Unit -4

Estimation and Testing of hypothesis,

Large sample tests

sampling theory:-

population :- population is totality of statistical data forming a subject of investigation.

Eg:- ① The population of heights of Indians.

② . The population of nationalised banks in India.

size of population :- size of population is number of observations in population and is denoted "N"

population may be classified into two ways. They are

(i) finite population.

(ii) Infinite population.

(i) **Finite population** :-

The no. of observations in the population is finite then the population is called finite population.

(ii) **Infinite population** :-

The no. of observations in the population is infinite then the population is called infinite population.

Ex :- ① The heights of residence in a certain city is a finite population.

② If there are 600 students in a school that we classified according to blood type, here we say that we have a population of size 600 is called finite population.

③ The observations obtained by measuring the atmospheric pressure every day from the past on into the future, is the example of infinite population.

<u>Sampling</u> :- The process of selection of a sample is called sampling.

<u>Ex</u> :- To assess the quality of a bag of rice, we examine only a portion of it by taking handful of it from the bag and then decide to purchase it (or) not

Thus in estimating the characteristics of the population, instead of enumerating entire population, only the individuals in the sample are examined. Then the sample characteristics are utilised to estimate the population.

To eliminate any possibility of bias in the sampling procedure it is desirable to choose a random sample in the sense that the observations are made independently and at random.

<u>sample</u> :- A sample is a subset of population and the size of the sample is denoted by "n"

samples are classified in to two types. They are

(i) small sample

(ii) Large sample

(i) <u>small sample</u> :- If the no. of items in the sample $(n) < 30$ items then the sample is called small sample

(ii) <u>Large sample</u> :- If the no. of items in the sample ie $n \geq 30$ items then the sample is called large sample.

Types of sampling :-

The process of selecting the sample from the population in four ways. They are

i). Random sampling
ii). stratified sampling
iii). systematic sampling
iv). purposive sampling

(i) Random sampling :-

Random sampling is the process of drawing a sample from a population in such a way that each member of the population has an equal chance of being included in the sample.

The random sampling is classified into two ways. They are

(a) Simple random sampling with replacement (SRSWR)
(b) Simple random sampling without replacement (SRSWOR)

(a) SRSWR :- one sample is selected from the population and replace the sample into population. Again the second sample is selected from the population and second sample is replaced in the population before drawing the 3^{rd} sample. continuing the same process for all selections is known as simple random sampling with replacement.

Let "N" be the size of the population and "n" be the size of the sample then the no. of samples are drawn in SRSWR is N^n

(b) SRSWOR :- A sample is selected from the population, the first sample excludes the population and the 2^{nd} sample is drawn from the remaining population. The 3^{rd} sample is selected from the population excludes sample 1 and sample 2

This process is continue for selection of all samples, is known as simple random sampling with out replacement.

Let "N" be the size of the population and "n" be the size of the sample then the no. of samples are selected from the population in SRSWOR is N_{C_n}

Ex :- ① From a finite population of size 5, number of samples of size "2" can be drawn in SRSWR is $N^n = 5^2 = 25$

② From a finite population of size 5, number of samples of size "2" can be drawn in SRSWOR is $N_{C_n} = 5_{C_2} = \frac{5 \times 4}{2 \times 1} = 10$

(ii) stratified sampling :- In this type of sampling the population is first subdivided in to several parts (or small groups) according to some relevant characteristics. so that each part is less homogeneous and each part is called stratum. from each stratum the samples are selected at random and combined together to give information about the population is called stratified sampling.

(iii) systematic sampling :- In systematic random sampling, population is arranged in some order. From the list we randomly select the first sample Element from the first "K" Elements on the population list. There after, we select every K^{th} Element on the list. In this method only the first member is choosen at random.

(iv) Purposive sampling :- In this sampling, the sample elements are choosen according to convenience and personal choice of the individual who selects the sample. This method is suitable when sample is small. Here the investigator must have the rough knowledge about the population.

For example, In a factory, suppose different types of items are produced, suppose there are some complaints against a product. Then they consider only the products or individual who have complaints and ignore the others.

Parameter :- Any statistical measure computed from population data is known as parameter

For example the statistical constants of the population namely population mean μ and population variance σ^2 are usually referred as the parameters.

Statistic :- Any statistical measure computed from sample data is known as statistic.

For example the statistical constants of the sample namely sample mean $\bar{x}$ and sample standard deviation "s" are the statistics

Note :- value of a statistic varies from sample to sample ($\because$ units of samples are not same), but parameters always remains constant

Sampling Distribution (or) sampling Distribution of a statistic :-

sampling distribution of a statistic is called sampling distribution

It helps us

(i) To estimate unknown population parameter from known statistic.

(ii) To set confidence limits of parameter with in which the parameter values are expected to lie

(iii) To test a hypothesis and to draw a statistical inference from it.

central limit theorem :-

If $\bar{x}$ be mean of a sample size n, drawn from a population with mean μ and standard deviation "σ" then

$Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ is a random variable whose distribution function approaches that of standard normal distribution $N(0,1)$ as $n \to \infty$

*** Note :- If a random sample of size "n" is taken from infinite population having mean μ and standard deviation "σ" then the expected value of sample mean is population mean i.e $E(\bar{x}) = \mu$ and variance $V(\bar{x}) = \dfrac{\sigma^2}{n}$, standard deviation of sample mean $S.D(\bar{x}) = \sqrt{\text{variance}} = \dfrac{\sigma}{\sqrt{n}}$

standard error (S.E) :-

The standard deviation of sampling distribution of a statistic is known as its standard error and is denoted by S.E

The standard error of means is : $S.E(\bar{x}) = \dfrac{\sigma}{\sqrt{n}}$

Sample Mean :-

If $x_1, x_2, x_3 ---- x_n$ represents the random sample size "n" then the sample mean is defined by the statistic

$$\bar{x} = \frac{x_1 + x_2 + --- + x_n}{n}$$

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

<u>sample variance</u> :-

If $x_1, x_2, x_3 \cdots x_n$ represents the random sample of size "n" then the sample variance is denoted by "s^2" and is defined as

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \quad \text{when } \sigma \text{ is known (or) for } n \geq 30$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \quad \text{when } \sigma \text{ is unknown (or) for } n < 30$$

<u>NOTE</u> :- 1. Mean of the population is denoted by "μ" and is defined as

$$\mu = \frac{\sum x_i}{N} \quad \text{where "N" is size of population}$$

2. Variance of the population is denoted by "σ^2" and is defined as

$$\sigma^2 = \frac{1}{N}\sum(x_i-\mu)^2$$

where μ is mean of population

<u>Theorem</u> :- If a random sample of size "n" is taken from infinite population having mean μ and variance σ^2 then the mean of sample mean is population mean i.e $E(\bar{x}) = \mu_{\bar{x}} = \mu$ and variance of the sample mean is

$$\sigma_{\bar{x}}^2 = V(\bar{x}) = \frac{\sigma^2}{n}$$

<u>Proof</u> :- Let infinite population mean be μ and variance σ^2

Let $x_1, x_2 \cdots x_n$ be "n" independent variables has taken from the same population with mean μ and variance σ^2

Mean of the sample mean

$$\mu_{\bar{x}} = E(\bar{x}) = E\left[\frac{x_1+x_2+\cdots+x_n}{n}\right]$$

$$= \frac{1}{n}\left[E(x_1)+E(x_2)+\cdots+E(x_n)\right]$$

$$= \frac{1}{n}\left[\mu+\mu+\cdots+\mu\right]$$

$$= \frac{1}{\cancel{n}}(\cancel{n}\mu)$$

$$\mu_{\bar{x}} = E(\bar{x}) = \mu$$

variance of the sample mean

$$\sigma_{\bar{x}}^2 = V(\bar{x}) = V\left[\frac{x_1+x_2+\cdots+x_n}{n}\right]$$

$$= \frac{1}{n^2}\left[V(x_1)+V(x_2)+\cdots+V(x_n)\right]$$

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

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

$$\sigma_{\bar{x}}^2 = \frac{1}{n}\sigma^2$$

$$= \frac{\sigma^2}{n}$$

standard deviation of sample mean

$$S.D(\bar{x}) = \sigma_{\bar{x}} = \sqrt{\text{variance}}$$

$$= \sqrt{\sigma_{\bar{x}}^2}$$

$$= \sqrt{\sigma^2/n}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

The sampling distribution of the mean (when σ known)

i) For finite population :-

suppose samples are drawn from a finite population without replacement (or) sampling is done without replacement then

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

standard deviation of sample means $\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1} \frac{\sigma^2}{n}}$

$$\Rightarrow \sigma_{\bar{x}}^2 = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}$$

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

ii) For infinite population:-

Suppose samples are drawn from a infinite population with replacement (or) sampling is done with replacement then

Mean of sample means $\mu_{\bar{x}} = \mu$

variance of sample means $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$

S.D of sample means $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Here $\frac{\sigma}{\sqrt{n}}$ is standard error of means.

Eg-1:- A population consists of 5 numbers 2, 3, 6, 8 and 11. consider all possible samples of size 2, that can be drawn with replacement from this population. find.

i. The mean of the population.

ii. The standard deviation of the population.

iii. The mean of the sampling distribution of means

iv. The standard deviation of sampling distribution of means

v. verify the result

Sol: Given population units are 2, 3, 6, 8 and 11

Here population size $N=5$ and sample size $n=2$

(i) mean of the population $\mu = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$

(ii) variance of the population $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$

$$= \frac{1}{5}\left[(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2\right]$$

$$= \frac{1}{5}\left[(-4)^2+(-3)^2+0+(2)^2+(5)^2\right]$$

$$= \frac{1}{5}\left[16+9+4+25\right]$$

$$= \frac{1}{5}\times 54$$

$$\sigma^2 = 10.8$$

$$\sigma = \sqrt{10.8} = 3.28$$

∴ population strandard deviation, $\sigma = \sqrt{10.8} = 3.29$

(iii) sampling with Replacement (Infinite population)

The total No. of samples sampling with Replacement is N^n

$N^n = 5^2 = 25$

25 samples of size 2.

Listing all possible samples of size 2 from population with replacement we get 25 samples i.e., {(2,2), (2,3), (2,6), (2,8), (2,11) (3,2), (3,3), (3,6), (3,8), (3,11), (6,2), (6,3), (6,6), (6,8), (6,11), (8,2) (8,3), (8,6), (8,8), (8,11), (11,2), (11,3), (11,6), (11,8), (11,11)}

The sample means are 2, 2.5, 4, 5, 6.5, 2.5, 3, 4.5, 5.5, 7, 4, 4.5, 6, 7, 8.5, 5, 5.5, 7, 8, 9.5, 6.5, 7, 8.5, 9.5, 11 $\left[\because \frac{2+2}{2}=2, \frac{2+3}{2}=2.5\right.$

mean of sampling distribution of means

$$\mu_{\bar{x}} = \frac{\text{sum of all sample means}}{25}$$

$$= \frac{2+2.5+4+5+6.5+2.5+3+4.5+5.5+7+4+4.5+6+7+8.5+5+5.5+7+8+9.5+6.5+7+8.5+9.5+11}{25}$$

$\mu_{\bar{x}} = \frac{150}{25} = 6$

(iv) The varience of sampling distribution of means is $\sigma_{\bar{x}}^2$

$$\sigma_{\bar{x}}^2 = \frac{1}{25}\{(2-6)^2+(2.5-6)^2+(4-6)^2+(5-6)^2+(6.5-6)^2+(2.5-6)^2+(3-6)^2+(4.5-6)^2+(5.5-6)^2+(7-6)^2+(4-6)^2+(4.5-6)^2+(6-6)^2+(7-6)^2+(8.5-6)^2+(5-6)^2+(5.5-6)^2+(7-6)^2+(8-6)^2+(9.5-6)^2+(6.5-6)^2+(7-6)^2+(8.5-6)^2+(9.5-6)^2+(11-6)^2\}$$

$$= \frac{1}{25}\{(-4)^2+(-3.5)^2+(-2)^2+(-1)^2+(0.5)^2+(-3.5)^2+(-3)^2+(-1.5)^2+(-0.5)^2+(1)^2+(-2)^2+(-1.5)^2+0+(1)^2+(2.5)^2+(-1)^2+(-0.5)^2+(+1)^2+(2)^2+(3.5)^2+(0.5)^2+(1)^2+(2.5)^2+(3.5)^2+(5)^2\}$$

$$= \frac{1}{25}[16+12.25+4+1+0.25+12.25+9+2.25+0.25+1+4+2.25+1+6.25+1+0.25+1+4+12.25+0.25+1+6.25+12.25+25] \quad (+0.25+1)$$

$$\sigma_{\bar{x}}^2 = \frac{1}{25}(136.25) = 5.45$$

$$\sigma_{\bar{x}} = \sqrt{\sigma_{\bar{x}}^2} = \sqrt{5.45} = 2.32$$

(v) verify the result (sampling with replacement) :-

(1) we know that $\mu_{\bar{x}} = \mu = 6$

(2) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.29}{\sqrt{2}} = 2.32$

2) Solve the above problem without replacement process

Sol:- Given population units 2, 3, 6, 8 and 11

Here population size $N=5$ and sample size $n=2$.

