UNIT – III

Estimation and Testing of Hypothesis

Sampling distribution of means

Population

It consists of the totality of the observations with which we are concerned.

The number of observations in the population is defined to be the size of the population.

Population can be finite or infinite

Sample

It is a subset of population.

Sampling

It is the selection of a subset of individuals from within a population to estimate characteristics of the whole population.

Random sampling is the one in which each unit of the population has an equal chance of being included in it.

Classification of samples:

Samples are classified in two ways.

1. Large sample : If the sample size $(n) \geq 30$
2. small sample : $(n) < 30$

In sampling with replacement , each member of the population may be chosen more than once , since the member is replaced in the population.

Thus sampling from finite population with replacement can be considered as sampling from infinite population.

In sampling without replacement, an element of the population cannot be chosen more than once , as it is not replaced.

Therefore the sampling from finite population without replacement can be considered as sampling from finite population only.

Statistic:

Any measures computed from sample observations are known as statistics.

Example :

mean $\left(\bar{x}\right)$, variance $\left(s^2\right)$

Parameters

Any measures computed from Population observations are known as Parameters.

Example :

mean (μ), variance (s^2)

The sample mean

If $x_1, x_2, x_3, ----- x_n$ represent a random sample o size 'n' then the sample mean is defined by the

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

The sample variance

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

sample standard deviation is positive square root of sample variance

Sampling distribution:

Sampling distribution of a statistic helps us to get information about the corresponding parameter.

Definition

Sampling distribution of a statistic is called a sampling distribution

If we draw a sample of size ‘n’ from a given finite population of size ‘N’; the total number of possible sample is $NC_n = \dfrac{N!}{n!(N-n)!}$

For each of these samples we can compute some statistic (sample mean $\bar{x}$, variance s^2)

The set of values of the statistic s obtained, one for each sample, constitutes the sampling distribution of the statistic

Standard Error

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

The standard error = $\dfrac{\sigma}{\sqrt{n}}$

Sampling distribution of means (known):

The probability distribution of $\bar{x}$ is called the sampling distribution of the mean.

Infinite population:

Sampling is done with replacement

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

Variance $\left(\sigma_{\bar{x}}\right)^2 = \dfrac{\sigma^2}{n}$

standard deviation = $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

Central Limit theorem

If $\overline{x}$ is the mean of a random sample of size ‘n’ taken from a population and finite variance $(\sigma)^2$, then the limiting form of the distribution of

$$z = \frac{\overline{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

as $n \to \infty$, is the standard normal distribution (N(0,1))

Finite population:

Consider a finite population of size N with mean μ

and standard deviation σ

Draw all possible samples of size 'n' without replacement , from this population.

Then the mean of the sampling distribution of

means (for N>n) is $\mu_{\bar{x}} = \mu$

Variance :

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

standard deviation :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\left(\sqrt{\frac{N-n}{N-1}}\right)$$

$\left(\frac{N-n}{N-1}\right)$ is called finite population correction factor

Problem:1

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

Solution

Correction factor $\left(\frac{N-n}{N-1}\right)=\left(\frac{N-n}{N-1}\right)$

$$=\left(\frac{1000-10}{1000-1}\right)=\frac{990}{999}=0.991$$

Problem :2

The variance of the population is 2. The size if the sample collected from the population is 169. what is the standard error of mean?

Solution

$$\sigma = \sqrt{2}$$

n=169

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

= $\dfrac{\sqrt{2}}{\sqrt{169}} = 0.185$

Problem:3

A random sample of size 100 is taken from an infinite population having the mean 76 and the variance 256. what is the probability that $\bar{x}$ will be between 75 and 78.

Solution

n=100, sample size

mean of the population=76

variance=256

By Central limit Theorem

$$z = \frac{\overline{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

When $\overline{x} = 75$

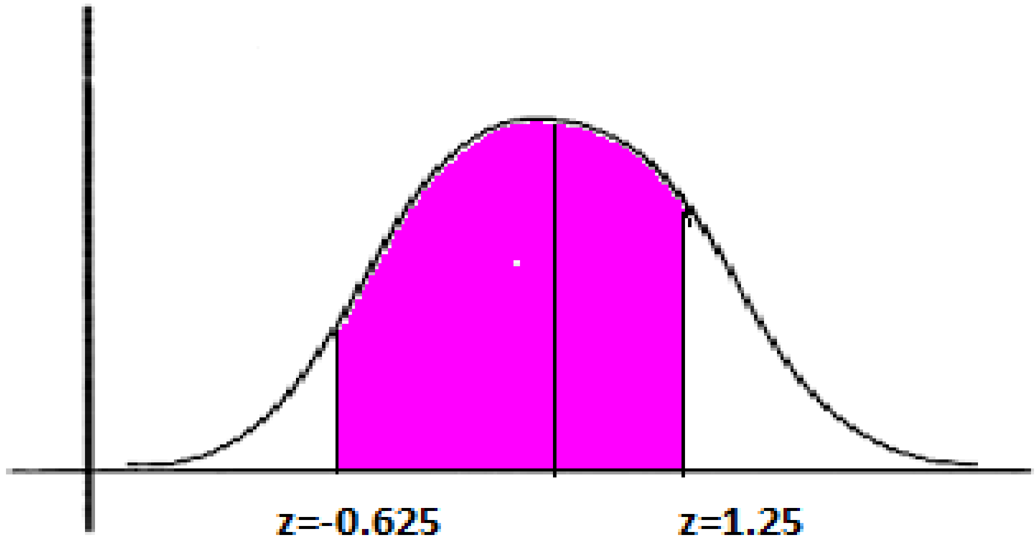
$$z_1 = \frac{75 - 76}{\left(\frac{16}{\sqrt{100}}\right)} = -0.625$$

When $\overline{x} = 78$

$$z_1 = \frac{78 - 76}{\left(\frac{16}{\sqrt{100}}\right)} = 1.25$$

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

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



$$=0.2334+0.3944=0.628$$

Problem:4

A population consists of five numbers 2,3,6,8and 11.Consider all possible samples of size 2 that can be drawn with replacement from this population.

Find

a) The mean of the population

b) The standard deviation of the population

c) The mean of the sampling distribution of means
d) The standard deviation of the sampling distribution of means(i.e., the standard error of means) .

Solution

a) Mean of the population $\mu = \dfrac{2+3+6+8+11}{5} = 6$

a) Variance of the population

$$\sigma^2 = \sum \frac{\left(x_i - \bar{x}\right)^2}{n}$$

$$= \frac{(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2}{5} = 10.8$$

Standard deviation of the population

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

a) Sampling with replacement(infinite population):

The total number of samples with replacement is $N^n = 5^2 = 25$

25 samples of size 2 are

(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)

Now compute the statistic the arithmetic mean for each of these 25 samples . The set of 25 means $\bar{x}$ of these 25 samples, give rise to the distribution of means of the samples known as sampling distribution of means.

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

The mean of sampling distribution of means is means of these 25 means.

$$\mu_{\bar{x}} = \frac{sum\ of\ all\ sample\ means}{25} = \frac{150}{25} = 6$$

This shows that $\mu_{\bar{x}} = \mu$

a) For finite population involving sampling with replacement , variance of the sampling distribution of means

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

standard deviation of the sampling distribution of means

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

Problem:5

A population consists of five numbers 2,3,6,8and 11.Consider all possible samples of size 2 that can be drawn without replacement from this population.

Find

a) The mean of the population

b) The standard deviation of the population

c) The mean of the sampling distribution of means

d) The standard deviation of the sampling distribution of means(i.e., the standard error of means) .

Solution

a) Mean of the population $\mu = \dfrac{2+3+6+8+11}{5} = 6$

b) Variance of the population

$$\sigma^2 = \sum \frac{\left(x_i - \bar{x}\right)^2}{n}$$

$$= \frac{(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2}{5} = 10.8$$

Standard deviation of the population

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

c) sampling without replacement(finite population)

the total number of samples without replacement is $NC_n = 5C_2 = 10$ samples of size 2.

The 10 samples are

(2, 3) (2, 6) (2, 8) (2,11)

(3, 6) (3, 8) (3, 11)

(6, 8) (6, 11)

(8,11)

In this case the selection (2,3) is considered same as (3,2)

The corresponding sample means are

2.5 4 5 6.5

4.5 5.5 7

7 8.5

9.5

The mean of sampling distribution of means is

$$\mu_{\bar{x}} = \frac{sum\ of\ all\ sample\ means}{10} = \frac{60}{10} = 6$$

This shows that $\mu_{\bar{x}} = \mu$

d) For finite population involving sampling with out replacement , variance of the sampling distribution of means

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

standard deviation of the sampling distribution of means

$$\left(\sigma_{\bar{x}}\right) = \frac{\sigma}{\sqrt{n}}\left(\sqrt{\frac{N-n}{N-1}}\right) \quad = \frac{3.29}{\sqrt{2}}\left(\sqrt{\frac{5-2}{5-1}}\right) = 2.01$$

Problem:6

The mean of certain normal population is equal to the standard error of the mean of the samples of 64 from that distribution. Find the probability that the mean of the sample size 36 will be negative.

Solution

The standard error of means= $\frac{\sigma}{\sqrt{n}}$

Sample size n=64

Mean = μ $=$ standard error= $\frac{\sigma}{8}$

(sample size=64)

We know $z = \dfrac{\bar{x} - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)}$

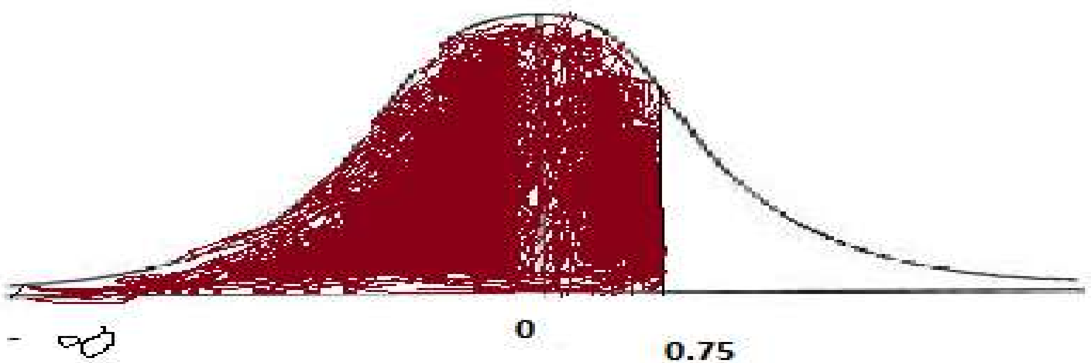
$$= \frac{\bar{x} - \left(\dfrac{\sigma}{8}\right)}{\left(\dfrac{\sigma}{6}\right)}$$ **(sample size =36)**

$$= \frac{\overline{6x}}{\sigma} - \left(\frac{3}{4}\right)$$

If $z < 0.75$, $\overline{x}$ is negative

$$P(Z < 0.75) = P(-\infty < z < 0.75) =$$

0.5+0.2734=0.7734



Problem:7

A normal population has mean of 0.1 and standard deviation of 2.1. Find the probability that mean of a sample of size 900 will be negative

Solution

Mean=0.1

standard deviation=2.1

sample size= 900

we know that

$$z=\frac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}=\frac{\bar{x}-0.1}{\left(\frac{2.1}{\sqrt{900}}\right)}=\frac{\bar{x}-0.1}{\left(\frac{2.1}{30}\right)}=\frac{\bar{x}-0.1}{0.07}$$

