.257 > 1.96$

$\therefore$ The null hypothesis is rejected.

i.e, $P \neq 0.04$.

7. 20 people were attacked by a disease and only 18 survived. Will you reject the hypothesis that the survival rate if attacked by this disease is 85% in favour of the hypothesis i.e, more at 5% level.

Given that,

$n = 20$

$x = 18$

$p = \frac{x}{n} = \frac{18}{20} = 0.9$

$P = 85\% = \frac{85}{100} = 0.85$

$Q = 1 - P = 1 - 0.85 = 0.15$.

1. Null hypothesis

$H_0 : P = p$ i.e, $P = 0.85$

2. Alternative hypothesis

$H_1 : P > 0.85$ (one tailed)

3. level of significance

$\alpha = 5\% = 0.05$

4. Test of statistic

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$$= \frac{0.9 - 0.85}{\sqrt{\frac{0.85(0.15)}{20}}}$$

$= 0.626$.

5. Conclusion

At $\alpha = 5\%$, $Z_\alpha = 1.65$

$|Z| = 0.626$.

$0.5 - \frac{0.05}{2}$

$= 0.5 - 0.05$

$= 0.45$.

$|Z| < Z_\alpha$

∴ The null hypothesis is accepted.

∴ $P = 0.85$

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 a corresponding true percentage.

Given that,

$$n = 160, x = 24$$

Confidence interval $= P \pm E$

$= [P - E, P + E]$

$$\boxed{C.I = \left[P - Z_{\alpha/2}\sqrt{\frac{PQ}{n}}, P + Z_{\alpha/2}\sqrt{\frac{PQ}{n}}\right]}$$

$$P = \frac{x}{n} = \frac{24}{160} = 0.15$$

$$Q = 1 - P = 1 - 0.15 = 0.85.$$

$$0.5 - \frac{0.01}{2}$$

$$= 0.5 - 0.005$$

$$= 0.495.$$

~~E(1) +~~

$Z_{\alpha/2}$ at 99% Confidence $= 2.58$.

$$E = Z_{\alpha/2}\sqrt{\frac{PQ}{n}}$$

$$= 2.58\sqrt{\frac{0.15(0.85)}{160}}$$

$$= 0.0728$$

$$= 0.073.$$

$$C.I = [0.15 - 0.073, 0.15 + 0.073]$$

$$= [0.077, 0.223].$$

Test of hypo significance for difference of proportions:

Let p_1, p_2 be the sample proportions in large random samples of sizes n_1 and n_2 drawn from two populations having proportions P_1 and P_2.

To test whether the two samples have been drawn from the same population.

1. Null hypothesis

$H_0: P_1 = P_2$

2. Alternative hypothesis

$H_1: P_1 \neq P_2$ (or) $P_1 > P_2$ (or) $P_1 < P_2$

3. level of significance α

4. Test of statistic

there are two ways of finding test statistic z.

i. When the population proportions P_1 and P_2 are known.

In this case $Q_1 = 1 - P_1$

$Q_2 = 1 - P_2$

small

$$\therefore z = \frac{p_1 - p_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

ii. when the population proportions P_1 and P_2 are not known but sample proportions p_1 and p_2 are known. In this case, by the method of pooling sample proportions of two samples or estimated value of P.

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

$$p_1 = \frac{x_1}{n_1}, \quad p_2 = \frac{x_2}{n_2}$$

$$\Rightarrow P = \frac{x_1 + x_2}{n_1 + n_2}.$$

$q = 1 - P$

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

5. Conclusion

i. If $|z| < Z_\alpha$ then we accept Null hypothesis H_0

ii. If $|z| > Z_\alpha$ then we reject Null hypothesis H_1.

1. In two large population there are 30% and 25% respectively are fair haired people. Is this difference likely to be hidden in the samples are 1200 and 900 respectively from the two populations.

Given that,

$$n_2 = 900 \qquad n_1 = 1200$$

P_1 = proportion of fair haired people in first Population $= 30\% = 0.3$.