(i) mean of the population $\mu = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$

(ii) varience of the population $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$

$= \frac{1}{5}\left[(2-6)^2+(8-6)^2+(6-6)^2+(8-6)^2+(11-6)^2\right]$

$= \frac{1}{5}\left[16+9+4+25\right] = \frac{1}{5}\times 54$

$= 10.8$

$\sigma^2 = 10.8$

$\sigma = \sqrt{10.8} = 3.29$

iii) Sampling without Replacement :-

The total No. of samples sampling without Replacement is N_{C_n}

$N_{C_n} = 5_{C_2} = 10$

10 samples of size 2

listing all possible samples of size 2 from population without replacement we get 10 samples

∴ The samples are {(2,3), (2,6), (2,8), (2,11), (3,6), (3,8), (3,11), (6,8), (6,11), (8,11)}

The sample means are 2.5, 4, 5, 6.5, 4.5, 5.5, 7, 7, 8.5, 9.5

mean of sampling distribution of means.

$$\mu_{\bar{x}} = \frac{\text{Sum of all sample means}}{10}$$

$$= \frac{2.5+4+5+6.5+4.5+5.5+7+7+8.5+9.5}{10}$$

$$\therefore \mu_{\bar{x}} = \frac{60}{10} = 6$$

iv) variance of sampling distribution of means is

$$\sigma^2_{\bar{x}} = \frac{1}{10}\Big[(2.5-6)^2+(4-6)^2+(5-6)^2+(6.5-6)^2+(4.5-6)^2+(5.5-6)^2+(7-6)^2+(7-6)^2+(8.5-6)^2+(9.5-6)^2\Big]$$

$$= \frac{1}{10}\Big[(-3.5)^2+(-2)^2+(-1)^2+(0.5)^2+(-1.5)^2+(-0.5)^2+(1)^2+(1)^2+(2.5)^2+(3.5)^2\Big]$$

$$= \frac{1}{10}\left[12.25+4+1+0.25+2.25+0.25+1+1+6.25+12.25\right]$$

$= 40.50/10 = 4.05$

$\therefore \sigma_{\bar{x}} = 2.0125$

(v) verify the result (sampling without replacement)

i) $\mu_{\bar{x}} = \mu = 6$

ii) $\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}} = \sqrt{\frac{5-2}{5-1} \cdot \frac{10.8}{2}}$

$= \sqrt{\frac{3}{4} \cdot \frac{10.8}{2}}$

$= \sqrt{\frac{3 \times 10.8}{8}}$

$= \sqrt{\frac{32.4}{8}}$

$= \sqrt{4.05}$

$= 2.0125$ //

3) A population consists of 6 numbers 4, 8, 12, 16, 20, 24 consider all possible samples of size 2 which can be drawn without replacement and with replacement from this population. Find

i) The population mean

ii) The population S.D

iii) The mean of sampling distribution of means.

iv) The S.D of sampling distribution of means.

v) verify 3 & 4 from 1 & 2

Sol:- with replacement :—

Given population units 4, 8, 12, 16, 20, 24

Here population size $N=6$ and sample size $n=2$

i) mean of the population $\mu = \frac{4+8+12+16+20+24}{6} = \frac{84}{6} = 14$

ii) varience of the population $\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$

$= \frac{1}{6} \left[(4-14)^2 + (8-14)^2 + (12-14)^2 + (16-14)^2 + (20-14)^2 + (24-14)^2 \right]$

$$= \frac{1}{6}\left[(-10)^2+(-6)^2+(-2)^2+(2)^2+(6)^2+(10)^2\right]$$

$$= \frac{1}{6}\left[100+36+4+4+36+100\right]$$

$$\sigma^2 = \frac{280}{6} = 46.66$$

$$\sigma = 6.8308$$

$\therefore$ population of S.D, $\sigma = \sqrt{46.66} = 6.83$

(iii) Sampling with Replacement (Infinite population)

The total no. of samples sampling with Replacement is N^n

$$N^n = 6^2 = 36$$

36 samples of size 2.

Listing all possible samples of size 2 from population with Replacement we get 36 samples.

i.e, {(4,4), (4,8), (4,12), (4,16), (4,20), (4,24),
(8,4), (8,8), (8,12), (8,16), (8,20), (8,24),
(12,4), (12,8), (12,12), (12,16), (12,20), (12,24),
(16,4), (16,8), (16,12), (16,16), (16,20), (16,24),
(20,4), (20,8), (20,12), (20,16), (20,20), (20,24),
(24,4), (24,8), (24,12), (24,16), (24,20), (24,24)}

The sample means are {4, 6, 8, 10, 12, 14, 6, 8, 10, 12, 14, 16, 8, 10, 12, 14, 16, 18, 10, 12, 14, 16, 18, 20, 12, 14, 16, 18, 20, 22, 14, 16, 18, 20, 22, 24}

mean of sampling distribution of means

$$\mu_{\bar{x}} = \frac{4+6+8+10+12+14+6+8+10+12+14+16+8+10+12+14+16+18+10+12+14+16+18+20+12+14+16+18+20+22+14+16+18+20+22+24}{36}$$

$$= \frac{504}{36} = 14$$

(iv) The varience of sampling distribution of means is

$$\sigma_{\bar{x}}^2 = \frac{1}{36}\Big[(4-14)^2+(6-14)^2+(8-14)^2+(10-14)^2+(12-14)^2+(14-14)^2+(6-14)^2+(8-14)^2+(10-14)^2+(12-14)^2+(14-14)^2+(16-14)^2+(8-14)^2+(10-14)^2+(12-14)^2+(14-14)^2+(16-14)^2+(18-14)^2+(10-14)^2+(12-14)^2+(14-14)^2+(16-14)^2+(18-14)^2+(20-14)^2+(12-14)^2+(14-14)^2+(16-14)^2+(18-14)^2+(20-14)^2+(22-14)^2+(14-14)^2+(16-14)^2+(18-14)^2+(20-14)^2+(22-14)^2+(24-14)^2\Big]$$

$$= \frac{1}{36}\Big[(-10)^2+(-8)^2+(-6)^2+(-4)^2+(-2)^2+0+(-8)^2+(-6)^2+(-4)^2+(-2)^2+0+(2)^2+(-6)^2+(-4)^2+(-2)^2+0+(2)^2+(4)^2+(-4)^2+(-2)^2+0+(2)^2+(4)^2+(-4)^2+(-2)^2+0+(2)^2+(4)^2+(6)^2+(-2)^2+0+(2)^2+(4)^2+(6)^2+(8)^2+0$$

$+(2)^2+(4)^2+(6)^2+(8)^2+(10)^2]$

$$=\frac{1}{36}[100+64+36+16+4+0+64+36+16+4+0+4+36+16+4+0+4+16+16+4+0+4+16+36+4+0+4+16+36+64+0+4+16+36+64+100]$$

$$\sigma_{\bar{x}}^2 = \frac{840}{36} = 23.33$$

$$\sigma_{\bar{x}} = \sqrt{23.33} = 4.830$$

(v) Verify the result (sampling with replacement) :—

(1) W.K.T $\mu_{\bar{x}} = \mu = 14$

(2) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{6.83}{\sqrt{2}} = 4.830$

When the above problem asking in without replacement :-

Given population units are 4, 8, 12, 16, 20, 24

Here, population size $N=6$ and sample size $n=2$.

Write the above (1) & (2) are the same as with replacement.

(iii) Sampling without replacement

The total No. of samples sampling without replacement is N_{c_n}

$$N_{c_n} = 6_{c_2} = 15$$

15 samples of size 2.

Listing all possible samples of size 2 from population without replacement we get 15 samples.

i.e, {(4,8), (4,12), (4,16), (4,20), (4,24),
(8,12), (8,16), (8,20), (8,24)
(12,16), (12,20), (12,24),
(16,20), (16,24), (20,24)}

The samples means are {6, 8, 10, 12, 14, 10, 12, 14, 16, 14, 16, 18, 18, 20, 22}

mean of sampling distribution of means

$$\mu_{\bar{x}} = \frac{\text{Sum of all sample means}}{15}$$

$$\mu_{\bar{x}} = \frac{6+8+10+12+14+10+12+14+16+14+16+18+18+20+22}{15}$$

$$= \frac{210}{15} = 14$$

iv, variance of sampling distribution of means is

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

$$= \frac{1}{15}\Big[(6-14)^2+(8-14)^2+(10-14)^2+(12-14)^2+(14-14)^2+(10-14)^2+(12-14)^2+(14-14)^2+$$
$$(16-14)^2+(14-14)^2+(16-14)^2+(18-14)^2+(18-14)^2+(20-14)^2+(22-14)^2\Big]$$

$$= \frac{1}{15}\Big[(-8)^2+(-6)^2+(-4)^2+(-2)^2+0+(-4)^2+(-2)^2+0+(2)^2+0+(2)^2+(4)^2+(4)^2+$$
$$(6)^2+(8)^2\Big]$$

$$= \frac{1}{15}\left[64+36+16+4+0+16+4+0+4+0+4+16+16+36+64\right]$$

$$= \frac{1}{15}\left[280\right]$$

$$\sigma_{\bar{x}}^2 = 18.66$$

$$\sigma_{\bar{x}} = \sqrt{18.66} = 4.319$$

v, verify the result (sampling without replacement) :—

i, $\mu_{\bar{x}} = \mu = 14$

ii, $$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}\cdot\frac{\sigma^2}{n}} = \sqrt{\frac{6-2}{6-1}\cdot\frac{46.66}{2}}$$

$$= \sqrt{\frac{\cancel{4}\,2}{5}\cdot\frac{46.66}{\cancel{2}}}$$

$$= \sqrt{\frac{2\times 46.66}{5}}$$

$$= \sqrt{\frac{93.32}{5}}$$

$$= \sqrt{18.66}$$

$$= 4.319$$

//

4, If the population is 3, 6, 9, 15, 17

a, listing all possible samples of size 3 that can be taken without replacement from the finite population.

b, calculate the mean of each of sampling distribution of means.

c, find the S.D of sampling distribution of means.

Sol:- Given population units 3, 6, 9, 15, 17

Here population size $N = 5$ and sample size $n = 3$

The total No. of samples without replacement is $N_{C_n} = 5_{C_3} = 10$

i, listing all possible samples of size 3 from population without replacement we get 10 samples.

∴ The samples are [(3, 6, 9), (3, 6, 15), (3, 6, 17), (3, 9, 15), (3, 9, 17), (3, 15, 17), (6, 9, 15), (6, 9, 17), (6, 15, 17), (9, 15, 17)]

The sample means are {6, 8, 8.67, 9, 9.67, 11.67, 10, 10.67, 12.67, 13.67} [i.e. ex $\frac{3+6+9}{3} = 6$

ii, The mean of sampling distribution of means

$$\mu_{\bar{x}} = \frac{100.02}{10} = \frac{100}{10} = 10$$

iii, The variance of sampling distribution of means

$$\sigma^2_{\bar{x}} = \frac{1}{10}\left[(6-10)^2 + (8-10)^2 + (8.6-10)^2 + (9-10)^2 + (9.6-10)^2 + (11.6-10)^2 + (10-10)^2 + (10.6-10)^2 + (12.6-10)^2 + (13.6-10)^2\right]$$

$$= \frac{1}{10}\left[(-4)^2 + (-2)^2 + (-1.4)^2 + (-1)^2 + (-0.4)^2 + (1.6)^2 + 0 + (0.6)^2 + (2.6)^2 + (3.6)^2\right]$$

$$= \frac{1}{10}\left[16+4+1.96+1+0.16+2.56+0.36+6.76+12.96\right]$$

$$\frac{1}{10}(45.45)$$

$$= 4.54$$

$$\sigma_{\bar{x}} = \sqrt{4.54} = 2.10.$$

NOTE :- (1) $\mu = 10$
(2) $\sigma^2 = 28$

5. Find the value of finite population correction factor for $n=10$ and $N=1000$

Sol: Given size of the population $N=1000$
and size of samples, $n = 10$

$\therefore$ finite population correction factor $= \frac{N-n}{N-1}$

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

$$= \frac{990}{999}$$

$$= 0.9910$$

6. The varience of a population is 2. The size of the sample collected from the population is 169. what is the standard Error of mean

Sol: Given sample size $n = 169$
population varience $\sigma^2 = 2$

w.k.n standard Error of mean $= \frac{\sigma}{\sqrt{n}} = \frac{1.414}{\sqrt{169}} = \frac{1.414}{13}$

$$= 0.1088$$

7. A random sample of size 100 is taken from an infinite population having the mean $\mu = 76$ and varience $\sigma^2 = 256$. what is the probability that $\bar{x}$ (sample mean) will be between 75 & 78