Which gives

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

The required probability, that the sample mean is negative is given by

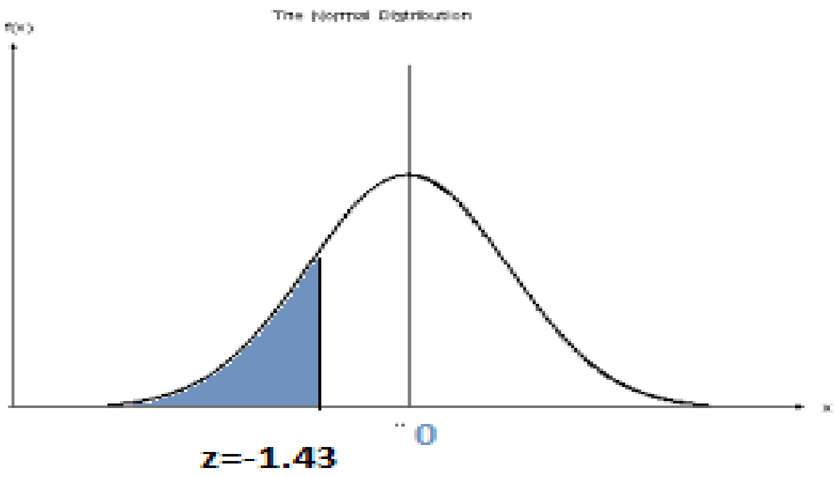
$$P\left(\bar{x} < 0\right) = P\left(\left(0.1 + 0.07Z\right) < 0\right)$$

$$P\left(\bar{x} < 0\right) = P\left(\left(0.07Z\right) < -0.1\right)$$

$$P\left(\bar{x} < 0\right) = P\left(\left(Z\right) < \frac{-0.1}{0.07}\right)$$

$$P\left(\bar{x} < 0\right) = P\left(\left(Z\right) < -1.43\right)$$

= 0.5-0.4236=0.0764



The Normal Distribution
f(x)
x
0
z=-1.43

EXTRA PROBLEMS

PROBLEM: 8

Find the value of finite population correction factor for $n=10$ and $N=100$.

N=1000

n=10

Correction factor = $$\frac{N-n}{N-1}=\frac{1000-10}{1000-1}=0.991$$

PROBLEM:9

Let S={1, 5, 6, 8} find the probability distribution of the sample mean for random sample of size 2 drawn with out replacement.

Let S= { 1, 5 ,6,8}

Size of the sample =2

Without replacement

(1,5) , (1,6) , (1,8)

(5,6) (5,8)

(6,8)

Number of samples = n=6

Sampling distribution of mean

3 , 3.5 , 4.5 , 5.5 , 6.5 , 7

Mean of the sampling distribution

$$\mu_{\bar{x}} = \frac{3+3.5+4.5+5.5+6.5+7}{6} = 5$$

Standard deviation of sampling distribution of means

$$\sigma = \sqrt{\frac{(3-5)^2+(3.5-5)^2+(4.5-5)^2+(5.5-5)^2+(6.5-5)^2+(7-5)^2}{5}}$$

$$= 1.612$$

PROBLEM:10

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

Solution

Size of the sample = n=100

Mean of the population= μ =76

Variance of the population = $\sigma^2 =$ 256

standard deviation = σ =16

$$z = \frac{\overline{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

When $\overline{x_1} = 75$

$$z_1 = \frac{\overline{x_1} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{75 - 76}{\frac{16}{\sqrt{100}}} = -0.625$$

When $\overline{x_2} = 78$

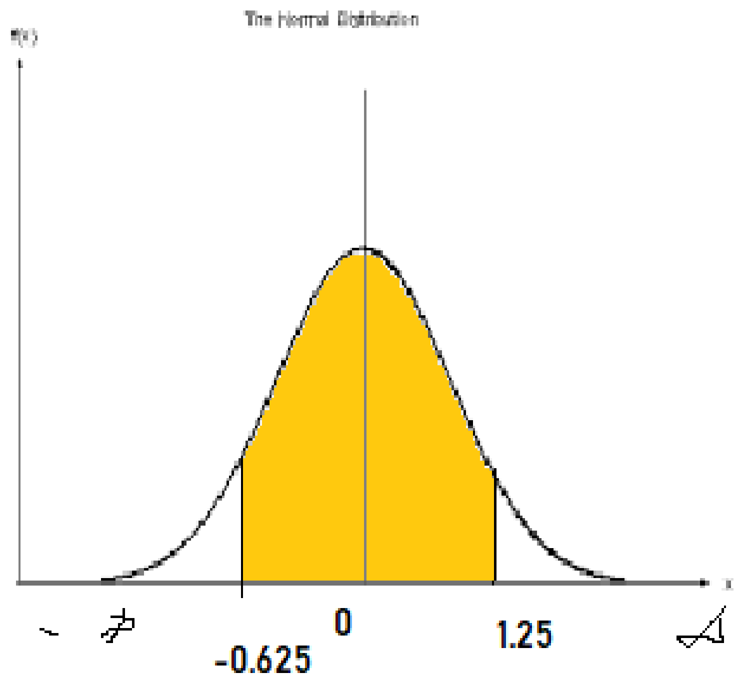
$$z_2 = \frac{\overline{x_2} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{78 - 76}{\frac{16}{\sqrt{100}}}$$

$$= 1.25$$

$$P\left(75 \leq \overline{x} \leq 78\right) = P\left(-0.625 \leq z \leq 1.\right.$$

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

$$0.2334 + 0.3944 = 0.628$$



The Normal Distribution
f(x)
x
0
1.25
-0.625

PROBLEM:11

What is the effect on standard error, if a sample is taken from an infinite population of sample size is increased from 400 to 900.

SOLUTION

sample taken from an infinite population

Standard error of mean = $\dfrac{\sigma}{\sqrt{n}}$

sample size= n = n_1 =400

standard error = $\dfrac{\sigma}{\sqrt{n}}$ = $\dfrac{\sigma}{\sqrt{400}} = \dfrac{\sigma}{20}$ -----(1)

If sample size n= n_2 = 900

standard error = $\dfrac{\sigma}{\sqrt{n}}$ = $\dfrac{\sigma}{\sqrt{900}} = \dfrac{\sigma}{30}$ ______(2)

$$\frac{2}{3}\left(\frac{\sigma}{20}\right) = \frac{\sigma}{30}$$

PROBLEM :12

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.

SOLUTION

sample taken from an infinite population

Standard error of mean = $\dfrac{\sigma}{\sqrt{n}}$

sample size= n = n_1 =800

standard error = $\dfrac{\sigma}{\sqrt{n}}$ = $\dfrac{\sigma}{\sqrt{800}} = \dfrac{\sigma}{20\sqrt{2}}$ -----(1)

If sample size n= n_2 = 200

standard error = $\dfrac{\sigma}{\sqrt{n}}$ = $\dfrac{\sigma}{\sqrt{200}} = \dfrac{\sigma}{10\sqrt{2}}$ ______(2)

$$2\left(\frac{\sigma}{20\sqrt{2}}\right) = \frac{\sigma}{10\sqrt{2}}$$

PROBLEM:13

The mean height of students in a college is 155cms and standard deviation is 15. What is the probability that the mean height of 36 students is less than 157cms.

SOLUTION

Mean of the population = $\mu = 155$

Standard deviation of the population $\sigma = 15$

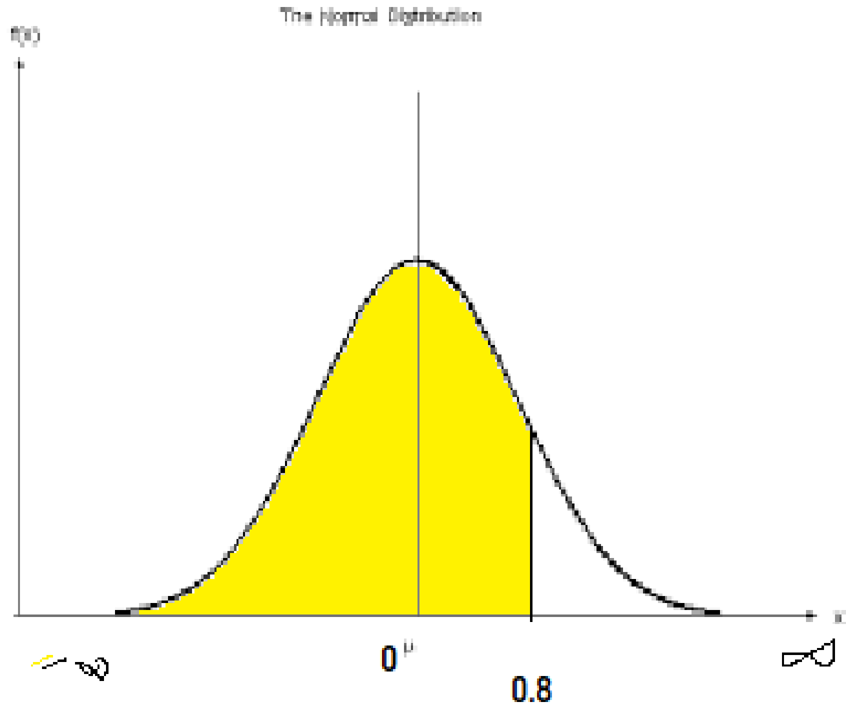
sample size = n=36

Mean of the sample $\overline{x} = 157$

$$z = \frac{\overline{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{157 - 155}{\left(\frac{15}{\sqrt{36}}\right)} = \frac{2}{\left(\frac{15}{6}\right)} = \frac{12}{15} = 0.8$$

$$P\left(\overline{x} \leq 157\right) = P\left(z \leq 0.8\right) = 0.5 + P\left(0 \leq z \leq 0.8\right)$$

=0.5 +0.2881=0.7881


The Normal Distribution
f(x)
x
0
0.8

PROBLEM:14

The variance of a population is 2. The size of the collected from the population is 169. what is the standard error of mean.

Distributions

<u>Sampling Distribution of the mean</u> (σ <u>Unknown):</u>

In case of sampling distribution of the mean with known standard deviation the information about population standard deviation ., σ must be known.

But for large sample of size $(n \geq 30)$, even if standard deviation σ of population is not known , it does not make any difference. Since we can substitute the sample standard deviation "S" in place of σ .

where $$S^2 = \sum_{i=1}^{n} \frac{\left(x_i - \overline{x}\right)^2}{n-1}$$

For small sample of size (n<30), when σ is unknown , it can be substituted by S, provided we make the assumption that the sample is taken from normal population.