P_2 = proportion of fair haired people in second population $= 25\% = 0.25$.

$$Q_1 = 1 - P_1 = 1 - 0.3 = 0.7$$

$$Q_2 = 1 - P_2 = 1 - 0.25 = 0.75$$

1. Null hypothesis : $H_0 : P_1 = P_2$
2. Alternative hypothesis : $H_1 : P_1 \neq P_2$ (Two tailed).
3. level of significance $\alpha = 5\% = 0.05$.
4. Test of statistic

$$z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n}}}$$

$$= \frac{0.3 - 0.25}{\sqrt{\frac{(0.7)(0.3)}{1200} + \frac{(0.25)(0.75)}{900}}}$$

$$z = \cancel{7.1533} \cdot 2.553$$

5. Conclusion

$$|z| = \cancel{2.1533}\ 2.553$$

At $\alpha = 5\%$, $Z_\alpha = 1.96$.

$$|z| > Z_\alpha$$

$\therefore$ The null hypothesis is rejected

$\therefore P_1 \neq P_2$

1. In a city A 20% of random sample of 900 school boys has a certain slight physical defect. In another city B 18.5% of a random sample of 1600 school boys has the same defect. Is the difference between the proportions is significant 0.05 level of significance.

Given, $n_1 = 900$, $n_2 = 1600$

$$x_1 = 20\% \text{ of } 900 = \frac{20}{100} \times 900 = 180$$

$$x_2 = 18.5\% \text{ of } 1600 = \frac{18.5}{100} \times 1600 = 296$$

$$P_1 = \frac{x_1}{n_1} = \frac{180}{900} = \frac{1}{5} = 0.2 \;(\Rightarrow q_1 = 1 - P_1 = 0.8$$

$$P_2 = \frac{x_2}{n_2} = \frac{296}{1600} = 0.185 \Rightarrow q_2 = 1 - P_2 = 1 - 0.185 = 0.815)$$

1. Null hypothesis $H_0: P_1 = P_2$

2. Alternative hypothesis $H_1: P_1 \neq P_2$ (two tailed).

3. level of significance $\alpha = 0.05 = 5\%$.

4. Test of statistic

By method of pooling,

$$P = \frac{n_1P_1 + n_2P_2}{n_1 + n_2} = \frac{180 + 296}{900 + 1600} = 0.1904.$$

$$q = 1 - P = 1 - 0.1904 = 0.8096.$$

Now,
$$z = \frac{P_1 - P_2}{\sqrt{\frac{pq}{n_1} + \frac{pq}{n_2}}}$$

$$= \frac{0.2 - 0.185}{\sqrt{(0.1904)(0.8096)\left[\frac{1}{900} + \frac{1}{1600}\right]}}$$

$$= 0.9169.$$

5. Conclusion

At $\alpha = 5\%$, $z_\alpha = 1.96$

$|z| = 0.9169$.

$|z| < z_\alpha$

$\therefore$ The null hypthesis is accepted

i.e, $P_1 = P_2$.

3. In a sample of 600 students of a certain college, 400 are found to use ball pens. In another college from a sample of 900 students, 450 were found to use ball pens. Test whether two colleges are significantly different with respect to the habit of using ball pens.

Given that,

$n_1 = 600$ $\quad$ $n_2 = 900$

$x_1 = 400$ $\quad$ $x_2 = 450$

$$P_1 = \frac{x_1}{n} = \frac{400}{600} = 0.66$$

$$P_2 = \frac{x_2}{n_2} = \frac{450}{900} = 0.5.$$

By method of pooling,

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{400+450}{600+900} = 0.563$$

$$q = 1 - P = 1 - 0.563 = \cancel{0.37}\ 0.44$$

1. Null hypothesis $H_0 : P_1 = P_2$
2. Alternative hypothesis $H_1 : P_1 \neq P_2$ (two tailed)
3. level of significance $\alpha = 5\% = 0.05$
4. Test of statistic

$$z = \frac{P_1 - P_2}{\sqrt{Pq\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

$$= \frac{0.66 - 0.5}{\sqrt{(0.56)(0.44)\left[\frac{1}{600} + \frac{1}{900}\right]}}$$

$$= 6.39.$$

5. Conclusion

At $\alpha = 5\%$, $z_\alpha = 1.96$

$|z| = 6.39$

$|z| > z_\alpha$

$\therefore$ Null hypothesis is rejected.

i.e. $P_1 \neq P_2$

4. A random sample of 1000 persons from a town A, 400 are found to be consumers of wheat. From a sample of 800 from town B, 400 are found to be consumers of wheat. Do this data reveal a significant difference b/w town A and town B, so far the proportion of wheat consumers is considered.