Sol: Given sample size $n = 100$ & population mean $\mu = 76$ and

Population varience $\sigma^2 = 256$

$\sigma = 16$

We know that $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{\bar{x} - 76}{\frac{16}{\sqrt{100}}} = \dfrac{\bar{x} - 76}{\frac{16}{10}}$

$$z = \frac{\bar{x} - 76}{1.6}$$

Now $P(75 \le \bar{x} \le 78)$

when $\bar{x} = 75$, then $z = \dfrac{75 - 76}{1.6} = \dfrac{-1}{1.6} = -0.625$

when $\bar{x} = 78$, then $z = \dfrac{78 - 76}{1.6} = \dfrac{2}{1.6} = 1.25$

$P(75 \le \bar{x} \le 78) = P(-0.625 \le z \le 1.25)$

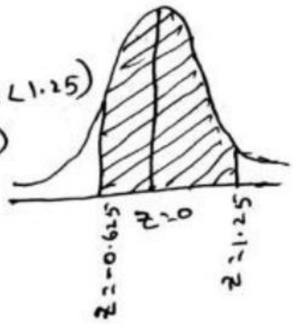
$= P(-0.625 \le z \le 0) + P(0 \le z < 1.25)$

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

$= A(0.63) + A(1.25)$

$= 0.2357 + 0.3944$

$= 0.6301$



2, Find $P(\bar{x} > 66.75)$. If a random sample of size 86 is drawn from an infinite population with mean $\mu = 63$ and S.D $\sigma = 9$

Sol: Given sample size $n = 36$

mean $\mu = 63$, S.D $\sigma = 9$

W.K.T $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{\bar{x} - 63}{\frac{9}{\sqrt{36}}} = \dfrac{\bar{x} - 63}{\frac{9}{6}} = \dfrac{\bar{x} - 63}{1.5}$

$P(\bar{x} > 66.75)$

when, $\bar{x} = 66.75$, $z = \dfrac{66.75 - 63}{1.5}$

$= \dfrac{3.75}{1.5} = 2.5$

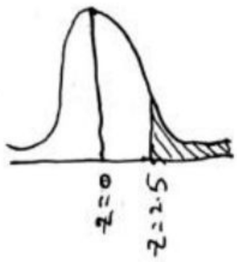
Now, $P(\bar{x} > 66.75) = P(z > 2.5)$

$= \frac{1}{2} - P(0 \leq z \leq 2.5)$

$= \frac{1}{2} - A(2.5)$

$= 0.5 - 0.4938$

$= 0.0062$



9, An electronic company manufacturer resistors that have a mean resistor that 100Ω and s.D of 10Ω. The distribution of resistance is normal. find the probability that 25 resistors will have an average resistence has less than 95Ω

Sol: Given, sample size $n = 25$

mean $\mu = 100\Omega$, S.D $\sigma = 10\Omega$

w.k.T $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x} - 100}{\frac{10}{\sqrt{25}}} = \frac{\bar{x} - 100}{\frac{10}{5}}$

$P(\bar{x} < 95)$ $\qquad z = \frac{\bar{x} - 100}{2}$

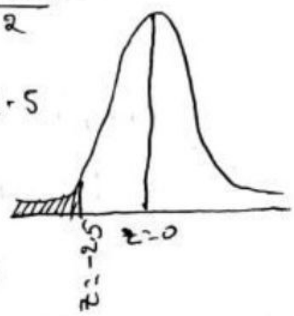
when $\bar{x} = 95$, $z = \frac{95 - 100}{2} = -\frac{5}{2} = -2.5$

$P(\bar{x} < 95) = P(z < -2.5)$

$= \frac{1}{2} - P(0 \leq z \leq -2.5)$

$= \frac{1}{2} - A(-2.5) = 0.5 - A(2.5)$

$= 0.5 - 0.4938 = 0.0062$



10, A normal population has a mean of 0.1 and s.D of 2.1. Find the probability that mean of a sample of size 900 will be negative.

Sol:- Given mean $\mu = 0.1$ and S.D $\sigma = 2.1$

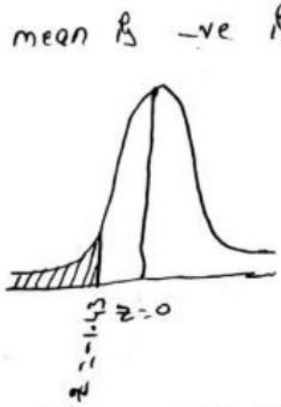
Sample size, $n = 900$

The standard normal variable $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x} - 0.1}{\frac{2.1}{\sqrt{900}}}$

$= \frac{\bar{x} - 0.1}{0.07}$

$\bar{x} = 0.1 + 0.07z$

The required probability that the sample mean is -ve is given by $P(\bar{x} < 0) = P(0.1 + 0.07z < 0)$



$= P\left(z < \frac{-0.1}{0.07}\right)$

$= P(z < -1.43)$

$= \frac{1}{2} - P(-1.43 \le z \le 0)$

$= \frac{1}{2} - A(-1.43) = \frac{1}{2} - A(1.43) = 0.5 - 0.4236$

$= 0.0764$

ii, Determine the expected no. of samples whose mean

i, lies b/w 66.8 and 68.3

ii, is less than 66.4 from a normal population of 3000 with mean 68 and S.D 3. with the size of sample is 25 from 80 samples.

Sol: Given population mean $\mu = 68$

S.D $\sigma = 3$.

Sample size $n = 25$

No. of sample from population $= 80$

population size $N = 3000$.

i, $P(66.8 < \bar{x} < 68.3)$

w.k.T $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x} - 68}{\frac{3}{\sqrt{25}}} = \frac{\bar{x} - 68}{0.6}$

when $\bar{x} = 66.8$ then $z = \frac{66.8 - 68}{0.6} = \frac{-1.20}{0.6} = -2$

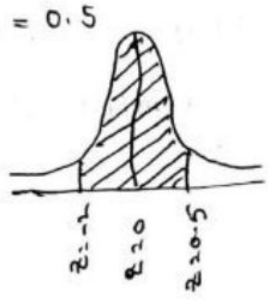
when $\bar{x} = 68.3$ then $z = \frac{68.3 - 68}{0.6} = \frac{0.3}{0.6} = 0.5$

Now, $P(-2 < z < 0.5) = A(0.5) + A(-2)$

$= A(0.5) + A(2)$

$= 0.1916 + 0.4772$

$= 0.6688$



$\therefore$ expected No. of samples whose mean lies b/w 66.8 and 68.3 is No. of samples * probability

$= 80 \times 0.6688$

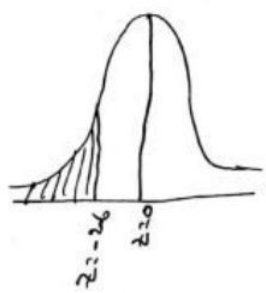
$= 53.504$

$\simeq 54$

iv) $P(\bar{x} < 66.4)$

when $\bar{x} = 66.4$ then $z = \frac{66.4-68}{0.6}$

$= -\frac{1.6}{0.6} = -2.6$



$P(z < -2.6) = \frac{1}{2} - P(-2.6 \le z \le 0)$

$= 0.5 - A(-2.6)$

$= 0.5 - A(2.6) = 0.5 - 0.4953$

$= 0.0047$

Expected No. of samples whose mean is less than 66.4 is No. of samples * probability $= 80 \times 0.0038 = 0.3040$

12) Determine the mean and S.D of the sampling distribution of means of 300 random samples each of size $n = 36$ are drawn from a population of 1500 which is normally distributed with mean $\mu = 22.4$ and S.D of 0.048, If sampling is drawn with replacement and without replacement.

Sol:- Given size of population $N = 1500$

size of sample size, $n = 36$

mean of the population $\mu = 22.4$

S.D of the population $\sigma = 0.048$.

with replacement :-

i) Given $\mu = 22.4$

W.K.T, The mean of the population = mean of the sampling distribution of means.

i.e, $\mu_{\bar{x}} = \mu = 22.4$

ii) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.048}{\sqrt{36}} = \frac{0.048}{6} = 0.0080$

Without replacement :—

Given, $\mu = 22.4$

w.k.t The mean of the population = mean of sampling distribution of means

ie, $\mu_{\bar{x}} = \mu = 22.4$

ii, $\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}}$

$= \sqrt{\frac{1500-36}{1500-1} \cdot \frac{(0.048)^2}{36}}$

3.3642

53111

$= \sqrt{\frac{1464}{1499} \cdot \frac{0.0023}{36}}$

$= \sqrt{0.0001}$

$= 0.0100$

NOTE :- 1) when a sample is taken from an infinite population what happen to the standard error of the mean if the sample size is decreased from 800 to 200.

Sol :- The standard error of the mean $SE(\bar{x}) = \frac{\sigma}{\sqrt{n}}$

sample size = n

let $n_1 = 800$ then standard error of mean

$S.E_1 = \frac{\sigma}{\sqrt{n_1}} = \frac{\sigma}{\sqrt{800}} = \frac{\sigma}{20\sqrt{2}}$

when n_1 is reduced to 200

let $n_2 = 200$ then $S.E_2 = \frac{\sigma}{\sqrt{n_2}} = \frac{\sigma}{\sqrt{200}} = \frac{\sigma}{10\sqrt{2}}$

Standard Error $S.E_2 = \frac{\sigma}{10\sqrt{2}} = 2\left(\frac{\sigma}{20\sqrt{2}}\right)$ $\left[\frac{S.E_2}{S.E_1} = 2\right]$

$= 2 \times S.E_1$

∴ If sample size is reduced from 800 to 200, s.e of mean will be multiplied by 2.

Q) A random sample is taken from $N(30, 12)$ how larger sample is should be taken if the sample mean is to lie b/w 25 & 35 with prob 0.95

A) Given $\mu = 30$

$$\sigma^2 = 12 \Rightarrow \sigma = 2\sqrt{3}$$

and also sample mean is lie b/w 25 & 35

i.e, $P(25 \le \bar{x} \le 35) = 0.95$

W.K.T $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{\bar{x} - 30}{\sqrt{\frac{12}{n}}}$

when $\bar{x} = 25$, then $z = \dfrac{25 - 30}{\sqrt{\frac{12}{n}}} = \dfrac{-5}{\sqrt{\frac{12}{n}}} = \dfrac{-5\sqrt{n}}{\sqrt{12}} = -z_1$ (say)

$\bar{x} = 35$, then $z = \dfrac{35 - 30}{\sqrt{\frac{12}{n}}} = \dfrac{5}{\sqrt{\frac{12}{n}}} = \dfrac{5\sqrt{n}}{\sqrt{12}} = +z_1$ (say)

Now, $p(25 \le \bar{x} \le 35) = 0.95$

$\Rightarrow p(-z_1 \le z \le z_1) = 0.95$

$\Rightarrow p(-z_1 \le z \le 0) + p(0 \le z \le z_1) = 0.95$

$\Rightarrow p(0 \le z \le z_1) + p(0 \le z \le z_1) = 0.95$

$\Rightarrow 2\,p(0 \le z \le z_1) = 0.95$

$\Rightarrow p(0 \le z < z_1) = 0.4750$

when area 0.4750, then $z_1 = 1.96$

$$\frac{5\sqrt{n}}{\sqrt{12}} = 1.96$$

$$\sqrt{n} = \frac{1.96}{5} \times \sqrt{12}$$

$$\sqrt{n} = 1.3579$$

$$n = 1.8 \simeq 2$$

Sampling distribution of proportions :— Let 'p' be the probability of sucess and Q is the probability of failure.

case - i : i) For finite population :—

The mean of sampling distribution of proportions is $\mu_p = p$

ii :- The standard deviation of proportions $\sigma_p = \sqrt{\frac{N-n}{N-1} \cdot \frac{pq}{n}}$

ii) **For infinite population :—**

(i) The mean of sampling distribution of proportions is $\mu_p = p$

(ii) The standard deviation of sampling distribution of propositions,

$$\sigma_p = \sqrt{\frac{pq}{n}}$$

Ex:-1 Find the probability that in 120 tosses of a fair coin

i) b/w 40% & 60% will be heads.

ii) 5/8 more will be heads.