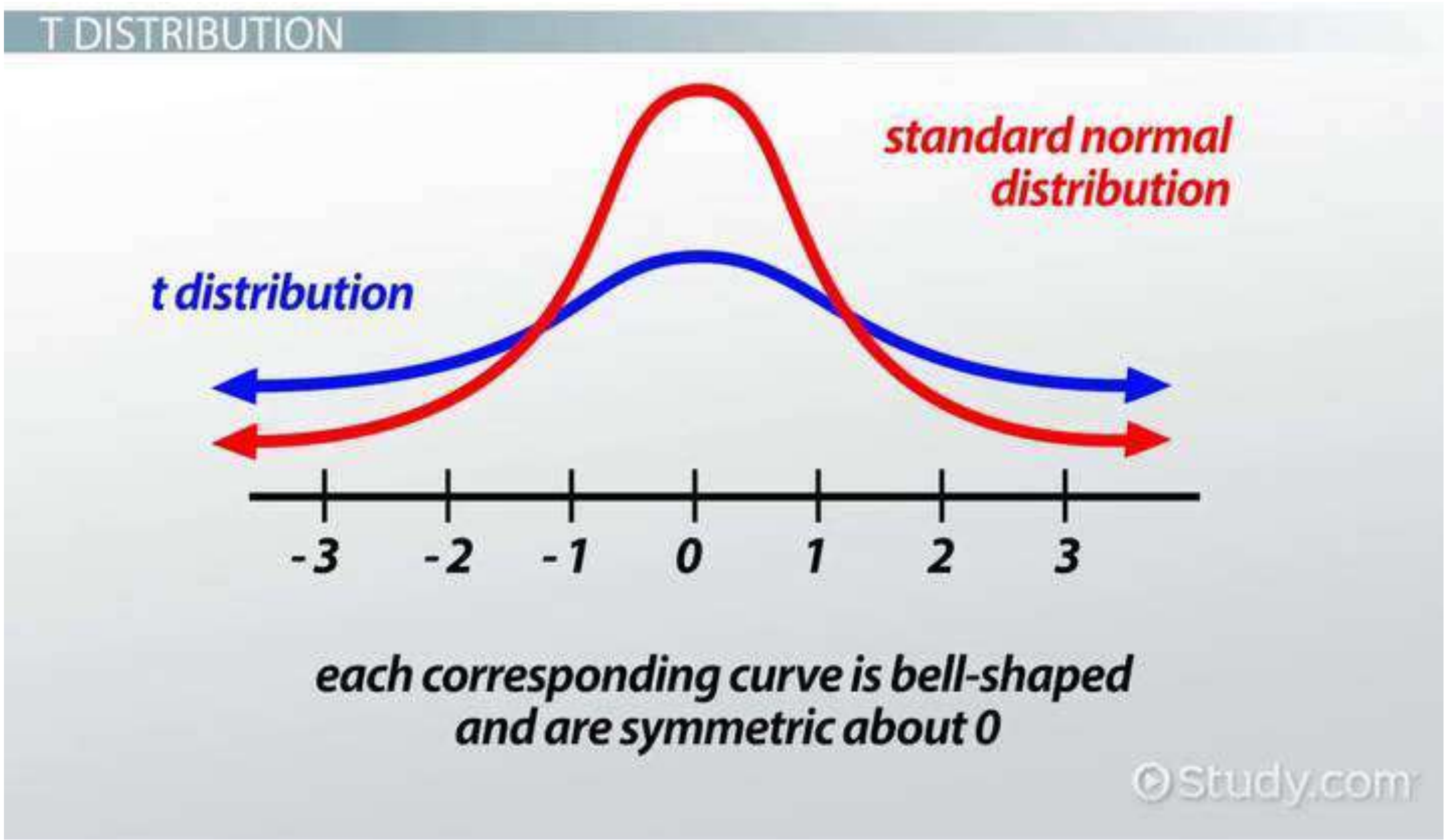
t- distribution (or) Student's t- distribution

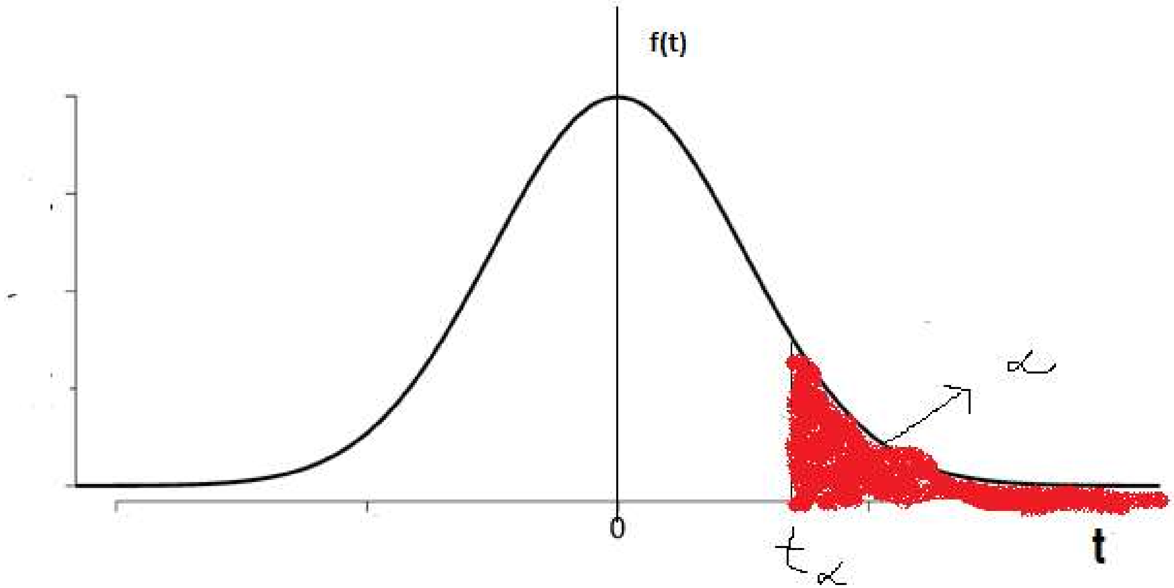
Let $\overline{x}$ be the mean of random sample of size 'n', taken from a normal population having the mean μ and variance σ^2, and $S^2 = \sum_{i=1}^{n} \frac{\left(x_i - \overline{x}\right)^2}{n-1}$, then

$t = \frac{\overline{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}$ is a random variable having the t-distribution with $\nu = n - 1$ degrees of freedom

The degrees of freedom refer to the number of independent observations in a set of data.

Hence, the distribution of the t statistic from samples of size 8 would be described by a t distribution having 8 - 1 or 7 degrees of freedom.


T DISTRIBUTION
standard normal distribution
t distribution
-3
-2
-1
0
1
2
3
each corresponding curve is bell-shaped and are symmetric about 0
© Study.com



f(t)
0
t_α
α
t

The selected values of t_α for various values of "v" can be obtained from the table of t-distribution , where t_α denotes the area under t-distribution to its right is equal to "α".

Degrees of freedom (ν)	Amount of area in one tail (α)							
	0.0005	0.001	0.005	0.010	0.025	0.050	0.100	0.200
1	636.6192	318.3088	63.65674	31.82052	12.70620	6.313752	3.077684	1.376382
2	31.59905	22.32712	9.924843	6.964557	4.302653	2.919986	1.885618	1.060660
3	12.92398	10.21453	5.840909	4.540703	3.182446	2.353363	1.637744	0.978472
4	8.610302	7.173182	4.604095	3.746947	2.776445	2.131847	1.533206	0.940965
5	6.868827	5.893430	4.032143	3.364930	2.570582	2.015048	1.475884	0.919544
6	5.958816	5.207626	3.707428	3.142668	2.446912	1.943180	1.439756	0.905703
7	5.407883	4.785290	3.499483	2.997952	2.364624	1.894579	1.414924	0.896030
8	5.041305	4.500791	3.355387	2.896459	2.306004	1.859548	1.396815	0.888890
9	4.780913	4.296806	3.249836	2.821438	2.262157	1.833113	1.383029	0.883404
10	4.586894	4.143700	3.169273	2.763769	2.228139	1.812461	1.372184	0.879058
11	4.436979	4.024701	3.105807	2.718079	2.200985	1.795885	1.363430	0.875530
12	4.317791	3.929633	3.054540	2.680998	2.178813	1.782288	1.356217	0.872609
13	4.220832	3.851982	3.012276	2.650309	2.160369	1.770933	1.350171	0.870152
14	4.140454	3.787390	2.976843	2.624494	2.144787	1.761310	1.345030	0.868055
15	4.072765	3.732834	2.946713	2.602480	2.131450	1.753050	1.340606	0.866245
16	4.014996	3.686155	2.920782	2.583487	2.119905	1.745884	1.336757	0.864667
17	3.965126	3.645767	2.898231	2.566934	2.109816	1.739607	1.333379	0.863279

The shape of t-distribution is bell shaped , which is similar to that of a normal distribution and is symmetrical about mean.

the mean of standard normal distribution and as well as t-distribution is zero but the variance of t-distribution depends upon the parameter “V” which is called the degrees of freedom.

Chi-squared (χ^2) Distribution

Chi –squared distribution is continuous probability distribution of a continuous random variable X with probability density function given by

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} e^{\frac{-x}{2}} & ; \quad for \quad x>0 \\ 0 & otherwise \end{cases}$$

Where $"\nu"$ is a positive integer is the only single parameter of the distribution, also known as degrees of freedom .

Properties of Chi-squared (χ^2) Distribution:

1. Chi-squared (χ^2) Distribution curve is not symmetrical, lies entirely in the first quadrant , and hence not a normal curve , since χ^2 varies from 0 to ∞
2. It depends only on the degrees of freedom.
3. α denotes the area under the chi-square distribution to the right of χ_{α}^{2}

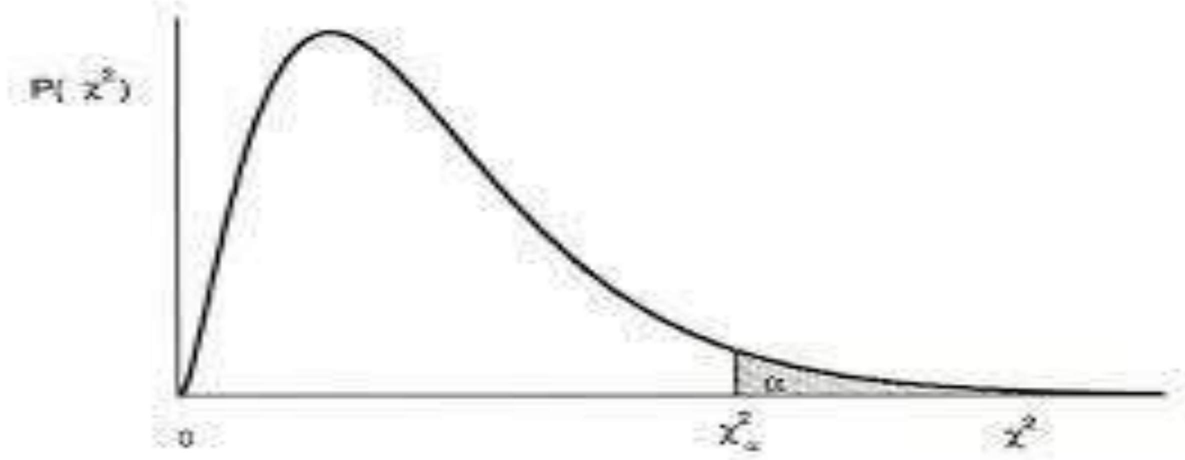


Figure J.1: The χ^2 distribution

Sampling distribution of varianceS^2 :

The theoretical distribution of the sample variance for random samples from normal population is related to the chi-squared distribution

Let S^2 be the sample variance of a random sample of size 'n' , taken from a normal population having the variance σ^2.

Then $\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{\sigma^2}$ is a value of a random variable having the χ^2- distribution with (n-1) degrees of freedom.

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

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

F-Distribution (Sampling Distribution of the Ratio of two Sample Variances:

Let S_1^2, S_2^2 be the sample variance of independent sample of size n_1 , n_2 drawn from a normal population, with variances σ_1^2 , σ_2^2

To determine whether the two samples come from two populations having equal variances,

Consider the sampling distribution of the ratio of the variances of the two independent random samples defined by

$$F = \frac{\left(\frac{S_1^2}{\sigma_1^2}\right)}{\left(\frac{S_2^2}{\sigma_2^2}\right)}$$

This follows F-distribution with $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$ degrees of freedom.

Under the assumption that two normal population have the same variance $\sigma_1^2 = \sigma_2^2$

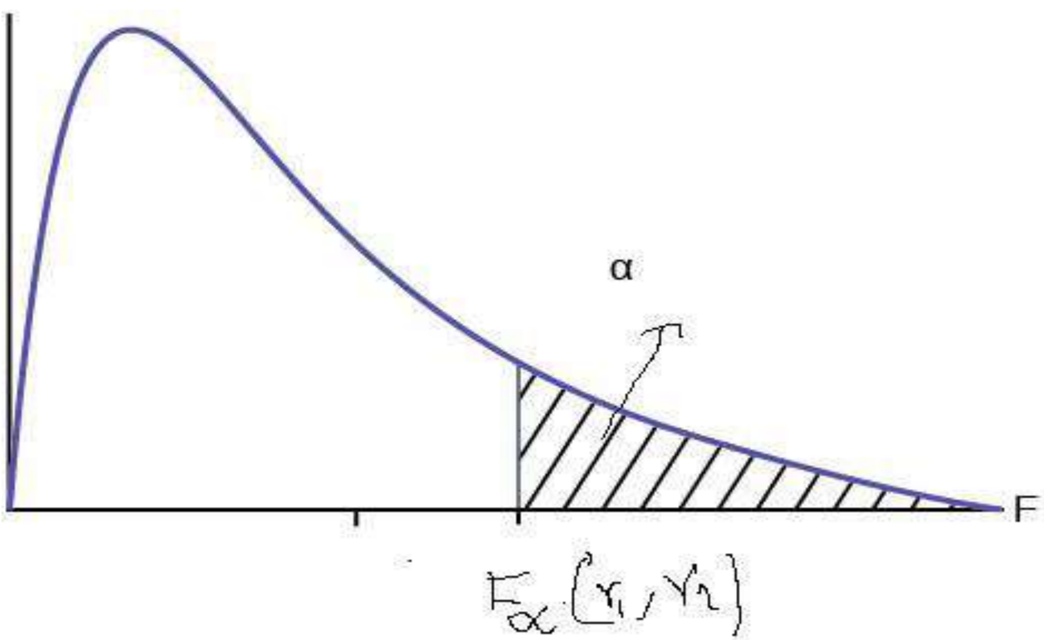
We have $F = \frac{(S_1^2)}{(S_2^2)}$, this determines whether the ratio of two sample variances S_1 and S_2 is too small or too large.

$F = \frac{\left(S_1^{\ 2}\right)}{\left(S_2^{\ 2}\right)}$ is always positive number. It practice , it is customary , to take the large sample variance as the numerator.

Properties of F-Distribution

1) F-distribution curve lies entirely in first quadrant.
2) The F-curve depends not only on the two parameters ν_1, ν_2 but also on the order in which they are stated.
3) $$F_{1-\alpha}(\nu_1, \nu_2) = \frac{1}{F_\alpha(\nu_2, \nu_1)}$$

   where $F_\alpha(\nu_1, \nu_2)$ is the value F with ν_1, ν_2 degrees of freedom such that the area under the F-distribution curve to the right of F_α is α


α
F
$F_\alpha(r_1, r_2)$

F - Distribution ($\alpha = 0.01$ in the Right Tail)

Denominator Degrees of Freedom df_2 \ df_1	Numerator Degrees of Freedom 1	2	3	4	5	6	7	8	9
1	4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5928.4	5981.1	6022.5
2	98.503	99.000	99.166	99.249	99.299	99.333	99.356	99.374	99.388
3	34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158
6	13.745	10.925	9.7795	9.1483	8.7459	8.4661	8.2600	8.1017	7.9761
7	12.246	9.5466	8.4513	7.8466	7.4604	7.1914	6.9928	6.8400	6.7188
8	11.259	8.6491	7.5910	7.0061	6.6318	6.3707	6.1776	6.0289	5.9106
9	10.561	8.0215	6.9919	6.4221	6.0569	5.8018	5.6129	5.4671	5.3511
10	10.044	7.5594	6.5523	5.9943	5.6363	5.3858	5.2001	5.0567	4.9424
11	9.6460	7.2057	6.2167	5.6683	5.3160	5.0692	4.8861	4.7445	4.6315
12	9.3302	6.9266	5.9525	5.4120	5.0643	4.8206	4.6395	4.4994	4.3875
13	9.0738	6.7010	5.7394	5.2053	4.8616	4.6204	4.4410	4.3021	4.1911
14	8.8616	6.5149	5.5639	5.0354	4.6950	4.4558	4.2779	4.1399	4.0297
15	8.6831	6.3589	5.4170	4.8932	4.5556	4.3183	4.1415	4.0045	3.8948
16	8.5310	6.2262	5.2922	4.7726	4.4374	4.2016	4.0259	3.8896	3.7804
17	8.3997	6.1121	5.1850	4.6690	4.3359	4.1015	3.9267	3.7910	3.6822
18	8.2854	6.0129	5.0919	4.5790	4.2479	4.0146	3.8406	3.7054	3.5971
19	8.1849	5.9259	5.0103	4.5003	4.1708	3.9386	3.7653	3.6305	3.5225
20	8.0960	5.8489	4.9382	4.4307	4.1027	3.8714	3.6987	3.5644	3.4567
21	8.0166	5.7804	4.8740	4.3688	4.0421	3.8117	3.6396	3.5056	3.3981
22	7.9454	5.7190	4.8166	4.3134	3.9880	3.7583	3.5867	3.4530	3.3458
23	7.8811	5.6637	4.7649	4.2636	3.9392	3.7102	3.5390	3.4057	3.2986
24	7.8229	5.6136	4.7181	4.2184	3.8951	3.6667	3.4959	3.3629	3.2560
25	7.7698	5.5680	4.6755	4.1774	3.8550	3.6272	3.4568	3.3239	3.2172
26	7.7213	5.5263	4.6366	4.1400	3.8183	3.5911	3.4210	3.2884	3.1818
27	7.6767	5.4881	4.6009	4.1056	3.7848	3.5580	3.3882	3.2558	3.1494
28	7.6356	5.4529	4.5681	4.0740	3.7539	3.5276	3.3581	3.2259	3.1195
29	7.5977	5.4204	4.5378	4.0449	3.7254	3.4995	3.3303	3.1982	3.0920
30	7.5625	5.3903	4.5097	4.0179	3.6990	3.4735	3.3045	3.1726	3.0665
40	7.3141	5.1785	4.3126	3.8283	3.5138	3.2910	3.1238	2.9930	2.8876
60	7.0771	4.9774	4.1259	3.6490	3.3389	3.1187	2.9530	2.8233	2.7185
120	6.8509	4.7865	3.9491	3.4795	3.1735	2.9559	2.7918	2.6629	2.5586
∞	6.6349	4.6052	3.7816	3.3192	3.0173	2.8020	2.6393	2.5113	2.4073

of the *F* Distribution

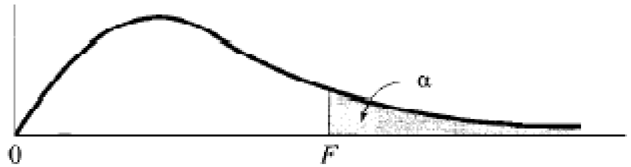


Table 1 $\alpha = 0.05$

Degrees of Freedom for Numerator

Degrees of Freedom for Denominator	1	2	3	4	5	6	7	8	9	10	15	20	25	30	40	50
1	161.4	199.5	215.8	224.8	230.0	233.8	236.5	238.6	240.1	242.1	245.2	248.4	248.9	250.5	250.8	252.6
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.44	19.46	19.47	19.48	19.48
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.63	8.62	8.59	8.58
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.77	5.75	5.72	5.70
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.52	4.50	4.46	4.44
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.83	3.81	3.77	3.75
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.40	3.38	3.34	3.32
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.11	3.08	3.04	3.02
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.89	2.86	2.83	2.80
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.73	2.70	2.66	2.64
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72	2.65	2.60	2.57	2.53	2.51
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.54	2.50	2.47	2.43	2.40
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53	2.46	2.41	2.38	2.34	2.31
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.39	2.34	2.31	2.27	2.24
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.28	2.25	2.20	2.18
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35	2.28	2.23	2.19	2.15	2.12
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31	2.23	2.18	2.15	2.10	2.08
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27	2.19	2.14	2.11	2.06	2.04
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.23	2.16	2.11	2.07	2.03	2.00
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	2.07	2.04	1.99	1.97
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.15	2.07	2.02	1.98	1.94	1.91
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.11	2.03	1.97	1.94	1.89	1.86
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.07	1.99	1.94	1.90	1.85	1.82
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.04	1.96	1.91	1.87	1.82	1.79
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93	1.88	1.84	1.79	1.76
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84	1.78	1.74	1.69	1.66
50	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Estimation of means and proportions

Estimation

Point Estimation

A point estimation of a parameter is a statistical estimation where the parameter is estimated by a single numerical value from sample data.

Point estimator

A point estimator is a statistic for estimating the population parameter θ and will be denoted by $\hat{\theta}$

Properties of Estimation

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

Unbiased Estimator

A point estimator $\hat{\theta}$ is said to be an unbiased estimator of the parameter θ if $E\left(\hat{\theta}\right)=\theta$

Most efficient estimator

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators of the same population parameter θ, and variance of the sampling distribution of estimators are $\left(\sigma_{\hat{\theta}1}\right)^2$, $\left(\sigma_{\hat{\theta}2}\right)^2$.

If $\left(\sigma_{\hat{\theta}1}\right)^2 < \left(\sigma_{\hat{\theta}2}\right)^2$, then $\hat{\theta}_1$ is more efficient estimator of θ than $\hat{\theta}_2$

Interval Estimate

Even the most efficient unbiased estimator cannot estimate the population parameter exactly. So instead of point estimate, it is preferable to determine an interval within which the value of the parameter. Such interval is called interval estimate.

An interval estimate of a population parameter θ

is an interval of the form $\hat{\theta}_L < \theta < \hat{\theta}_U$

From the sampling distribution of $\hat{\theta}$ we shall be able to find $\hat{\theta}_L$ and $\hat{\theta}_U$ such that

$P\left(\hat{\theta}_L < \theta < \hat{\theta}_U\right) = 1 - \alpha$ where $0 < \alpha < 1$

The interval $\hat{\theta}_L < \theta < \hat{\theta}_U$, computed from the selected sample , is called a $(1-\alpha)100\%$ confidence interval.

$(1-\alpha)$ is called degree of confidence, and the end points $\hat{\theta}_L$ and $\hat{\theta}_U$ are called the lower and upper limits.

Thus, when $(\alpha)=0.05$, we have 95% confidence interval, and when $(\alpha)=0.01$ we have 99% confidence interval.

Maximum error of estimate E for Large Samples:

Since the sample mean estimate very rarely equals to the mean of population μ , a point estimate is generally accompanied with a statement of error which gives difference between estimate and the quantity to be estimated , the estimator.