Sol:- Here the coins are fair $p = 1/2$ and

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

Here $n = 120$

$\left[\because z = \frac{x-\mu}{\sigma} \right]$

W.K.T the mean of sampling distribution of proportions $\mu_p = P$

$\mu_p = 1/2$

and S.D of sampling distribution of proportions

$$\sigma_p = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{1/2 \cdot 1/2}{120}} = \sqrt{\frac{1}{480}} = 0.0456$$

W.K.T Standard binomial distribution $z = \frac{P - \mu_p}{\sigma_p}$

$$z = \frac{P - 1/2}{0.0456}$$

i, when 40% S.U then $z = \frac{0.4 - 1/2}{0.0456} = \frac{-0.10}{0.0456} = -2.19$

when 60% S.U then $z = \frac{0.6 - 0.5}{0.0456} = \frac{0.1}{0.0456} = 2.19$

$\therefore$ The required probability, $P(-2.19 < z < 2.19)$

$= P(0 < z < 2.19) + P(-2.19 < z < 0)$

$= A(2.19) + A(-2.19)$

$= 0.4857 + 0.4857$

$= 0.9714$

ii, when 5/8 S.U then $z = \frac{\frac{5}{8} - \frac{1}{2}}{0.0456} = \frac{0.62 - 0.5}{0.0456} = \frac{0.12}{0.0456}$

$= 2.74$

$\therefore$ The required probability, $P(z \geq 2.74)$

$= 0.5 - P(0 < z < 2.74)$

$= 0.5 - A(2.74)$

$= 0.5 - 0.4969$

$= 0.0031$ //

Sampling distribution of differences and sums :—

Sums :— The mean of sampling distribution of sums of means are given by $\mu_{\bar{x}_1 + \bar{x}_2} = \mu_1 + \mu_2$

2, The standard deviation of S.D of means are given by

$$\sigma_{\bar{x}_1+\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \qquad \left[\because \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \text{ or } \sqrt{\frac{\sigma^2}{n}}\right].$$

Differences :—

1, The mean of sampling distribution of differences of means are given by $\mu_{\bar{x}_1-\bar{x}_2} = \mu_1-\mu_2$.

2, The standard deviation of sampling distribution of differences of means are given by $\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

NOTE :- 1, The mean of sampling distribution of differences of proportions are given by $\mu_{P_1-P_2} = P_1-P_2$.

2, The S.D of sampling distribution of differences of proportions are given by $\sigma_{P_1-P_2} = \sqrt{\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{n_2}}$

Ex 1:- Let $u_1=\{2,7,9\}$ and $u_2=\{3,8\}$ find 1, μ_{u_1} 2, μ_{u_2} 3, σ_{u_1} 4, σ_{u_2} 5, $\mu_{u_1-u_2}$ 6, $\sigma_{u_1-u_2}$ 7, verify the result.

i.e, $\mu_{u_1-u_2} = \mu_{u_1}-\mu_{u_2}$

$\sigma_{u_1-u_2} = \sqrt{\sigma_{u_1}^2+\sigma_{u_2}^2}$

Sol:-

Given $u_1=\{2,7,9\}$ and $u_2=\{3,8\}$

1, $\mu_{u_1} = \frac{2+7+9}{3} = 6$

2, $\mu_{u_2} = \frac{3+8}{2} = \frac{11}{2} = 5.5$

3, $\sigma_{u_1} = \sqrt{\frac{1}{3}\left[(2-6)^2+(7-6)^2+(9-6)^2\right]}$

$= \sqrt{\frac{1}{3}\left[(-4)^2+(1)^2+(3)^2\right]}$

$= \sqrt{\frac{1}{3}(16+1+9)} = \sqrt{\frac{26}{3}} = \sqrt{8.66} = 2.94.$

(4) $\sigma_{U_2} = \sqrt{\frac{1}{2}\left[(3-5.5)^2 + (8-5.5)^2\right]}$

$= \sqrt{\frac{1}{2}\left\{(-2.5)^2 + (2.5)^2\right]}$

$= \sqrt{\frac{1}{2}\left\{(6.25) + (6.25)\right\}}$

$= \sqrt{\frac{12.5}{2}} = 2.5$

(5) $\mu_{U_1-U_2}$
population consisting of the difference of any member of U_1 and any member of U_2 is

$2-3,\ 2-8,\ 7-3,\ 7-8,\ 9-3,\ 9-8.$

i.e, $U_1 - U_2 = \{-1, -6, 4, -1, 6, 1\}$

$\mu_{U_1-U_2} = \frac{-1+(-6)+4-1+6+1}{6}$

$= \frac{3}{6} = 0.5$

(6) $\sigma_{U_1-U_2} = \sqrt{\frac{1}{6}\left\{(-1-0.5)^2 + (-6-0.5)^2 + (4-0.5)^2 + (-1-0.5)^2 + (6-0.5)^2 + (1-0.5)^2\right.}$

$= \sqrt{\frac{1}{6}\left\{(-1.5)^2 + (-6.5)^2 + (3.5)^2 + (-1.5)^2 + (5.5)^2 + (0.5)^2\right.}$

$= \sqrt{\frac{1}{6}\left[2.25 + 42.25 + 12.25 + 2.25 + 30.25 + 0.25\right.}$

$= \sqrt{\frac{89.5}{6}} = \sqrt{14.91} = 3.86$

(7) verify the result

(i) $\mu_{U_1-U_2} = \mu_{U_1} - \mu_{U_2}$

$= 6 - 5.5 = 0.5$

(ii) $\sigma_{U_1-U_2} = \sqrt{\sigma_{U_1}^2 + \sigma_{U_2}^2} = \sqrt{(2.94)^2 + (2.5)^2}$

$= \sqrt{8.64 + 6.25}$

$= \sqrt{14.89} = 3.85$ //

2, let $U_1 = \{3, 7, 8\}$ and $U_2 = \{2, 4\}$ find

(1) μ_{U_1} (2) μ_{U_2} (3) σ_{U_1} (4) σ_{U_2}

(5) The mean of sampling distribution of sums of means.

(6) The S.D of sampling distribution of sums of means.

(7) verify the result.

Sol:- Given $U_1 = \{3, 7, 8\}$ and $U_2 = \{2, 4\}$

i) $\mu_{U_1} = \frac{3+7+8}{3} = \frac{18}{3} = 6$ $\quad [\because \mu = \frac{x_1+x_2+\cdots+x_n}{n}$

ii) $\mu_{U_2} = \frac{2+4}{2} = \frac{6}{2} = 3$

iii) $\sigma_{U_1} = \sqrt{\frac{1}{3}\left[(3-6)^2+(7-6)^2+(8-6)^2\right]}$ $\quad [\because \sigma^2 = \frac{1}{n}\Sigma(x_i-\mu)^2$

$= \sqrt{\frac{1}{3}[9+1+4]}$

$= \sqrt{\frac{1}{3}\times 14} = 2.160$

iv) $\sigma_{U_2} = \sqrt{\frac{1}{2}\left[(2-3)^2+(4-3)^2\right]}$

$= \sqrt{\frac{1}{2}[1+1]} = 1$

v) $\mu_{U_1+U_2}$

$U_1+U_2 = \{3+2, 3+4, 7+2, 7+4, 8+2, 8+4\}$

$= \{5, 7, 9, 11, 10, 12\}$

$\mu_{U_1+U_2} = \frac{5+7+9+11+10+12}{6}$

$= 9$

vi) $\sigma_{U_1+U_2} = \sqrt{\frac{1}{6}\left[(5-9)^2+(7-9)^2+(9-9)^2+(11-9)^2+(10-9)^2+(12-9)^2\right]}$

$= \sqrt{\frac{1}{6}[16+4+0+4+1+9]}$

$$=\sqrt{\frac{1}{6}(34)} = 2.380$$

vii, verify the result.

$$\mu_{v_1+v_2} = \mu_{v_1} + \mu_{v_2} = 6+3 = 9$$

$$\sigma_{v_1+v_2} = \sqrt{\sigma_{v_1}^2 - \sigma_{v_2}^2} = \sqrt{(2.160)^2 + (1)^2} = \sqrt{4.665+1}$$

$$= \sqrt{5.665}$$

$$= 2.380$$

3, The mean life time of electric bulbs produced by company is 1500 hrs and S.D of 150 hrs. Find the probability that electric bulb will take place for a, atleast 5000 hrs

b, atmost 4200 hrs, if 3 bulbs are connected such that when 1 bulb burns out, another bulb will go on.

Assume that life time is normal distributed.

Sol:- Let the mean lifetime of electric bulbs L_1, L_2, L_3 be

$$\mu_{L_1}, \mu_{L_2}, \mu_{L_3}$$

The mean of these 3 electric bulbs is $\mu_{L_1+L_2+L_3}$

$$\therefore \mu_{L_1+L_2+L_3} = \mu_{L_1} + \mu_{L_2} + \mu_{L_3}.$$

$$= 1500 + 1500 + 1500$$

$$= 4500 \text{ hrs.}$$

$$\sigma_{L_1+L_2+L_3} = \sqrt{\sigma_{L_1}^2 + \sigma_{L_2}^2 + \sigma_{L_3}^2}$$

$$= \sqrt{(150)^2 + (150)^2 + (150)^2}$$

$$= \sqrt{67500}$$

$$= 259.80$$

$\therefore$ let x be the 3 bulbs are connected.

$$z = \frac{x-\mu}{\sigma} = \frac{x-4500}{259.80}$$

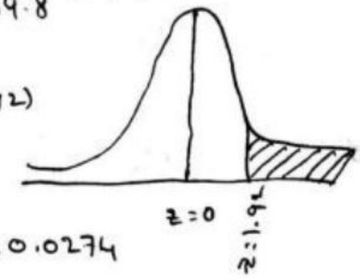
i) P(atleast 5000) = $P(x \geq 5000)$

when $x = 5000$ then $z = \frac{5000-4500}{259.8} = \frac{500}{259.8} = 1.92$.

$P(x \geq 5000) = P(z \geq 1.92) = \frac{1}{2} - P(0 \leq z \leq 1.92)$

$= 0.5 - A(1.92)$

$= 0.5 - 0.4726 = 0.0274$



ii) P(atmost 4200) = $P(x \leq 4200)$

when $x = 4200$ then $z = \frac{4200-4500}{259.8} = \frac{-300}{259.8} = -1.1547 = -1.15$

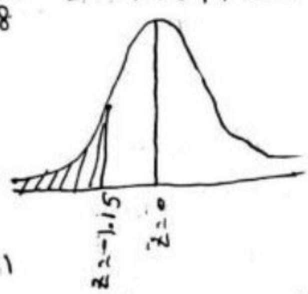
$P(x \leq 4200) = P(z \leq -1.15)$

$= \frac{1}{2} - P(-1.15 \leq z \leq 0)$

$= 0.5 - A(-1.15)$

$= 0.5 - A(1.15)$

$= 0.5 - 0.3749 = 0.1251$



4. The mean voltage of a battery is 15v and standard deviation is 0.2v what is the probability that 4 such batteries connected in series will have a combine voltage of 60.8 or more volts.

Sol: let mean voltage of batteries 1, 2, 3, 4 be v_1, v_2, v_3, v_4

∴ The mean of the series of the 4 batteries are connected

is $\mu_{v_1+v_2+v_3+v_4} = \mu_{v_1} + \mu_{v_2} + \mu_{v_3} + \mu_{v_4}$

$= 15 + 15 + 15 + 15$

$= 60v$

and $\sigma_{v_1+v_2+v_3+v_4} = \sqrt{\sigma_{v_1}^2 + \sigma_{v_2}^2 + \sigma_{v_3}^2 + \sigma_{v_4}^2}$

$= \sqrt{(0.2)^2 + (0.2)^2 + (0.2)^2 + (0.2)^2}$

$= \sqrt{0.16}$

$= 0.4$

let x be the combined voltage of the series.

$$\therefore Z = \frac{x-\mu}{\sigma} = \frac{x-60}{0.4}$$

when $x = 60.8$ then $Z = \frac{60.8-60}{0.4} = \frac{0.8}{0.4} = 2.$

i, $P(60.8$ or more volts$) = P(x \geq 60.8)$

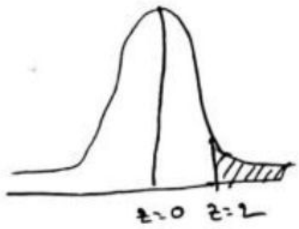
$= P(Z \geq 2)$

$= \frac{1}{2} - P(0 \leq Z \leq 2)$

$= \frac{1}{2} - A(2)$

$= 0.5 - 0.4772$

$= 0.0228$



5, find the mean and S.D of sampling distribution of variences for the population 2, 3, 4, 5 for drawing sampling size 2.

i, without replacement ii, with replacement.

Sol:

Given population units 2, 3, 4, 5, Here $N=4$, $n=2$

The total No. of samples without replacement is $N_{C_n} = {}^4C_2 = 6$

The samples are $(2,3), (2,4), (2,5), (3,4), (3,5), (4,5)$ $\left[\because \frac{2+3}{2} = \frac{5}{2} = 2.5\right.$

The sample means are $\{2.5, 3, 3.5, 3.5, 4, 4.5\}$

calculating the sample variance for each of the 6 samples are

$\frac{(2-2.5)^2+(3-2.5)^2}{2} = 0.25$

$\frac{(2-3)^2+(4-3)^2}{2} = 1$

$\frac{(3.5-2)^2+(3.5-5)^2}{2} = 2.25$

$\frac{(3-3.5)^2+(4-3.5)^2}{2} = 0.25$

$\frac{(3-4)^2+(5-4)^2}{2} = 1$

$\frac{(4-4.5)^2+(5-4.5)^2}{2} = 0.25.$

The variences of each of the 6 samples are,

0.25, 1, 2.25, 0.25, 1, 0.25

I, mean of the sampling distribution of variences $\mu_{s^2} = \frac{0.25+1+2.25+0.25+1+0.25}{6}$

$$= \frac{5}{6} = 0.8$$

II, The S.D of sampling distribution of variences

$$\sigma_{s^2} = \sqrt{\frac{1}{6}\{(0.25-0.8)^2+(1-0.8)^2+(2.25-0.8)^2+(0.25-0.8)^2+(1-0.8)^2+(0.25-0.8)^2\}}$$

$$= \sqrt{\frac{1}{6}\{0.302+0.040+2.102+0.302+0.040+0.302\}}$$

$$= \sqrt{0.5147} = 0.7174.$$

H.W

with replacement :-

Test of Hypothesis (Large Sample)

Hypothesis (or) Satistical Hypothesis :—

A statistical hypotheses is a statement about the parameter of one or more populations.