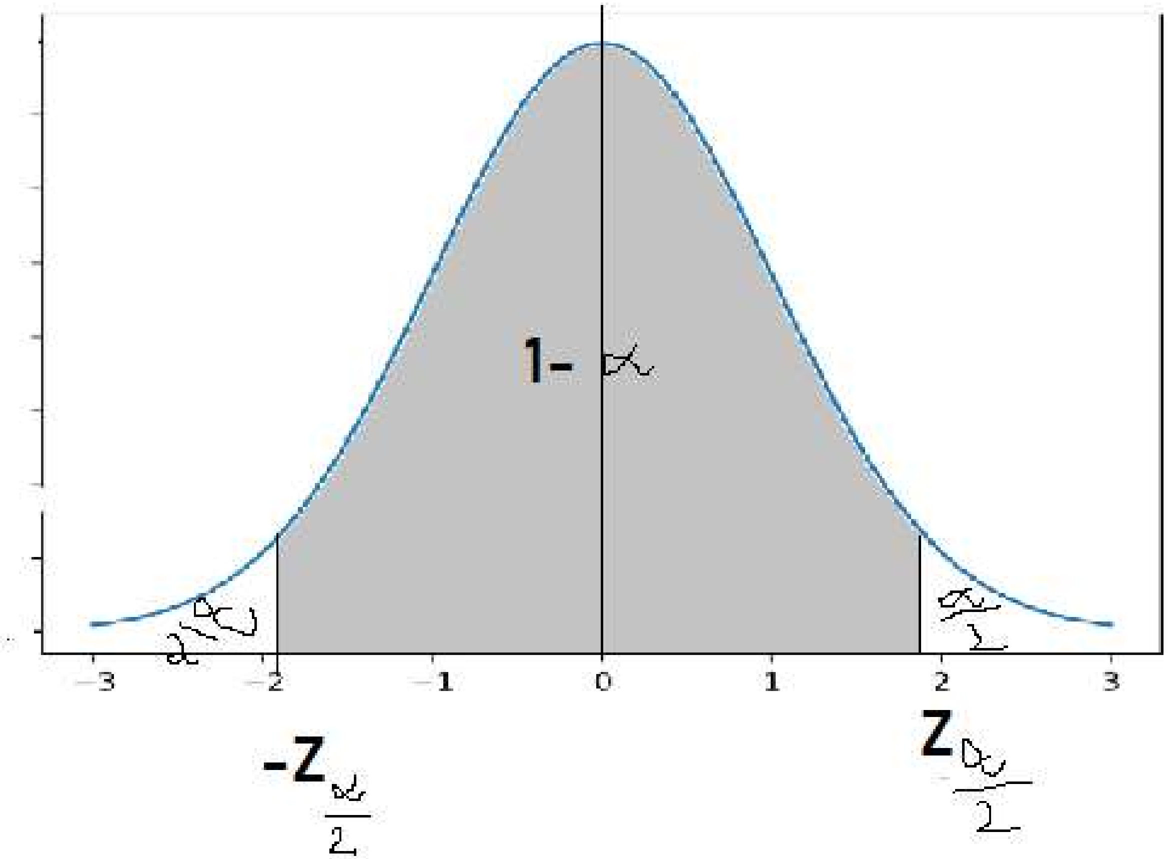
Thus error $\bar{x} - \mu$.

For large n, the random variable $\frac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$ is anormal variate approximately.

Then $P\left(-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}\right) = 1-\alpha$

where $Z = \frac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$

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



1- α
α/2
α/2
−3
−2
−1
0
1
2
3
$-Z_{\frac{\alpha}{2}}$
$Z_{\frac{\alpha}{2}}$

Multiplying each term in the inequality by $\frac{\sigma}{\sqrt{n}}$, and then subtracting $\bar{x}$ from each term and multiplying by -1

$$\therefore \quad P\left(\bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Confidence interval of μ, α known:

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

$$\left(\bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

Where $Z_{\frac{\alpha}{2}}$ is the Z-value leaving an area of $\frac{\alpha}{2}$ to the right.

So, the maximum error of estimate E with $(1-\alpha)$ probability is given by

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

Thus in the point estimation of population mean μ with sample mean $\bar{x}$ for a large random sample $(n \geq 30)$, one can assert with probability $(1-\alpha)$ that the error $|\bar{x}-\mu|$ will not exceed $Z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$.

Sample size:

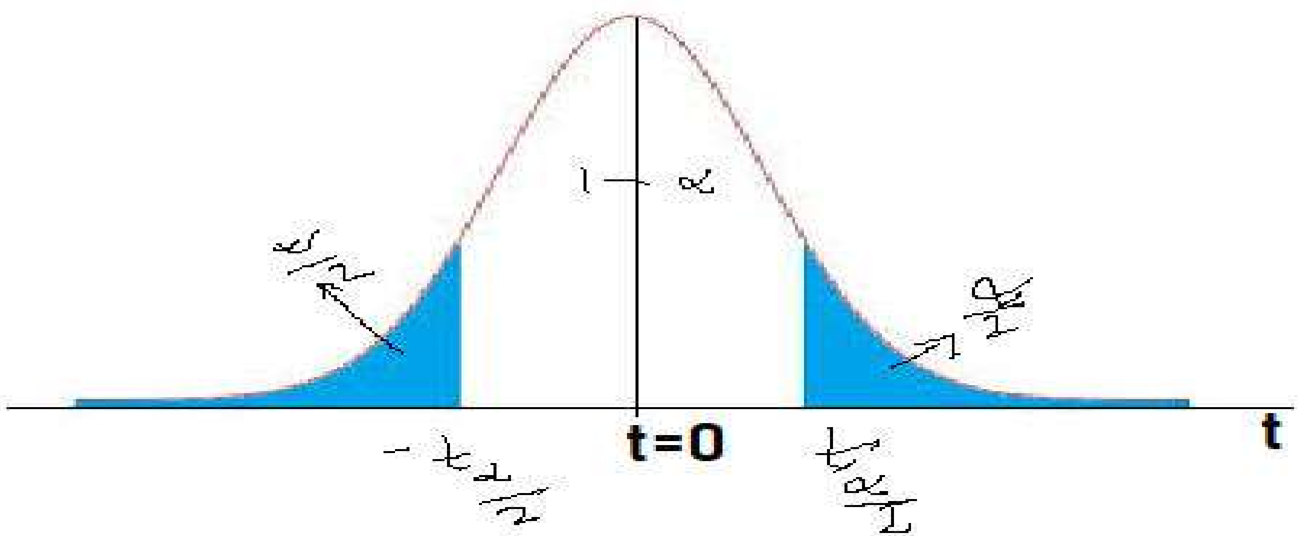
When σ , E are known , the sample size ‘n’ is given by $$n = \left(\frac{Z_{\frac{\alpha}{2}} \sigma}{E} \right)^2$$

Maximum Error of estimate for Small Sample:

When $n < 30$, small sample, we use S , the standard deviation of sample to determine E. When σ is unknown , t can be used to construct a confidence interval as μ.

$$P\left(-t_{\frac{\alpha}{2}} < T < t_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

Where $-t_{\frac{\alpha}{2}}$ is the t -value within (n-1) degrees of freedom



1 − α
α/2
α/2
−t α/2
t=0
t α/2
t

Substitute for T,

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

Multiplying each term in the inequality by $\frac{S}{\sqrt{n}}$ and the "n" subtracting $\overline{x}$ from each term and multiplying by -1 , we obtain .

$$\therefore \quad P\left(\overline{x} - t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} < \mu < \overline{x} + t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

Confidence interval of μ ; σ unknown

If $\bar{x}$ and S are the mean and standard deviation of a random sample from a normal population with unknown variance σ^2, a $(1-\alpha)100\%$ confidence interval for μ

$$\left(\bar{x}-t_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}<\mu<\bar{x}+t_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

Where $-t_{\frac{\alpha}{2}}$ is the t -value within (n-1) degrees of freedom ,leaving an area to the right.

So,the maximum error of estimate E with $(1-\alpha)$ probability is given by

$$E = t_{\frac{\alpha}{2}}\left(\frac{S}{\sqrt{n}}\right)$$

Single Proportion(Large sample)

Suppose a large sample of size 'n' is taken from a normal population. The confidence interval for population proportion "P" is given by

$$p - 3\sqrt{\frac{pq}{n}} < P < p + 3\sqrt{\frac{pq}{n}}$$

where "p" is the sample proportion and "P" is the population proportion .

PROBLEM:1

The mean and S.D of a population are 11,795 and 14054 respectively. What can one assert with 95% confidence about the maximum error if $\bar{x} = 11{,}795$ **and n =50. And also construct 95% confidence interval for the true mean.**

Solution:

Mean of Population $\mu = 11795$

S.D of population $\sigma = 14054$

$\bar{x} = 11795$

n = sample size =50

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

Maximum Error $E = Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} = 1.96 \times \dfrac{14054}{\sqrt{50}} = 3899$

Confidence interval = $\left(\bar{x} - Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right)$

= (11795-3899,11795+3899)

= (7896,15694)

PROBLEM:2

A random Sample of size 81 was taken whose variance is 20.25 and mean is 32 , construct 98% confidence interval.

Solution:

Given Sample mean $\bar{x} = 32$

$\sigma^2 = 20.25 \Rightarrow \sigma = 4.5$

n = 81

$Z_{\alpha/2}$ = 2.33 (for 98%)

We know that confidence interval =

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

$$Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 2.33\frac{4.5}{\sqrt{81}} = 1.165$$

Therefore Confidence interval =

(32-1.165, 32-1.165)

= (30.835,33.165)

PROBLEM3. It is desired to estimate the mean time of continuous use until an answering machine will first required service. If it can be assumed that $\sigma = 60$ days, how large a sample is needed so that one will be able to assert with 90% confidence that the sample mean is off by at most 10 days.

Solution:

Maximum error E= 10 hours

$\sigma = 60$ days

n = Sample size?

$Z_{\alpha/2}$ = 1.645 (for 90%)

$$n = \left[\frac{Z_{\alpha/2}\sigma}{E}\right]^2 = \left[\frac{1.645 \times 60}{10}\right]^2 = 72$$

PROBLEM: 4

A random sample of size 100 has a standard deviation of 5. What can you say about the maximum error with 95% confident.

Solution:

Given $\sigma = 5, n = 100$

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

We Know that

Maximum Error $E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1.96 \times \frac{5}{\sqrt{100}} = 0.98$

PROBLEM:5

The dean of a college wants to use the mean of a random sample to estimate the average amount of time students take to get from one class to the next and she wants to be able to assert with 99% confidence that the error is at most 0.25 minute . if it can be presumed from experience that $\sigma = 1.40$ minutes. How large a sample will she have to take?

Solution:

Maximum error E= 0.25 minutes

Standard deviation $\sigma = 1.40$

n = Sample size?

$Z_{\alpha/2}$ = 2.575 (for 99%)

$$n = \left[\frac{Z_{\alpha/2}.\sigma}{E}\right]^2 = \left[\frac{2.575 \times 1.40}{0.25}\right]^2 = 208$$

Sample size = 208

PROBLEM:6

It is desired to estimate the mean number of hours of continuous use until a certain computer will first required repairs. If it can be assumed that $\sigma = 48$ hours, how large a sample be needed so that one will be able to assert with 90% confident that the sample mean is off by at most 10 hours.

Solution:

Maximum error E= 10 hours

$\sigma = 48$ hours

n = Sample size?