Note :- statement = Hypothesis

Test of hypothesis :-

i.e The procedure which enables us to decide on the basis of sample results whether a hypothesis is true or not, is called Test of hypothesis (or) Test of significance

There are two types of hypothesis. They are

i) Null hypothesis (H_0)

ii) Alternative hypothesis (H_1)

(i) Null hypothesis :-

Null hypothesis is denoted by H_0. It is a definite statement which asserts that there is no significant difference between statistic and the population parameter

so null hypothesis $H_0 : \mu = \mu_0$

ii) Alternative hypothesis :-

Alternative hypothesis is denoted by H_1. It is always opposite (or) complementary to null hypothesis.

i.e If we want to test the null hypothesis that the population has a specified mean μ_0 ie $H_0 : \mu = \mu_0$ then the alternative hypothesis could be

1) $H_1: \mu \neq \mu_0$ It is called two tailed alternative hypothesis
2) $H_1: \mu > \mu_0$ It is called right-tailed alternative hypothesis
3) $H_1: \mu < \mu_0$ It is called left-tailed alternative hypothesis

Errors in sampling :-

To accept (or) reject the hypothesis gives the some errors (or) risks

There are two types of errors (or) risks in the testing of hypothesis. They are

1) Type - I error
2) Type - II error

1) Type - I error :- Reject H_0 (null hypothesis) when it is true, ie rejecting a correct hypothesis. The probability of type - I error is denoted by α

ie P(type - I error) = P(reject H_0 when it is true) = α

where α is the size of the type - I error and is reffered as producers risk (error)

2) Type - II error :-

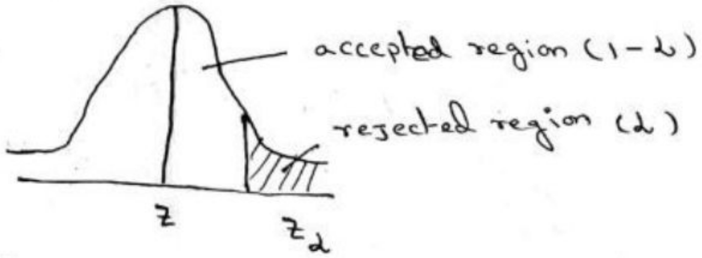
Accept H_0 when it is false ie Accepting a wrong hypothesis. The probability of type - II error is denoted by "β"

i.e P(type - II error) = P(Accept H_0 when it is false) = β

where β is the size of the type - II error and is reffered as consumer's risk.

Level of significance :-

Level of Significance of a test denoted by 'α' is the probability of Commiting Type-I error. Thus LOS measures the amount of risk or error associated in taking decissions. We take generally α as 0.05 (ie 5%) (or) 0.01 (ie 1%). Thus α = 5% means there are 5 chances in 100 that Null hypothesis is rejected when it is true (or) 95% confident that a right decission is made.



Critical region :-

Critical region is the region of rejection of null Hypotheses. The area of critical region equals to the level of Significance 'α'. Note that critical region always lies on the tail(s) of distribution. Depending nature of Alternate Hypotheses Critical region may lie on the one side (or) both side of the tails.

Critical values (or) significant values :-

The value of the test statistic which separates the critical region and the acceptance region is called the critical value (or) significant value and is denoted by "z_α". This value is dependent on

(i) the level of significance used and
(ii) the alternative hypothesis, whether it is one-tailed (or) two-tailed.

One tailed and Two tailed tests :-

One tailed tests :-

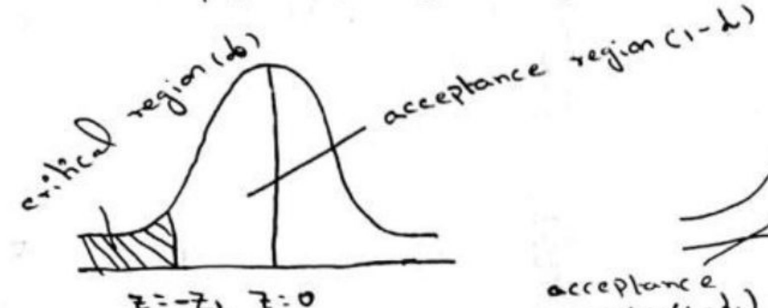
A test of any statistical hypothesis where the alternative hypothesis (H_1) is expressed by the the symbols $<$ (or) $>$ is called as one-tailed test.

Ex :- $H_0 : \mu = \mu_0$

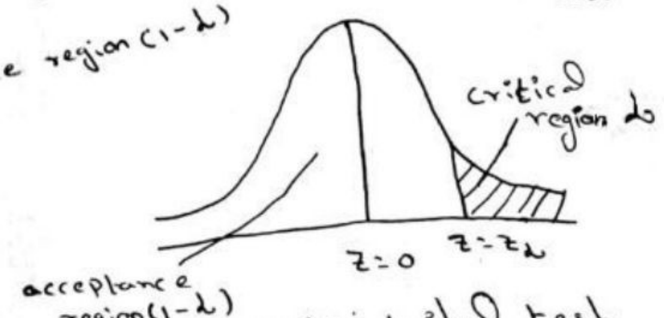
$H_1 : \mu < \mu_0$ (left tailed)

(or)

$H_1 : \mu > \mu_0$ (Right tailed) is called one-tailed test.



Left tailed test



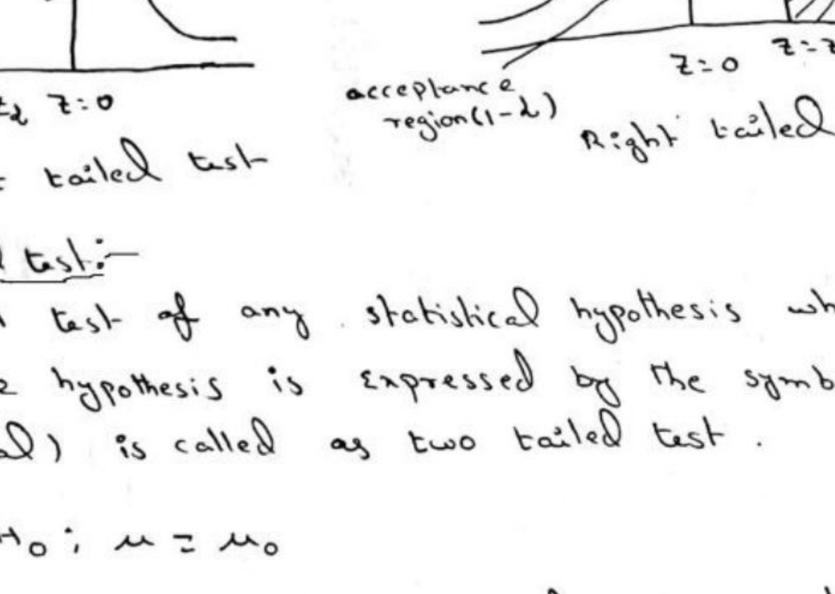
Right tailed test

Two tailed test :-

A test of any statistical hypothesis where the alternative hypothesis is expressed by the symbol "$\neq$" (not equal) is called as two tailed test.

Ex :- $H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$ is called as two-tailed test.



Large and small samples :- If the size of the sample (n) is ≥ 30 then the sample is called large sample otherwise it is called small sample

ie. sample size $n < 30$ then the sample is small sample

Simple and composite Hypothesis :-

If the parameters are completely specified in the hypothesis then the hypothesis is called "simple hypothesis" otherwise it is called "composite hypothesis"

Table values of z at the level of significance :-

critical values (z_{tab})	Level of significance (α)		
	1% (0.01) (99% confidence)	5% (0.05) (95% confidence)	10% (0.10) (90% confidence)
Two tailed $\lvert z_{\alpha/2}\rvert$	$\lvert z_{\alpha/2}\rvert = 2.58$	1.96	1.645
Right tailed z_α	$z_\alpha = 2.33$	1.645	1.28
left tailed z_α	$z_\alpha = -2.33$	-1.645	-1.28

Procedure for testing of hypothesis :-

various steps in testing of hypothesis are given below.

1. Null hypothesis :- (H_0)

set up a null hypothesis H_0 in clear terms from the given data

2. Alternative hypothesis (H_1) :-

set up alternative hypothesis H_1 which is based on null hypothesis

so that we could decide whether we shall use one tailed (right or left) test (or) two tailed test we use

$H_1 : \mu \neq \mu_0$ for two tailed test

$H_1 : \mu > \mu_0$ for right tailed test

$H_1 : \mu < \mu_0$ for left tailed test

3. Level of significance :-

choose the appropriate level of significance (α)
Here we have table value of z ie z_{tab}

If 'α' is not given in problem, then we choose 5% level of significance.

4. Test statistic :-

compute the test statistic $z = \dfrac{t - E(t)}{S.E(t)}$

under the null hypothesis H_0

where 't' sample statistic and S.E is standard error of t.

5. conclusion :-

Here we give the conclusion of the given problem by the comparison of statistical value z_{cal} and tablated value z_{tab}

(i) If $|z_{cal}| < |z_{tab}|$ ie if the calculated value of z is less than tabulated value of z at the level of significance then we accept the null hypothesis H_0.

(ii) If $|z_{cal}| > |z_{tab}|$ i.e if the calculated value of z is greater than tabulated value of z at the Level of significance then we reject the null hypothesis H_0 (or) accept the alternative hypothesis H_1.

Test of significance for large samples :-

under large samples we see four important tests of significance. They are

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

1. Testing of significance for single mean :-

If the population variance "σ^2" is known then to test the hypothesis we use the test statistic

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

where

$\bar{x}$ is sample mean

μ is population mean

σ is population S.D

n is sample size

Here $\sigma/\sqrt{n}$ is the standard Error of sample mean $\bar{x}$

If the population variance "σ^2" is unknown then to test the hypothesis we use the test statistic

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

where "s" is sample standard deviation

Note :- The confidence limits (fiducial limits) for single mean at the confidence are $\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}\right)$ where $z_{\alpha/2}$ is the value of Z at the % of confidence

P) A sample of 400 items is taken from a population whose standard deviation 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 for the population

Sol:- Given

sample size $n = 400$

mean of sample $\bar{x} = 40$

mean of population $\mu = 38$

standard deviation of population $\sigma = 10$

1. Null hypothesis: $H_0 : \mu = 38$

2. Alternative hypothesis: $H_1 : \mu \neq 38$ (two tailed test)

3. Level of significance:

given Level of significance $\alpha = 5\%$ (confidence 95%)

The table value of z at level of significance

$\alpha = 5\%$ is $z_{tab} = 1.96 \Rightarrow |z_{tab}| = 1.96$

4. Test statistic:

The test statistic is

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

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

$$z_{cal} = 4$$

5. Conclusion:

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

i.e calculated value of z is greater than the tabulated value of z at level of significance $\alpha = 5\%$.

Hence we reject null hypothesis H_0

$\therefore$ The sample is not from the population whose mean is 38

<u>confidence limits</u> :-

95% confidence value of z is $z_{\alpha/2} = 1.96$

The confidence interval of the population mean "μ" is

$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

$$\left(40 - 1.96\frac{10}{\sqrt{400}},\ 40 + 1.96\frac{10}{\sqrt{400}}\right)$$

$$(39.02,\ 40.98)$$

5M

(2) The mean and standard deviation of a populations are 225 and 278 respectively. what can we assert with 95% confidence about maximum error if $\bar{x} = 225$ and $n = 100$? ⟨2015⟩

<u>Sol</u> Given $\mu = 225$, $\sigma = 278$, $\bar{x} = 225$; $n = 100$.

Confidence limit = 95%.

$Z_{\alpha/2}$ for 95% Confidence = 1.96.

$\therefore$ Maximum Error $= Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = \frac{1.96 \times 278}{10} = 54.488$

95% Confidence interval for μ is $\left(\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$

$$(225 - 54.488,\ 225 + 54.488)$$

$$(170.512,\ 279.488)$$

③ A Sample of 64 students has a mean weight of 70 Kgms. Can this be regarded as a sample from a population with mean weight 56 Kgms and standard deviation 25 Kgms with level of significance 0.05?

Sol Given $\bar{x}$ = Sample mean = 70, Sample Size = n = 64.

population mean μ = 56, population S.D = 25.

Level of Significan $\alpha = 0.05 = 5\%$.

(i) Null Hypothesis H_0: $\mu = 56$ (i.e a sample from a population with Mean weight $\mu = 56$)

(ii) Alternative Hypothesis H_1: $\mu \neq 56$. ⟨two tailed Test⟩

(iii) Level of significance α : $\alpha = 0.05 = 5\%$.

(iv) Test Statistic : z_{cal} (or) $z = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{70-56}{\frac{25}{\sqrt{64}}} = 4.48$.

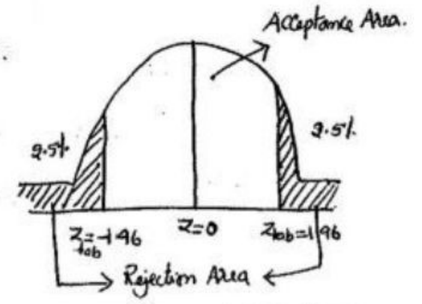
$\therefore |z_{cal}| = 4.48$.

(v) Conclusion : Say the Table

$z_{tab} = 1.96 \Rightarrow |z_{tab}| = 1.96$.

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

$\therefore$ H_0 is rejected.



④ The daily out put of a chemical plant has an average of 590 tons. To know whether this average has changed in recent months, a random selection of 100 days has been made. The average and standard deviation are $\bar{x} = 585$ tons and $s = 16.5$ tons. Test the appropriate hypothesis using $\alpha = 0.05$.

Sol Given $\mu = 590$, $n = 100$, $\bar{x} = 585$, $s = 16.5$

(1) Null Hypothesis $H_0: \mu = 590$

(2) Alternative Hypothesis $H_1: \mu \neq 590$ ⟨two tailed Test⟩

(3) Level of Significance α: $\alpha = 0.05 = 5\%$.

(4) Test Statistic : Z_{cal} or $Z = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ (∵ σ is not given

$$\therefore Z = \frac{585 - 590}{\frac{16.5}{10}} = -3.03.$$

$$\therefore |Z_{cal}| = 3.03.$$

(5) Conclusion : By Table $Z_{tab} = 1.96 \Rightarrow |Z_{tab}| = 1.96$

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

$\therefore$ H_0 is rejected

ie The average has changed in recent months.

(5) In a random sample of 60 workers, the average time taken by the to get to work 33.8 minutes with standard deviation of 6.1 minutes. Can we reject the null Hypotheses $\mu = 32.6$ minutes in favour of alternative null Hypotheses $\mu > 32.6$ at $\alpha = 0.05$.

Sol Given $n = 60$, $\bar{x} = 33.8$, $\mu = 32.6$ and $s = 6.1$

(1) Null Hypotheses $H_0: \mu = 32.6$

(2) Alternative Hypothesis $H_1: \mu > 32.6$

(3) Level of Significance : $\alpha = 0.05 = 5\%$.

(4) Test statistic : $Z = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{33.8 - 32.6}{\frac{6.1}{\sqrt{60}}} = 1.5238$

$$\therefore |Z_{cal}| = 1.5338.$$

(5) Conclusion : By Table $Z_{tab} = 1.645 \Rightarrow |Z_{tab}| = 1.645$

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

$\therefore$ H_0 is accepted.

(6) A Sample of 900 members has mean of 3.4 Cms and S.D of 2.61 Cms. Is the Sample from a large population of mean 3.25 Cm and S.D 2.61 Cms. and find 95% Confidence limits of true mean.

Sol Given $n = 900$, $\bar{x} = 3.4$, $S = 2.61$, $\mu = 3.25$ and $\sigma = 2.61$.

(1) Null Hypothesis $H_0 : \mu = 3.25$ (ie The sample has been drawn from the population mean $\mu = 3.25$)

(2) Alternative Hypothesis $H_1 : \mu \neq 3.25$ <two tailed Test>

(3) Level of Significance : $\alpha = 5\%$.

(4) Test Statistic : $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{3.4 - 3.25}{\frac{2.61}{30}} = 1.724$.

$\therefore |z_{cal}| = 1.724$

(5) Conclusion : By the Table $z_{tab} = 1.96 \Rightarrow |z_{tab}| = 1.96$

$\therefore |z_{cal}| < |z_{tab}|$

$\therefore H_0$ is accepted

i.e The Sample has been drawn from the population mean $\mu = 3.25$.

Confidence limit $= 95\%$.

$z_{\alpha/2}$ for 95% Confidence $= 1.96$

95% Confidence Interval for mean μ is given by

$$\left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \ , \ \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

$$\left(3.4 - \frac{1.96 \times 2.61}{30} \ , \ 3.4 + \frac{1.96 \times 2.61}{30} \right)$$

$$(3.4 - 0.1705 \ , \ 3.4 + 0.1705)$$

$$(3.2295 \ , \ 3.57)$$ //

Ke

(7) A Tracking firm Suspects the claim that the average life time of certain types at least 28,000 miles. To check the claim the firm puts 40 of these tyres on its trucks and gets a mean life of 27,463 miles with S.D of 1,348 miles. what we conclude if the probability of Type-I error is to be at most 0.01?

Sol Given $n = 40$, $\bar{x} = 27,463$, $\mu = 28,000$, $s = 1.348$, $\alpha = 0.01 = 1\%$.

(1) Null Hypothesis $H_0 : \mu = 28000$

(2) Alternative Hypothesis $H_1 : \mu \neq 28000$ < Two Tailed Test >

(3) Level of Significance : $\alpha = 0.01 = 1\%$.

(4) Test statistic : $z = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{27463 - 28000}{1.348/\sqrt{40}} = -2.52$

$\therefore |z_{cal}| = 2.52$

(5) Conclusion : By Table $z_{tab} = 2.58 \Rightarrow |z_{tab}| = 2.58$

$\therefore |z_{cal}| < |z_{tab}|$

$\therefore$ H_0 is accepted

K

(8) The mean life time of sample of 100 bulbs produced by a Company is Computed as 1570 h with a standard deviation of 120h. If μ is the mean life time of all bulbs produced by the Company, Test the hypothesis $\mu = 1600$ h against AH $\mu \neq 1600$ h using a LOS of as (a) 0.05 (b) 0.01 Also (c) find p-value of test.

Sol Given $n = 100$, $\bar{x} = 1570$, $s = 120$, $\mu = 1600$

(a)

(1) Null Hypothesis $H_0 : \mu = 1600$

(2) Alternative Hypothesis $H_1 : \mu \neq 1600$

(3) Level of Significance : $\alpha = 0.05 = 5\%$.

(4) Test statistic : $z = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

$\therefore z = \dfrac{1570 - 1600}{120/10} = -2.50$

$\therefore |z_{cal}| = 2.50$

(5) Conclusion : By Table

$z_{tab} = 1.96 \Rightarrow |z_{tab}| = 1.96$

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

$\therefore$ H_0 is rejected

(b)

(1) Null Hypothesis $H_0 : \mu = 1600$

(2) Alternative Hypothesis $H_1 : \mu \neq 1600$

(3) Level of Significance : $\alpha = 0.01 = 1\%$.

(4) Test Statistic : $z = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

$\therefore z = \dfrac{1570 - 1600}{120/10} = -2.50$

$\therefore |z_{cal}| = 2.50$

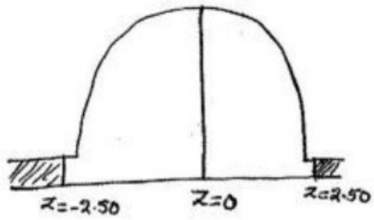
(5) Conclusion: By Table

$z_{tab} = 2.58 \Rightarrow |z_{tab}| = 2.58$

$|z_{cal}| < |z_{tab}|$

$\therefore$ H_0 is accepted.

(c) P- Value of the Test is given by



$$P(z \le -2.50) + P(z \ge 2.50)$$
$$= 0.5 - A(-2.50) + 0.5 - A(2.50)$$
$$= 0.5 - A(2.50) + 0.5 - A(2.50)$$
$$= 1 - 2A(2.50)$$ < See z-Table 2.5 at 0
$$= 1 - 2(0.4938) = 1 - 0.9876$$
$$= 0.0124 \,//$$

Ke
⑨ The breaking strength of cabels produced by a manufacturer has mean 1800 lb and standard deviation 100lb by a new technique. It is claimed that The breaking strength can be increased. To test this claim, a sample of 50 cabels is tested and it is found that the mean breaking strength is 1850 lb.

(a) Can we support the claim at 0.01 LOS?
(b) What is the p-Value of Test?

Sol Given $\mu = 1800$, $\sigma = 100$, $n = 50$, $\bar{x} = 1850$

(1) Null Hypothesis $H_0: \mu = 1800$

(2) Alternative Hypothesis $H_1: \mu > 1800$ < Right Tailed Test >

(3) LOS : $\alpha = 0.01 = 1\%$

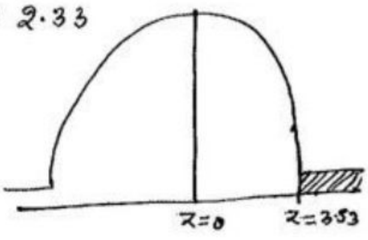
(4) Test statistic : $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{1850 - 1800}{\frac{100}{\sqrt{50}}} = \dfrac{\cancel{50}}{\cancel{100}_2} \times \sqrt{50} = 3.53$

$$|z_{cal}| = 3.53$$

(5) Conclusion : $z_{tab} = 2.33 \Rightarrow |z_{tab}| = 2.33$

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

$\therefore$ H_0 is rejected.



(b) P- Value of the Test is given by
$$P(z \ge 3.53)$$
$$= 0.5 - A(3.53) = 0.5 - 0.4998$$
$$= 0.0002$$

(10) It is claimed that a random sample of 49 Tyres has mean life of 15200 Km. This sample was drawn from a population whose mean is 15150 Kms and standard deviation of 1200 Km. Test the significance at 0.05 level.

Sol Given $n=49$, $\bar{x}=15200$, $\mu=15150$, $\sigma=1200$, $\alpha=5\%$.

(1) Null Hypothesis $H_0: \mu=15150$,

(2) Alternative Hypothesis $H_1: \mu \neq 15150$,

(3) Level of Significance $\alpha=5\%$,

(4) Test Statistic : $z=\dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}=0.2917$

$$|z_{cal}|=0.2917$$

(5) Conclusion : $z_{tab}=1.96 \Rightarrow |z_{tab}|=1.96$.

$$\therefore |z_{cal}|<|z_{tab}|,$$

H_0 is Accepted.

(11) An Ambulence service claims that it takes on avarage less than 10 minutes to reach its destination in emergency Calls. A sample of 36 calls has a mean of 11 minutes and Variance of 16 minutes. Test the significance at 0.05 level.

Sol Given $n=36$, $\bar{x}=11$, $\mu=10$, $s^2=16 \Rightarrow s=4$, $\alpha=5\%$

(1) Null Hypothesis $H_0: \mu=10$,

(2) Alternative Hypothesis $H_1: \mu<10$,

(3) Level of Significance $\alpha=5\%$,

(4) Test Statistic : $z=\dfrac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}=1.5$,

$$z_{cal}=1.5,$$

(5) Conclusion : $z_{tab}=-1.645 \Rightarrow |z_{tab}|=1.645$

$$\therefore |z_{cal}|<|z_{tab}|,$$

$\therefore$ H_0 is accepted.

Test of Hypothesis for Two means < Large Samples >

Given $\mu_1 - \mu_2 = 0$ (populations are same means)	Given $\mu_1 - \mu_2 = \delta$ (populations have different means)
$H_0 : \mu_1 - \mu_2$	$H_0 : \mu_1 - \mu_2 = \delta$
$H_1 : \mu_1 \neq \mu_2$ (or) $\mu_1 < \mu_2$ (or) $\mu_1 > \mu_2$	$H_1 : \mu_1 - \mu_2 \neq \delta$ (or) $\mu_1 - \mu_2 < \delta$ (or) $\mu_1 - \mu_2 > \delta$
Test Statistic	Test Statistic
$z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ if σ_1, σ_2 known	$z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ if σ_1, σ_2 known
$z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ if σ_1, σ_2 unknown	$z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ if σ_1, σ_2 unknown

where $\bar{x}_1$ = first sample mean, $\bar{x}_2$ = Second Sample mean
n_1 = first Sample Size, n_2 = Second Sample Size
σ_1 = first population S.D, σ_2 = Second population S.D
s_1 = first Sample S.D, s_2 = Second Sample S.D.

① The means of two large Samples of Sizes 1000 and 2000 members are 67.5 inches and 68.0 inches respectively. Can the Samples be regarded as drawn from the Same population of S.D 2.5 inches.