$Z_{\alpha/2}$ = 1.645 (for 90%)

$$n = \left[\frac{Z_{\alpha/2}\sigma}{E} \right]^2 = \left[\frac{1.645 \times 48}{10} \right]^2 = 62.3$$

$$n = 62.3 \approx 62$$

PROBLEM:7

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 $\sigma^2 = 2.56$

Solution:

Here n = 64

The probability = 0.90

$\sigma^2 = 2.56 \quad \Rightarrow \sigma = \sqrt{2.56} = 1.6$

Confidence limit = 90%

$(1-\alpha)100 = 90 \Rightarrow 1-\alpha = 0.90$

$\alpha = 0.10, \alpha/2 = 0.05$

$Z_{\alpha/2}$ = 1-0.05 =0.95

area from $-\infty$ to $z_{\frac{\alpha}{2}}$ =0.5+0.45

corresponding to 0.45 ordinate is 1.645

$\Rightarrow Z_{\alpha/2} = 1.645$

Maximum Error $E = Z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}} = 1.645 \times \dfrac{1.6}{\sqrt{64}} =$ 0.329

PROBLEM:8

In a study of an automobile Insurance a random sample of 80 body repair costs had a mean of Rs. 472.36 and the S.D of Rs. 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 doesn't exceed Rs. 10?

Solution:

Size of a random sample = 80

The mean of random sample $\bar{x}$ = Rs. 472.36

Standard Deviation $\sigma = Rs.62.35$

Maximum error of estimation $E_{\max} = Rs.10$

$$E_{\max} = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\Longrightarrow \quad Z_{\alpha/2} = \frac{E_{\max}\sqrt{n}}{\sigma} = \frac{10\sqrt{80}}{62.35}$$

=1.43

$Z_{\alpha/2}$ = 0.4236 (from Normal Distribution table)

that is area from o to 1.43

there fore area from $-\infty$ ton1.43 is 0.5+0.4236=0.9236

($\frac{\alpha}{2}$ is the area , right of $Z_{\frac{\alpha}{2}}$ to ∞)

therefore $1-\alpha/2 = 0.9236$

$\alpha/2 = 1-0.9236$ =0.0764

$\alpha =$ 2(0.0764)=0.1528

(1-α)=1-0.1528=0.8472

Confidence level = $(1-\alpha)100\% = 84.72\%$

PROBLEM:9

The mean of random sample is an unbiased estimate of the mean of the population 3, 6, 9, 15, 27.

a) List of all possible samples of size 3 that 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 these $\bar{x}$ is equal to 12. Which is equal to the mean of population θ i.e $E(\bar{x}) = \theta$ i.e prove that $\bar{x}$ is an unbiased estimate of θ.

Solution:

(a) The possible samples of size 3 taken from 3, 6, 9, 15, 27 without replacement , are $5_{C_3} = 10$ samples i.e., (3,6,9) (3,6,15) (3,6,27) (6,9,15) (6,9,27) (3,9,15) (3,9,27) (9,15,27) (6,15,27) (3,15,27)

(b) Mean of the population $\mu = \dfrac{3 = 6 + 9 + 15 + 27}{5} = 12$

Mean of the samples = 6, 8, 12, 10, 14, 9, 13, 17 ,16, 15.

Probability assigned to each one is $\frac{1}{10}$ each

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

$$E(\bar{x}) = 6.\frac{1}{10} + 8.\frac{1}{10} + 12.\frac{1}{10} + 10.\frac{1}{10} + 14.\frac{1}{10} + 9.\frac{1}{10} + 13.\frac{1}{10} + 17.\frac{1}{10} + 16.\frac{1}{10} + 15.\frac{1}{10}$$

$= \frac{1}{10}$ x 120 = 12 = μ

$\therefore \quad E(\bar{x}) = \mu$

$\therefore \quad \bar{x}$ is an unbiased estimate of μ

I.e., the mean of a random sample is an unbiased estimator of the mean of the population.

PROBLEM:10

Find 95% confidence limits for mean of a normality distributed population from which the following sample was taken 15, 17, 10, 18, 16, 9, 7, 11, 13, 14.

Solution:

$$\bar{x} = \frac{15+17+10+18+16+9+7+11+13+14}{10} = 13$$

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

=

$$\frac{1}{9}\left[(15-13)^2+(17-13)^2+(10-13)^2+(18-13)^2+(16-13)^2+(9-13)^2+(7-13)^2+(11-13)^2+(13-13)^2+(14-13)^2\right]$$

$$= \frac{40}{3}$$

Since $t_{\alpha/2} = 2.26$ **(**$\alpha/2=$**0.05/2=0.025) with 10-**

degrees of freedom

We have $t_{\alpha/2}.\frac{\sqrt{S^2}}{\sqrt{n}} = 2.26.\frac{\sqrt{40}}{\sqrt{10}.\sqrt{3}} = 2.6$

$\therefore$ Confidence limits are ($\bar{x} - t_{\alpha/2}.\frac{S}{\sqrt{n}}$, $\bar{x} + t_{\alpha/2}.\frac{S}{\sqrt{n}}$)

= (13-2.6, 13+2.6)

= (10.4 , 15.6)

PROBLEM:11

Ten bearings made by a certain process have a mean diameter of 0.5060 cm with S.D of 0.0040 cm. Assuming that the data maybe taken as a random sample from a normal distribution, construct a 95% confidence interval for the actual average diameter of the bearings?

PROBLEM:12

A mong 900 people in a state 90 are found to be chapatti eaters. Construct 99% confidence interval for true proportion.

Sol. Given x= 90, n = 900

P = $\frac{x}{n} = \frac{90}{900} = 0.1$ and Q = 1- P = 0.9

Now $\sqrt{\frac{PQ}{n}} = \sqrt{\frac{0.1\times 0.9}{900}}$ =0.01

Confidence interval is (p - 3$\sqrt{\frac{PQ}{n}}$, p+ 3$\sqrt{\frac{PQ}{n}}$)

i.e (0.1 – 0.03, 0.1 + 0.03)

i.e (0.07,0.13)

PROBLEM:13.In a random sample of 160 workers exposed to a certain amount of radiation , 24 experienced some ill effects . construct a 99% confidence interval for the corresponding true percentage.

Sol. We have x = 24, n = 160 and P = $\frac{x}{n} = \frac{24}{160} = 0.15$, Q = 0.85.

Now $\sqrt{\frac{PQ}{n}}$ = $\sqrt{\frac{0.15\times0.85}{160}}$ = 0.028

Confidence interval is (p - $Z_{\propto/2}\sqrt{\frac{PQ}{n}}$, p + $Z_{\propto/2}\sqrt{\frac{PQ}{n}}$)

i.e (0.15 – 3 × 0.028 , 0.15 + 0.03)

i.e (0.065,0.234)

PROBLEM:14If 80 patients are treated with an antibiotic 59 got cured. Find a 99% confidence limits to true population of cure.

Sol. n = 80, x = 59 and p = $\frac{x}{n} = \frac{59}{80} = 0.7375$

Q = 1 – P = 0.2625

Now $\sqrt{\frac{PQ}{n}}$ = $\sqrt{\frac{0.7375 \times 0.2625}{80}}$ = 0.049

Confidence interval is (p - $Z_{\propto/2}\sqrt{\frac{PQ}{n}}$, p + $Z_{\propto/2}\sqrt{\frac{PQ}{n}}$)

(p - 3$\sqrt{\frac{PQ}{n}}$, p + 3$\sqrt{\frac{PQ}{n}}$)

i.e (0.7375 – 3 $\times$ 0.049 , 0.7375 + 3$\times$ 0.049)

i.e (0.59,0.88)

PROBLEM:15

Assuming that $\sigma = 20.0$,how large a random sample be taken to assert with probability 0.95 that the sample mean will not differ from the true mean by more than 3.0 points?

PROBLEM:16

A sample of 10 cam shafts intended for use in gasoline engines has an average eccentricity of 1.02 and a standard deviation of 0.044 inch. Assuming the data may be treated a random sample from a normal population, determine a 95% confidence interval for the actual mean eccentricity of the cam shaft?

PROBLEM:17

The mean & the standard deviation of a population are 11,795 & 14,054 respectively. If n = 50, find 95% confidence interval for the mean.

PROBLEM:18

A research worker wants to determine the average time it takes a mechanic to rotate the tyres of a car & he wants to be able to assert with 95% confidence that the mean of his sample is off by atmost 0.5 minutes. If he can presume from past experience that $\sigma = 1.6$minutes, how large a sample will have to take?

PROBLEM:19

A random sample of size 100 is taken from a population with $\sigma = 5.1$. Given that the sample mean is $\bar{x} = 21.6$. Construct a 95% confidence interval for the population mean μ.

PROBLEM:20. A random sample of 100 teachers in a large metropolitan area revealed a mean weekly salary of Rs. 487 with a S.D Rs.48. with what degree of confidence can be assert that the average weekly salary of all teachers in the metropolitan area is between 472 to 502?

PROBLEM:21

. In a random sample of 400 industrial accidents, it was found that 231 were due atleast partially to unsafe working conditions. Construct a 99% confidence interval for the corresponding true proportion.

SOLUTION

. We have x = 231, n = 400 and P = $\frac{x}{n} = \frac{231}{400} = 0.5775$ Q = 0.4225

Now $\sqrt{\frac{PQ}{n}}$ = $\sqrt{\frac{0.5775 \times 0.4225}{400}}$ = 0.0247

Confidence interval is (p - $Z_{\propto/2}\sqrt{\frac{PQ}{n}}$, p + $Z_{\propto/2}\sqrt{\frac{PQ}{n}}$)

(p - 3$\sqrt{\frac{PQ}{n}}$, p + 3$\sqrt{\frac{PQ}{n}}$)

i.e (0.5775 – 3 $\times$ 0.0247 , 0.5775 + 3$\times$ 0.0247)

i.e (0.5034,0.6516)

TEST OF SIGNICANCE OR SINGLE MEAN:

Suppose we want to test whether the given sample of size n has been drawn from a population with mean μ . we set up null hypothesis that there is no difference between $\bar{x}$ and μ is the sample mean.

The test statistics is, Z = $\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$, where $\sigma\ is\ S.D\ of\ population$

If the population S.D is not known,then use the statistics

Z = $\frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}}$, where $S\ is\ S.D\ of\ sample.$

Pb.1) According to the norms established for a mechanical aptitude test, persons who are 18 years old have an averge 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 the 0.01 level of significance.

Sol. Given n = 4, μ = 73.2, $\bar{x} = mean\ of\ the\ sample = 76.7$ and $\sigma = S.D\ of\ population = 8.6$

Null Hypothesis $\mathbf{H_0}$: μ = 73.2

Alternative Hypothesis $\mathbf{H_1}$: $\mu > 73.2$ (Right one tailed test)

Level of significance $\propto$ =0.01

Test statistics Z = $\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{76.7 - 73.2}{\frac{8.6}{\sqrt{4}}} = 0.814$

Since calculated I Z I= 0.814

Tabulated z at 1% level of significance is 2.33

Since calculated z < tabulated z, We accept the Null Hypothesis $\mathbf{H_0}$

That is, the mean value of population is 73.2.

Pb2) A sample of 64 students have a mean weight of 70kgs.Can this be regarded as a sample from a population with mean weight 56kgs and standard deviation 25kgs.

Sol. Given sample size n = 64,

$population\ mean\ \mu$ = 70kgs

Sample mean $\bar{x} = 56\ kgs$ and $\sigma = S.D\ of\ population = 25\ kgs$

Null Hypothesis $\mathbf{H_0}$: μ = 70 kgs,

i.e, the population mean μ = 70 kgs

Alternative Hypothesis $\mathbf{H_1}$: $\mu \neq\ \ 70kgs$ (two tailed test)

Level of significance ∝ =0.05

Test statistics Z $= \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{56 - 70}{\frac{25}{\sqrt{64}}} = 4.48$

Since calculated I Z I= 4.48

Tabulated z at 5% level of significance is 1.96

Since calculated z > tabulated z, We reject the Null Hypothesis $\mathbf{H_0}$

That is, the given sample cannot be regarded as from same population.