Sol Given $n_1 = 1000$, $n_2 = 2000$, $\bar{x}_1 = 67.5$, $\bar{x}_2 = 68$, $\sigma = 2.5 = \sigma_1 = \sigma_2$

(1) Null Hypotheses $H_0 : \mu_1 = \mu_2$ (i.e Samples drawn from Same population)

(2) Alternative Hypothesis $H_1 : \mu_1 \neq \mu_2$ < two tailed Test >

(3) Level of Significance : $\alpha = 5\%$.

(4) Test Statistic : $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} = \dfrac{67.5 - 68}{\sqrt{\frac{(2.5)^2}{1000} + \frac{(2.5)^2}{2000}}} = -5.16$

If Same Samples have been drawn from the population with Common S.D 'σ' then $\sigma_1 = \sigma_2 = \sigma$

$|z_{cal}| = 5.16$

(5) Conclusion : By Table $z_{tab} = 1.96 \Rightarrow |z_{tab}| = 1.96$

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

$\therefore$ Ho is rejected

$\therefore$ the Samples are drawn from the Same population of SD 2.5 inches

(2) A Company claims that the light bulbs are Superior to Those of its main Compilitor. If a study showed that a sample of $n_1 = 40$ of its bulbs has a mean life time of 647h of Continous use with standard deviation of 27h, while a Sample of $n_2 = 40$ bulbs made by the Competitor had a mean lifetime of 638h of Continous use with standard deviation of 31h. Does this Support the claim at 0.05 LOS?

Sol Given $n_1 = 40$, $\bar{x}_1 = 647$, $s_1 = 27$, $n_2 = 40$, $\bar{x}_2 = 638$, $s_2 = 31$ and $\alpha = 0.05 = 5\%$

(1) Null Hypothesis H_0: $\mu_1 = \mu_2$

(2) Alternative Hypothesis H_1: $\mu_1 > \mu_2$

(3) Level of Significance : $\alpha = 5\%$

(4) Test Statistic : $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \dfrac{647 - 638}{\sqrt{\dfrac{27^2}{40} + \dfrac{31^2}{40}}} = 1.38$

$\therefore |Z_{cal}| = 1.38$

(5) Conclusion : By the Table $Z_{tab} = 1.645 \Rightarrow |Z_{tab}| = 1.645$

$\therefore |Z_{cal}| < |Z_{tab}|$.

H_0 is accepted.

(3) A sample of height of 6400 Englishmen has a mean 67.85 inches and S.D of 2.56 inches while a sample of heights of 1600 Austrains has a mean of 68.55 inches and S.D of 2.52 inches. Do the data indicate the Austrians are on the average taller than the English? (use α as 0.01).

Sol Given $n_1 = 6400$, $\bar{x}_1 = 67.85$, $s_1 = 2.56$
$n_2 = 1600$, $\bar{x}_2 = 68.55$, $s_2 = 2.52$ and $\alpha = 0.01 = 1\%$ (1 - English, 2 - Austrians

(1) Null Hypotheses : H_0: $\mu_1 = \mu_2$ (2) Alternate Hypotheses H_1: $\mu_1 < \mu_2$

(3) Level of significance $\alpha = 0.01 = 1\%$. (4) Test Statistic : $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = -9.9$

$\therefore |Z_{cal}| = 9.9$

(5) Conclusion : By Table $Z_{tab} = -2.33 \Rightarrow |Z_{tab}| = 2.33$

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

$\therefore$ H_0 is rejected.

i.e Thus Austrains are Taller than Englishmen.

(4) Two types of new cars produced in U.S.A are tested for petrol mileage, one sample consisting of 42 cars averaged 15 Kmpl while the other sample consisting of 80 cars averaged 11.5 kmpl with population variances $\sigma_1^2 = 2.0$ and $\sigma_2^2 = 1.5$ respectively. Test whether there is any significance difference in the petrol consumption of these two types of cars (use $\alpha = 0.01$).

Sol Given $n_1 = 42$, $\bar{x}_1 = 15$, $\sigma_1^2 = 2$, $n_2 = 80$, $\bar{x}_2 = 11.5$, $\sigma_2^2 = 1.5$, $\alpha = 1\%$.

(1) $H_0: \mu_1 = \mu_2$ (2) $H_1: \mu_1 \neq \mu_2$ (3) $\alpha = 1\%$. (4) $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

$\therefore z = 13.587$

$\therefore |z_{cal}| = 13.587$

(5) Conclusion: $z_{tab} = 2.58 \Rightarrow |z_{tab}| = 2.58$.

$\therefore |z_{cal}| > |z_{tab}|$. $\therefore$ H_0 is rejected.

Hence there is significance difference in petrol consumption.

(5) It has been claimed that the resistance of electric wire can be reduced by more than 0.050 ohm by alloying. To test this claim, 32 values obtained for standard wire yielded $\bar{x}_1 = 0.136$ ohm and $s_1 = 0.004$ ohm and 32 values obtained alloyed wire yielded $\bar{x}_2 = 0.083$ ohm and $s_2 = 0.005$ ohm. At 0.05 ohm LOS, does this support the claim?

Sol Given $n_1 = 32$, $\bar{x}_1 = 0.136$, $\bar{x}_2 = 0.083$, $s_1 = 0.004$, $s_2 = 0.005$, $\alpha = 5\%$.

(1) Null Hypothesis $H_0: \mu_1 - \mu_2 = 0.050$ (2) Alternative Hypothesis $H_1: \mu_1 - \mu_2 > 0.050$

(3) Level of Significance: $\alpha = 5\%$. (4) Test Statistic: $z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

$$\therefore z_{cal} = \frac{(0.136 - 0.083) - 0.050}{\sqrt{\dfrac{(0.004)^2}{32} + \dfrac{(0.005)^2}{32}}} = 2.65$$

$\therefore |z_{cal}| = 2.65$

(5) Conclusion: $z_{tab} = 1.645 \Rightarrow |z_{tab}| = 1.645$

$\therefore |z_{cal}| > |z_{tab}|$ $\therefore$ H_0 is rejected.

Note: - P-value of Test is $P(z > 2.65) = 0.5 - A(2.65) = 0.5 - 0.4960 = 0.004$//

Test of Hypothesis for one proportion < Large Sample >

Test Statistic $z = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}}$ where p = Sample proportion
P = population proportion.
n = Sample Size.

Note: ① Confidence interval (or) Confidence limits of population proportion 'P' is

$$\left(p - z_{\alpha/2}\sqrt{\frac{pq}{n}},\ p + z_{\alpha/2}\sqrt{\frac{pq}{n}} \right).$$

① 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?

Sol Sample Size $n = 1000$

sample proportion of rice eaters $p = \frac{540}{1000} = 0.54$.

population proportion of rice eaters $P = \frac{1}{2} = 0.5$.

$\Rightarrow Q = 1 - P = 1 - 0.5 = 0.5$.

(1) Null Hypothesis H_0: $P = 0.5$ (i.e Both rice and wheat are equally popular in the State)

(2) Alternative Hypothesis H_1: $P \neq 0.5$. < two tailed Test >

(3) Level of Significance : $\alpha = 1\%$

(4) Test Statistic : $z = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}} = \dfrac{0.54 - 0.5}{\sqrt{\dfrac{0.5 \times 0.5}{1000}}} = 2.532$

$\therefore |z_{cal}| = 2.532$

(5) Conclusion: By the Table $z_{tab} = 2.58 \Rightarrow |z_{tab}| = 2.58$.

$\therefore |z_{cal}| < |z_{tab}|$

$\therefore$ H_0 is accepted

$\therefore$ Rice and wheat eaters are Equally popular in the State.

(2) 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? (use $\alpha = 0.01$)

Sol Given $n = 9000$, sample proportion $p = \frac{3220}{9000} = 0.3578$.

population proportion of getting 3 or 4 is $P = \frac{2}{6} = \frac{1}{3} = 0.3333$.

$\Rightarrow Q = 1 - P = 1 - 0.3333 = 0.6667$.

(1) Null Hypothesis $H_0 : P = \frac{1}{3}$ ⟨i.e the die is unbiased⟩

(2) Alternative Hypothesis $H_1 : P \neq \frac{1}{3}$ (3) Level of significance: $\alpha = 1\%$.

(4) Test Statistic: $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.3578 - 0.3333}{\sqrt{\frac{0.3333 \times 0.6667}{9000}}} = 4.94$

$\therefore |z_{cal}| = 4.94$.

(5) Conclusion: $z_{tab} = 2.58 \Rightarrow |z_{tab}| = 2.58$.

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

$\therefore$ H_0 is rejected.

i.e The die is biased.

(3) A Company prescribed drug for relieving nervous tension is believed to be only 60% effective. Experimental results with a new drug administered a random sample of 100 adults who had been suffering from nervous tension 70 received relief. Is this sufficient evidence to conclude that the new drug is better? use 0.05 LOS.

Sol Given $P = 0.6$, $\Rightarrow Q = 1 - P = 1 - 0.6 = 0.4$.

$p = \frac{70}{100} = 0.7$, $\alpha = 0.05 = 5\%$.

(1) Null Hypothesis $H_0 : P = 0.6$ (2) Alternative Hypothesis $H_1 : P > 0.6$

(3) Level of Significance: $\alpha = 5\%$. (4) Test Statistic: $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$

$\therefore z = \frac{0.7 - 0.6}{\sqrt{\frac{0.6 \times 0.4}{100}}} = 2.04$. $\therefore |z_{cal}| = 2.04$

(5) Conclusion: $z_{tab} = 1.645 \Rightarrow |z_{tab}| = 1.645$

$\therefore |z_{cal}| > |z_{tab}|$ $\therefore$ H_0 is rejected

i.e New drug is superior.

(4) In a big city 325 men out of 600 men were found to be smokers. Does the information support the conclusion that the majority of men in this city are smokers?

Sol

$n = 600$, $p = \frac{325}{600} = 0.5417$, Let $P = \frac{1}{2} \Rightarrow Q = 1 - P = 1 - \frac{1}{2} = \frac{1}{2}$

(1) Null Hypothesis H_0: $P = 0.5$ ⟨i.e the number of smokers and non-smokers are equal in the city⟩.

(2) Alternative Hypothesis H_1: $P > 0.5$ (3) Level of Significance: $\alpha = 5\%$.

(4) Test Statistic: $$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.5417 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{600}}} = 2.04$$

$\therefore |Z_{cal}| = 2.04$.

(5) Conclusion: $Z_{tab} = 1.645 \Rightarrow |Z_{tab}| = 1.645$

$\therefore |Z_{cal}| > |Z_{tab}|$ $\therefore$ H_0 is rejected

i.e The majority of men in the city are smokers.

(5) A manufacturer claimed that atleast 95% of the equipment which he supplied to a factory conformed to specifications. An Examination of a sample of 200 pieces of equipment revealed that 18 were faulty. Test his claim at 5% level of significance.

Sol Given $n = 200$, Number of pieces confirming to specifications $= 200 - 18 = 182$.

p = proportion of pieces confirming to specifications $= \frac{182}{200} = 0.91$

P = population proportion $= 95\% = 0.95$

$Q = 1 - P = 0.05$

(1) Null Hypothesis H_0: $P = 95\% = 0.95$ (2) Alternative Hypothesis: $H_1: P > 0.95$.

(3) Level of Significance: $\alpha = 5\%$. (4) Test Statistic, $Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$

$$\therefore Z = \frac{0.91 - 0.95}{\sqrt{\frac{0.95 \times 0.05}{200}}} = -2.59 \quad \therefore |Z_{cal}| = 2.59$$

(5) Conclusion: $Z_{tab} = 1.645 \Rightarrow |Z_{tab}| = 1.645$

$\therefore |Z_{cal}| > |Z_{tab}|$ $\therefore$ H_0 is rejected

Hence the manufacture claim is accepted

⑥ A manufacturer of a patents medicine claimed that it was 90% effective in relieving an allergy for a period of 8h. In sample of 200 people who had the allergy, the medicine provided relief for 160 people (a) Determine whether the manufacturer's claim is legimate by using 0.01 as the LOS

(b) find the P-Value of the test.

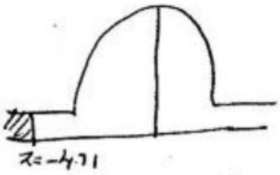
Sol Given $P = 90\% = 0.9$, $Q = 0.1$, $p = \frac{160}{200} = 0.8$, $n = 200$.

① Null Hypothesis: $H_0 : P = 0.9$ ② Alternative Hypothesis: $H_1 : P < 0.9$

③ Level of significance: $\alpha = 0.01 = 1\%$. ④ Test Statistic: $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$

$$\therefore z = \frac{0.8 - 0.9}{\sqrt{\frac{0.9 \times 0.1}{200}}} = -4.71 \quad \therefore |z_{cal}| = 4.71$$

⑤ Conclusion: $z_{tab} = -2.33 \Rightarrow |z_{tab}| = 2.33$



$\therefore |z_{cal}| > |z_{tab}|$ $\therefore$ H_0 is rejected

(b) P-Value of Test is $P(z \le -4.71) = 0.5 - A(4.71) = 0.5 - A(4.71)$
$= 0.5 - 0.5$
$= 0$

⑦ In a study designed to investigate whether certain detonators used with Explosives in coal mining meet the requirement that at least 90% will ignite the explosive when charged, it is found that 174 of 200 detonators function properly. Test the NH $P = 0.9$ against the AH $P < 0.9$ at the 0.05 LOS.