P

Pb3) An oceanographer wants to check whether the depth of the ocean in a certain region 57.4 fathoms,as had previously been recorded. What can he conclude at the 0.05 level of significance, if readings taken at 40 random locations in the given region yielded a mean of 59.1 athoms with a standard deviation of 5.2 fathoms

Sol. Given sample size n = 40,

$population\ mean\ \mu$ = 57.4

Sample mean $\bar{x} = 59.1$ and $\sigma = S.D\ of\ population = 5.2$

Null Hypothesis H_0 : μ = 57.4

i.e, the depth of ocean μ = 57.4 fathoms

Alternative Hypothesis H_1 : $\mu \neq 57.4$ (two tailed test)

Level of significance $\propto$ =0.05

Test statistics Z $= \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{59.1 - 57.4}{\frac{5.2}{\sqrt{40}}} = 2.067$

Since calculated I Z I= 2.067

Tabulated z at 5% level of significance is 1.96

Since calculated z > tabulated Z, We reject the Null Hypothesis **H_0**

Pb4) In a random sample of 60 workers, the average time taken by them to get to work is 33.8 minutes with a standard deviation of 6.1 minutes. Can we reject the null hypothesis

μ = 32.6 minutes in favour of alternative null hypothesis

μ > 32.6 at $\propto$ =0.05 level of significance.

Sol. Given sample size n = 60,

$population\ mean\ \mu$ = 32.6 minutes

Sample mean $\bar{x} = 33.8\ minutes$ and $\sigma = S.D\ of\ population = 6.1 minutes$

Null Hypothesis H_0 : μ = 32.6,

i.e, the population mean μ = 70 kgs

Alternative Hypothesis H_1 : $\mu > 32.6$ (two tailed test)

Level of significance $\propto$ =0.05

Test statistics $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{33.8 - 32.6}{\frac{6.1}{\sqrt{60}}} = 1.5238$

Since calculated I Z I= 1.5238

Tabulated z at 5% level of significance is 1.96

Since calculated z < tabulated z, We accept the Null Hypothesis $\mathbf{H_0}$

Pb5) A sample of 900 members has a mean of 3.4 cms and S.D 2.61 cms.Is this sample has been taken from a large population of mean 3.25 cm and S.D 2.61 cms. If the population is normal and its mean is unknown find the 95% fiducial limits of true mean.

Sol. Given sample size n = 900,

$population\ mean\ \mu$ = 3.25 cms

Sample mean $\bar{x} = 3.4\ cms$ and $\sigma = S.D\ of\ population = 2.61cms$

Null Hypothesis H_0 : μ = 3.25cms,

Alternative Hypothesis H_1 : $\mu \neq 3.25$ (two tailed test)

Level of significance $\propto$ =0.05

Test statistics $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{3.4 - 3.25}{\frac{2.61}{\sqrt{900}}} = 1.724$

Since calculated I Z I=1.724

Tabulated z at 5% level of significance is 1.96

Since calculated z < tabulated z, We accept the Null Hypothesis **H_0**

The sample has been drawn rom the population with mean μ = 3.25cms

95% confidence limits are given by

$$\boldsymbol{\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} = 3.4 \pm 1.96 \frac{2.61}{\sqrt{900}} = 3.4 \pm 0.1705}$$

i.e, 3.57 and 3.2295

Pb6) 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,

$population\ mean\ \mu$ = 38

Sample mean $\bar{x} = 40$ and $\sigma = S.D\ of\ population = 10$

Null Hypothesis H_0 : μ = 38,

Alternative Hypothesis H_1 : $\mu \neq 38$ (two tailed test)

Level of significance $\propto$ =0.05

Test statistics Z $= \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{40-38}{\frac{10}{\sqrt{400}}} = 4$

Since calculated I Z I=4

Tabulated z at 5% level of significance is 1.96

Since calculated z > tabulated z, We reect the Null Hypothesis **H_0**

The sample is not from the population whose mean μ = 38.

95% confidence limits are given by

$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right) = (40 - 1.96\frac{10}{\sqrt{400}}, 40 + 1.96\frac{10}{\sqrt{400}})$$

i.e, (39.02, 40.98)

Pb7) An ambulance service claims that it takes on the average less than 10 minutes to reach its destination in emergency calls. A sample of 36 calls has a mean of 11 minutes and the variance of 16 minutes. Test the claim at 0.05 level significance.

Sol. Given sample size n = 36,

$population\ mean\ \mu$ = 10

Sample mean $\bar{x} = 11$ and $variance\ \sigma^2 = 16, \sigma = S.D\ of\ population = \sqrt{16} = 4$

Null Hypothesis H_0 : μ = 10,

Alternative Hypothesis H_1 : $\mu > 10$ (right one tailed test)

L**evel of significance** $\propto$ **=0.05**

Test statistics Z $= \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{11 - 10}{\frac{4}{\sqrt{36}}} = 1.5$

Since calculated I Z I= 1.5

Tabulated z at 5% level of significance is 1.645

Since calculated z < tabulated z, We accept the Null Hypothesis $\mathbf{H_0}$

Pb8) it is claimed that a random sample of 49 tyres has a mean life of 15200km. This sample was drawn from a population whose mean is 15150 ms and a standard deviation of 1200 km. Test the significance at 0.05 level.

Sol. Given sample size n = 49,

$population\ mean\ \mu$ = 15150

Sample mean $\bar{x} = 15200$ and $\sigma = S.D\ of\ population = 1200$

Null Hypothesis H_0 : μ = 15150

Alternative Hypothesis H_1 : $\mu \neq \ 10$ (two tailed test)

Level of significance $\propto$ =0.05

Test statistics Z $= \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{15200\ -15150}{\frac{1200}{\sqrt{49}}} = 0.2917$

Since calculated I Z I= 0.2917

Tabulated z at 5% level of significance is 1.96

Since calculated z < tabulated z, We accept the Null Hypothesis **H_0**

TEST OF SIGNIFICANCE FOR DIFFERENCE OF MEANS

Let $\overline{x_1}$ be the mean of a sample of size n_1 from a population with mean μ_1 and variance σ_1^2.

Let be the mean of a sample of size n_2 from a population with mean μ_2 and variance σ_2^2 .

To test whether there is any significant diffence between $\overline{x_1}$ and $\overline{x_2}$, we have to use the statistics Z = $Z = \frac{\overline{x_1}-\overline{x_2}}{\sqrt{(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})}}$

Note: If the samples have been drawn from the same population then σ_1^2 = σ_2^2 $= \sigma^2$

Z = $\frac{\overline{x_1}-\overline{x_2}}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}}$

Pb1) 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. let μ_1 and μ_2 be the means of the two populations.

Given n_1 = 1000, n_2 = 2000 and $\overline{x_1}$ =67.5inches, $\overline{x_2}$ = 68 inches, $\sigma = S.D\ of\ population =$ 2.5 inches

Null Hypothesis H$_0$: $\mu_1 = \mu_2$

i.e, the samples have been drawn from the same population of S.D 2.5 inches.

Alternative Hypothesis H$_1$: $\mu_1 \neq \mu_2$

Level of significance $\propto$ **=0.05**

The Test Statistics : Z = $\frac{\overline{x_1}-\overline{x_2}}{\sqrt{\sigma^2(\frac{1}{n_1}+\frac{1}{n_2})}} = \frac{67.5-68}{\sqrt{2.5^2(\frac{1}{1000}+\frac{1}{2000})}}$ = -5.16

Since calculated I Z I= 5.16

Tabulated z at 5% level of significance is 1.96

Since calculated z > tabulated z, We rejectt the Null Hypothesis **H$_0$**

We conclude that the samples are not drawn from the same population of S.D 2.5 inches.

Pb2) The mean yield of wheat from a district A was 210 pounds with S.D 10 pounds per acre from a sample of 100 plots. In another district the mean yield was 220 pounds with S.D 12 pounds from a samp le of 150 plots. Asssumeing that the S.D of yield in the entire state was 11 pounds, test whether there is any significant difference between the mean yield of crop in the two districts.

Sol. let μ_1 and μ_2 be the means of the two populations.

Given n_1 = 100, n_2 = 150 and $\overline{x_1}$ = 210, $\overline{x_2}$ = 200, $\sigma = S.D\ of\ population = 11$

Null Hypothesis H$_0$: $\mu_1 = \mu_2$

i.e, there is no significant difference between μ_1 and μ_2 .

Alternative Hypothesis H$_1$: $\mu_1 \neq \mu_2$

Level of significance $\propto$ =0.05

The Test Statistics : $Z = \frac{\overline{x_1}-\overline{x_2}}{\sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{210-200}{\sqrt{11^2\left(\frac{1}{100}+\frac{1}{150}\right)}} = 7.04178$

Since calculated I Z I= 7.04178

Tabulated z at 5% level of significance is 1.96

Since calculated z > tabulated z, We rejected the Null Hypothesis **H$_0$**

We conclude that there is a significant difference between the mean yield of crops in the two districts.

Pb3) In a survey of buying habits, 400 women shoppers are chosen at random in super market A located in a certain section of the city. Their average weekly food expenditure is Rs 250 with a S.D of Rs. 40. For 400 women shoppers chosen at random in super market B in another section of the city, the average weely food expenditure is Rs.220 with a S.D of Rs. 55. Test at 1% level of significance whether the average weekly food expenditure of the two populations of shoppers are equal.

Sol. . let μ_1 and μ_2 be the means of the two populations.

Given n_1 = 400, n_2 = 400 and $\overline{x_1}$ = 250, $\overline{x_2}$ = 220, $S_1 = 40, S_2 = 55$

Null Hypothesis $\mathbf{H_0}$: $\mu_1 = \mu_2$

i.e, there is no significant difference between μ_1 and μ_2 .

Alternative Hypothesis $\mathbf{H_1}$: $\mu_1 \neq \mu_2$

Level of significance $\propto$ =0.01

The Test Statistics : Z = $\dfrac{\overline{x_1}-\overline{x_2}}{\sqrt{\left(\frac{{S_1}^2}{n_1}+\frac{{S_2}^2}{n_2}\right)}}$ = $\dfrac{250-220}{\sqrt{\left(\frac{40^2}{400}+\frac{55^2}{400}\right)}}$ = 8.82

Since calculated I Z I= 8.82

Tabulated z at 1% level of significance is 2.58

Since calculated z > tabulated z, We rejected the Null Hypothesis $\mathbf{H_0}$

Pb4) A sample of students were drawn from two universities and from their weights in kilograms, mean and standard deviations are calculated and shown below. Make a large sample test to test significance of the difference between the means.

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

Sol. Given n_1 = 400, n_2 = 100 and $\overline{x_1}$ = 55, $\overline{x_2}$ = 57, $S_1 = 10, S_2 = 15$

Null Hypothesis H_0: $\overline{x_1}$ = $\overline{x_2}$

i.e, there is no significant difference .

Alternative Hypothesis H_1 : $\overline{x_1} \neq \overline{x_2}$

Level of significance $\propto$ =0.05

The Test Statistics : Z = $\dfrac{\overline{x_1}-\overline{x_2}}{\sqrt{\left(\frac{{S_1}^2}{n_1}+\frac{{S_2}^2}{n_2}\right)}}$ = $\dfrac{55-57}{\sqrt{\left(\frac{100}{400}+\frac{225}{100}\right)}}$ = -1.26

Since calculated I Z I= 1.26

Tabulated z at 5% level of significance is 1.96

Since calculated z < tabulated z, We accept the Null Hypothesis **H_0**

STUDENTS "t" TEST for Single Mean

The statistic

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{(n-1)}}\right)}$$ with $(n-1)$ degrees of freedom

$\bar{x} = \frac{\sum x_i}{n}$ sample mean

μ = population mean

n = sample size

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

PROBLEM:1

A mechanist is making engine parts with axle diameters of 0.700 inch. A random sample of 10 parts shows a mean diameter of 0.742 inch with standard deviation of 0.040 inch . Compute the statistic you would use to test whether the work is meeting the specification.