Sol Given $P = 0.9$, $Q = 0.1$ $p = \frac{174}{200} = 0.87$, $n = 200$

① $H_0 : P = 0.9$, ② $H_1 : P < 0.9$ ③ $\alpha = 5\%$ ④ $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$

$$\therefore z = \frac{0.87 - 0.9}{\sqrt{\frac{0.9 \times 0.1}{200}}} = -1.41 \quad \therefore |z_{cal}| = 1.14$$

⑤ Conclusion: $z_{tab} = -1.645 \Rightarrow |z_{tab}| = 1.645$

$|z_{cal}| < |z_{tab}|$ $\therefore$ H_0 is accepted.

⑧ In a random sample of 400 Industrial accidents, it was found that 231 were due to unsafe working conditions. Construct a 99% confidence interval for the corresponding true proportion using Large sample formula. ⟨2015⟩ 2M

Sol Given $n = 400$, $p = \frac{231}{400} = 0.58 \Rightarrow q = 1 - p = 1 - 0.58 = 0.42$

Confidence limit = 99%.

$z_{\alpha/2}$ for 99% confidence = 2.58

$\therefore$ Confidence interval for P is given by $\left(p - z_{\alpha/2}\sqrt{\frac{pq}{n}},\ p + z_{\alpha/2}\sqrt{\frac{pq}{n}}\right)$

$$\left(0.58 - 2.58\sqrt{\frac{0.58\times0.42}{400}},\ 0.58 + 2.58\sqrt{\frac{0.58\times0.42}{400}}\right)$$

$$(0.58 - 0.064,\quad 0.58 + 0.064)$$

$$(0.516,\ 0.644)$$

⑨ Among 900 people in a state 90 are found to be chapati eaters. Construct 99% confidence interval for the true population. ⟨2017⟩

Sol Given $n = 900$, $p = \frac{90}{900} = 0.1$, $q = 1 - p = 0.9$.

Confidence limit = 99%, $z_{\alpha/2}$ for 99% confidence = 2.58.

Confidence interval for P is $\left(p - z_{\alpha/2}\sqrt{\frac{pq}{n}},\ p + z_{\alpha/2}\sqrt{\frac{pq}{n}}\right)$

$$\left(0.1 - 2.58\sqrt{\frac{0.1\times0.9}{900}},\ 0.1 + 2.58\sqrt{\frac{0.1\times0.9}{900}}\right)$$

$$(0.1 - 0.026,\ 0.1 + 0.026)$$

$$(0.074,\ 0.126)$$

⑩ In a random sample of 400 Industrial accidents, it was found that 231 were due to unsafe working conditions. What can we say 95% confidence about the maximum error.

Sol Given $p = \frac{231}{400} = 0.58$, $q = 1 - p = 0.42$, $n = 400$

$z_{\alpha/2}$ for 95% confidence = 1.96

$\therefore$ Maximum error $E = z_{\alpha/2}\sqrt{\frac{pq}{n}} = 1.96\sqrt{\frac{0.58\times0.42}{400}} = 0.048$

(11) A Random Sample of 500 apples was taken from a large consignment of 60 were found to be bad. Obtain the 98% Confidence limits for the percentage number of bad apples in consignment. <DEC2017>

Sol Given $n = 500$, $p = \frac{60}{500} = 0.12$, $q = 1 - p = 0.88$

Confidence limit $= 98\%$. < Here $\alpha = 2\%$

$z_{\alpha/2}$ for 98% confidence $= 2.33$.

$\therefore$ Confidence limits for bad apples.

$$\left(p - z_{\alpha/2}\sqrt{\frac{pq}{n}},\ p + z_{\alpha/2}\sqrt{\frac{pq}{n}}\right)$$

$$(0.12 - 2.33 \times 0.014,\ 0.12 + 2.33 \times 0.014) = (0.086, 0.1584)$$

Confidence limits for percentage of bad apples $= (8.7, 15.3)$.

2m
(12) If 80 patients are Treated with an antibiotic 59 got card. Find a 99% Confidence limits to the true population. <Dec 2017>

Sol Given $n = 80$, $p = \frac{59}{80} = 0.7375$

$$q = 1 - p = 1 - 0.7375 = 0.2625$$

Confidence limit $= 99\%$.

$z_{\alpha/2}$ for 99% Confidence $= 2.58$

$\therefore$ Confidence interval for P is $\left(p - z_{\alpha/2}\sqrt{\frac{pq}{n}},\ p + z_{\alpha/2}\sqrt{\frac{pq}{n}}\right)$

$$(0.7375 - 2.58 \times 0.049,\ 0.7375 + 2.58 \times 0.049)$$

$$(0.61,\ 0.86)$$

Test of Hypothesis of two proportions (Large Sample)

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

where $p = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$, $q = 1-p$.

n_1 = first sample size, p_1 = first sample proportion

n_2 = second sample size, p_2 = second sample proportion.

Note:-

① Suppose the population proportions P_1 and P_2 are given and $P_1 \neq P_2$. If we want to test the Hypothesis that the difference $(P_1 - P_2)$ in population proportions is likely to be hidden in samples of sizes n_1 and n_2 from two populations respectively. then

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

If sample proportions are not known then we use

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

① If 120 out of 200 patients suffering from a certain disease are cured by allopathy and 240 out of 500 patients are cured by homeopathy, is there reason enough to believe that allopathy is better than homeopathy in curing the disease? use $\alpha = 0.05$ LOS ⟨2014⟩

Sol Given $n_1 = 200$, $p_1 = \frac{120}{200} = 0.6$, $n_2 = 500$, $p_2 = \frac{240}{500} = 0.48$

① Null Hypothesis H_0: $P_1 = P_2$ (i.e There is no difference between allopathy and homeopathy)

② Alternative Hypothesis H_1: $P_1 > P_2$ (i.e Allopathy is better than homeopathy).

③ Level of Significance: $\alpha = 0.05 = 5\%$

④ Test statistic $z = \dfrac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

where $p = \dfrac{n_1p_1 + n_2p_2}{n_1+n_2} = \dfrac{200\times 0.6 + 500\times 0.48}{700} = \dfrac{120+240}{700} = 0.51.$

$\therefore q = 1-p = 1-0.51 = 0.49.$

$\therefore z = \dfrac{(0.6-0.48)}{\sqrt{(0.51)(0.49)\left(\frac{1}{200}+\frac{1}{500}\right)}} = 2.9$

$\therefore |z_{cal}| = 2.9.$

⑤ Conclusion: $|z_{tab}| = 1.645$

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

$\therefore$ H_0 is rejected.

$\therefore$ Allopathy is better than Homeopathy.

② In a random sample of 1000 persons from town A, 400 are found to be consumers of wheat. In sample of 800 from town B, 400 are found to be consumers of wheat. Do these data reveal a significance difference Town A and Town B, so far as the proportion of wheat consumers is concerned.

Sol Given $n_1 = 1000$, $p_1 = \frac{400}{1000} = 0.4$, $n_2 = 800$, $p_2 = \frac{400}{800} = 0.5$.

(i) Null Hypothesis H_0: $P_1 = P_2$ (ii) Alternative Hypotheses H_1: $P_1 \neq P_2$

(iii) Level of Significance: $\alpha = 5\%$ (iv) Test statistic $z = \dfrac{p_1-p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

where $p = \dfrac{n_1p_1 + n_2p_2}{n_1+n_2} = \dfrac{1000\times 0.4 + 800\times 0.5}{1800} = \dfrac{4}{9}$

$\therefore q = 1-p = 1-\frac{4}{9} = \frac{5}{9}.$

$\therefore z = \dfrac{(0.4-0.5)}{\sqrt{\frac{4}{9}\cdot\frac{5}{9}\left(\frac{1}{1000}+\frac{1}{800}\right)}} = -4.242 \quad \therefore |z_{cal}| = 4.242$

(v) Conclusion: $|z_{tab}| = 1.96$, $\therefore |z_{cal}| > |z_{tab}|$

$\therefore$ H_0 is rejected

(3) A Study of TV Viewers was Conducted to find the opinion about mega Serial Ramayana. If 56% of a sample 300 Viewers from South and 48% of 200 Viewers from north prefered the Serial, Test the claim at 0.05 LOS that

(a) There is difference of opinion between South and north

(b) Ramayana is prefered in south.

Sol Given $n_1 = 300$, $p_2 = 48\% = 0.48$, $n_2 = 200$. $\alpha = 0.05 = 5\%$.

$p_1 = 56\% = 0.56$

(a)

(i) Null Hypotheses H_0: $P_1 = P_2$

(ii) Alternative Hypotheses H_1: $P_1 \neq P_2$

(iii) Level of Significance: $\alpha = 0.05 = 5\%$.

(iv) Test Statistic:

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

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

$= \frac{200 \times 0.48 + 300 \times 0.56}{500}$

$p = 0.53$

$q = 1 - p = 0.47$

$$\therefore z = \frac{0.56 - 0.48}{\sqrt{0.53 \times 0.47\left(\frac{1}{300}+\frac{1}{200}\right)}} = \frac{0.08}{0.045}$$

$z = 1.78$ $\therefore |z_{cal}| = 1.78$

(v) Conclusion: $|z_{tab}| = 1.96$

$\therefore |z_{cal}| < |z_{tab}|$

$\therefore$ Ho is accepted.

(b) (i) Null Hypotheses H_0: $P_1 = P_2$

(ii) Alternative Hypotheses H_1: $P_1 > P_2$

(iii) Level of Significance: $\alpha = 0.05 = 5\%$.

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

Here $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = 0.53$

$q = 1 - P = 1 - 0.53 = 0.47$

$\therefore |z_{cal}| = 1.78$.

(v) Conclusion: $|z_{tab}| = 1.645$

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

$\therefore$ Ho is rejected.

$\therefore$ Ramayana is just prefered in the South.

(4) A random Sample of 300 shoppers at a Super market includes 204, who regularly use cents off Coupons. In another Sample of 500 Shoppers at a Super market includes 75, who regularly use cents off Coupons. Test the Significance difference of two proportions at 2% level. Construct Confidence interval for the probability that any one shopper in sample selected at random will use regularly Cents off Coupons. ⟨2017⟩

Sol $n_1 = 300 \qquad n_2 = 500$

(a) $P_1 = \frac{204}{300} = 0.68, \quad P_2 = \frac{75}{500} = 0.15$

$H_0: P_1 = P_2, \; H_1: P_1 \neq P_2$

$\alpha = 2\%$.

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

$\Rightarrow \boxed{p = 0.35}$

$\Rightarrow q = 0.65$.

$$Z = \frac{0.68 - 0.15}{\sqrt{(0.35)(0.65)\left(\frac{1}{300}+\frac{1}{500}\right)}} = 15.14.$$

$|Z_{cal}| = 15.14$.

Conclusion:- By Table $|Z_{tab}| = 2.33$.

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

$\therefore$ Ho is rejected

There is significant difference between two proportions

Confidence interval:-

$Z_{\alpha/2}$ for 98% confidence $= 2.33$.

$\therefore$ confidence interval for the probability that any one shopper in sample selected at random is

$$\left(p_1 - Z_{\alpha/2}\sqrt{\frac{p_1 q_1}{n_1}}, \; p_1 + Z_{\alpha/2}\sqrt{\frac{p_1 q_1}{n_1}}\right)$$

$$\left(0.68 - 2.33\sqrt{\frac{0.68 \times 0.32}{300}}, \; 0.68 + 2.33\sqrt{\frac{0.68 \times 0.32}{300}}\right)$$

$(0.68 - 0.063, \; 0.68 + 0.063)$

$(0.62, \; 0.74)$.

(5) Among the items produced by a factory out of 800, 65 were defective. In another sample out of 300, 40 were defective. Test the significance between the differences of two proportions at 1% level. ⟨Dec 2017⟩

Sol Given $n_1 = 800, \; n_2 = 300, \; P_1 = \frac{65}{800} = 0.081, \; P_2 = \frac{40}{300} = 0.1333$

Null Hypothesis $H_0: P_1 = P_2$, Alternative Hypothesis $H_1: P_1 \neq P_2$, Level of significance $= \alpha = 1\%$.

Test Statistic: $$Z = \frac{P_1 - P_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}, \quad p = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

$$p = \frac{800 \times 0.081 + 300 \times 0.1333}{800 + 300} = 0.095$$

$\Rightarrow q = 0.905$

$$Z = \frac{0.081 - 0.1333}{\sqrt{0.095 \times 0.905\left(\frac{1}{800}+\frac{1}{300}\right)}} = -2.6324$$

$\therefore |Z_{cal}| = 2.6324$.

Conclusion: By Table $|Z_{tab}| = 2.58$

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

$\therefore$ Ho is rejected.