Solution:

$n = 10 < 30$ sample size is small

$\bar{x} = 0.742$

$\mu = 0.700$

S.D= 0.040

Null Hypothesis: $H_0 : \mu = 0.700$

Alternative Hypothesis: $H_1 : \mu \neq 0.700$

Level of significance $\alpha = 0.05$

Critical region: $t > t_{0.05}$

Test statistic = t = $\dfrac{\bar{x} - \mu}{\left(\dfrac{s}{\sqrt{(n-1)}}\right)}$ with (n-1) degrees of freedom

$$t = \frac{0.742 - 0.700}{\left(\dfrac{0.040}{\sqrt{(10-1)}}\right)} = 3.15$$

The table value of 't' are 5% level with 9 degrees of freedom

$t_{0.05} = 2.26$

Since calculated value of 't ' > tabulated value of 't' , therefore H_0 is rejected.

PROBLEM:2

A machine is designed to produce insulating washers for electrical devices of average thickness of 0.025 cm. A random sample of 10 washers was found to have a thickness of 0.024 cm with a S.D of 0.002 cm . Test the significance of the deviation . Value of ‘t’ for 9 degrees of freedom at 5% level is 2.262.

Solution:

$n = 10 < 30$ sample size is small

$\bar{x} = 0.024$

$\mu = 0.025$

S.D= 0.002

Null Hypothesis: $H_0 : \mu = 0.025$

Alternative Hypothesis: $H_1 : \mu \neq 0.025$

Level of significance $\alpha = 0.05$

Critical region: $t > t_{0.05}$

Test statistic = t = $\dfrac{\bar{x} - \mu}{\left(\dfrac{s}{\sqrt{(n-1)}}\right)}$ with (n-1) degrees of freedom

$$t = \frac{0.024 - 0.025}{\left(\dfrac{0.002}{\sqrt{(10-1)}}\right)} = -1.5$$

$$|t| = 1.5$$

The table value of 't' are 5% level with 9 degrees of freedom $t_{0.05} = 2.26$

Since calculated value of 't ' < tabulated value of 't' , therefore H_0 is accepted.

PROBLEM: 3

A random sample from a company's very expensive iles shows that the orders for a certain kind of machinery were filled , respectively in 10,12, 19, 14, 15 ,18, 11, 13 days. Use the level of significance $\alpha = 0.01$ to test the claim that on the average such orders are filled in 10.5 days. Assume normality.

Solution:

$$n = 8$$

$$\bar{x} = \frac{1}{8}(10+12+19+14+15+18+11+13) = \frac{112}{8} = 14$$

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

$$= \frac{1}{7}\left[(10-14)^2 + (12-14)^2 + (19-14)^2 + ----\right]$$

=10.286

$s = \sqrt{10.286}$ =3.207

Null Hypothesis: $H_0 : \mu = 10.5$

Alternative Hypothesis: $H_1 : \mu \neq 10.5$

Level of significance $\alpha = 0.01$

Critical region: $t > t_{0.01}$

Test statistic = t = $\dfrac{\bar{x} - \mu}{\left(\dfrac{s}{\sqrt{(n-1)}}\right)}$ with (n-1) degrees of freedom

$$t = \frac{14 - 10.5}{\left(\frac{3.207}{\sqrt{(8-1)}}\right)} = 3.087$$

The table value of 't' are 5% level with 9 degrees of freedom

$t_{0.01} = 2.998$

Since calculated value of 't ' > tabulated value of 't' , therefore

H_0 is rejected.

EXERCISE

1. The height of 10 males of a given locality are found to be 70, 67, 62, 68, 61, 68, 70, 64, 64, 66 inches. Is it reasonable to belie that the average height is greater than 64 inches ? Test at 5% significance level assuming that for 9 degrees of freedom
2. A random sample of six steel beams has a mean compressive strength of 58.392p.s.i with a standard deviation of 648 p.s.i . Use this information and the level of significance $\alpha = 0.05$

   To test whether the true average compressive strength of the steel from which this sample came is 58,000 p.s.i

Student's 't' test for difference of means

To test the significant difference between two means $\bar{x}_1$ and $\bar{x}_2$ of samples of sizes n_1 and n_2

Statistic $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s\left(\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}\right)}$ degrees of freedom (n_1+n_2-2)

$$s^2 = \frac{\sum(x_1 - \bar{x}_1)^2 + \sum(x_2 - \bar{x}_2)^2}{(n_1 + n_2 - 2)}$$

OR

$$s^2 = \frac{1}{n_1 + n_2 - 2}\left(n_1 s_1{}^2 + n_2 s_2{}^2\right)$$

PROBLEM:1

Samples of two types of electric light bulbs were tested for length of life and following data were obtained

TYPE-1	TYPE-2
Sample number $n_1 = 8$	$n_2 = 7$
Sample means $\bar{x}_1 = 1234$ hours	$\bar{x}_2 = 1036$ hours
Sample S.D $s_1 = 36$ hrs	$s_2 = 40$ hrs

Is the difference in the means sufficient to warrant tat type-1 is superior to type-2 regarding length of life.

SOLUTION:

$n_1 = 8$, $n_2 = 7$

$$\bar{x} = \frac{1}{8}(11+11+13+11+15+9+12+14) = 12$$

$$\bar{y} = \frac{1}{7}(9+11+10+13+9+8+10) = 10$$

$$= s^2 = \frac{\sum(x_1 - \bar{x}_1)^2 + \sum(x_2 - \bar{x}_2)^2}{(n_1 + n_2 - 2)}$$

$$s^2 = \frac{1}{8+7-2}(26+16) = 3.23$$

S=1.8

Null Hypothesis: $H_0 : \mu_1 = \mu_2$

Alternative Hypothesis: $H_1 : \mu_1 \neq \mu_2$ (two tailed test)

Level of significance $\alpha = 0.05$

Critical region: $t > t_{0.05}$

Statistic $t = \dfrac{\bar{x} - \bar{y}}{s\left(\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}\right)}$ degrees of freedom (n_1+n_2-2)

$$t = \frac{12-10}{1.8\sqrt{\frac{1}{8}+\frac{1}{7}}} = 2.15$$

Degrees of freedom 8+7-2=13

Tabulated value of t for 13 degrees of freedom at 5% level of significance is 2.16(twotailed test)

Since calculated value of t < tabulated t we accept the null hypothesis

EXERCISE

1. The means of two random samples of sizes 9 and 7 are 196.42 and 198.82 respectively. The sum of the squares of the deviations from the mean are 26.94 and 18.73 respectively. Can the sample be considered to have been drawn from the same normal population
2. Two horses A and B were tested according to the time to run a particular track with the following results.

HorseA	28	30	32	33	33	29	34
HorseB	29	30	30	24	27	29	

Test whether the two horses have the same running capacity .

PAIRED SAMPLE t-test

Paired observations arise in a very special experimental situation where each homogeneous experimental unit receives both population conditions. As a result , each experimental unit has a pair of observations, one for each population . Thus the paired observations are on the same unit or matching units.

Example

To test the effectiveness of "insulin" some 10diabetic patients sugar level in blood is measured "before" and "after" the insulin is injected. Here the individual diabetic patient is the experimental unit and the two populations are blood sugar level 'before '' and 'after' the insulin is injected.

So for each observation is one sample, there is a corresponding observation in the other sample pertaining to the same character. Thus the two samples are not independent. Paired t-test is applied for 'n' paired observations(which are dependent) by taking the (signed) differences d_1 ,d_2, ………d_n of the paired data. To test whether the differences 'd' from a random sample from a population with $\mu_D = d_0$

Use large sample test or one sample test when sample is small (the one sample t-test in this case is known as the paired – sample t-test) .

The test statistic is $\dfrac{\bar{d}-\mu_d}{\left(\dfrac{s_d}{\sqrt{n}}\right)}$ with (n-1) degrees of freedom

and $\bar{d}$ and $s^2{}_d$ are the mean and variance of the differences d_1 ,d_2, ...d_n

PROBLEM:1

In a study of usefulness of yoga in weight reduction , a random sample of 16 persons undergoing yoga were examined of their weight before (without) and after (with) yoga with the following results;

Weight before	209	178	169	212	180	192	158	180	170	153	183	165	201	179	243	144
Weight after	196	171	170	207	177	190	159	180	164	152	179	162	199	173	231	140

Test whether yoga is useful in weight reduction at 0.01 level of significance.

SOLUTION

Let μ be the mean of population of differences,

1. Null Hypothesis: $\mu = 0$ (i.e ., not use ful)
2. Alternative Hypothesis: $\mu > 0$ (i.e ., yoga is useful in weight reduction)
3. Level of significance: $\alpha = 0.01$
4. Critical region : Right tailed test

   Reject Null Hypothesis if $t > t_{0.01}$ with 16-1 =15 degrees of freedom.

   From table $t_{0.01} = 2.602$

1. Calculation : differences d_i's are

13,7,-1,5,3,2,-1,0,6,1,4,3,2,6,12,4

$\bar{x} =$ mean of differences of sampled data = $\dfrac{66}{16} = 4.125$

$$s^2 = \frac{247.73}{15} = 16.516$$

$$s = 4.064$$

$$t = \frac{\bar{x} - \mu_0}{\left(\dfrac{s}{\sqrt{n}}\right)} = \frac{4.125 - 0}{\left(\dfrac{4.064}{\sqrt{16}}\right)} = 4.06$$

2. Decision: Reject Null Hypothesis since $t = 4.06 > 2.602$

i.e ., Yoga is useful in weight reduction

EXERCISE

1.The average weekly losses of man-hours due to strikes in an institute before and after a disciplinary program was implemented are as follows

before	45	73	46	124	33	57	83	34	26	17
after	36	60	44	119	35	51	77	29	24	11

Is there reason to believe that the disciplinary program is effective at 0.05 level of significance.

2.The blood pressure (B.P) of 5 women before and after intake of certain drug are given below

Before	110	120	125	132	125
After	120	118	125	136	121

Test at 0.01 level of significance whether there is significant change in B.P

F- Test

To test whether there is any significant difference between two estimates of population variance we use f-test

In this case

Null Hypothesis H_0 : $\sigma_1^2 = \sigma_2^2$ (i.e.,population variance are same)

Test statistic F= $\dfrac{S_1^2}{S_2^2}$

Where $S_1^2 = \dfrac{\sum(x-\bar{x})^2}{n_1 - 1}$ (n_1 first sample size)

$S_2^2 = \dfrac{\sum(y-\bar{y})^2}{n_2 - 1}$ (n_2 second sample size)

And $S_1^{\ 2} > S_2^{\ 2}$

The degrees of freedom are $\upsilon_1 = n_1 - 1 \quad , \upsilon_2 = n_2 - 1$

Note: take greater of the variance $S_1^{\ 2}$ or $S_2^{\ 2}$ in the numerator and adjust for the degree of freedom accordingly

i.e., F = $\dfrac{greater\ \mathrm{var}\,iance}{smaller\ \mathrm{var}\,iance}$

Note: If sample variance S^2 is given ,obtain population variance σ^2 by using the relation

$n\sigma^2 = (n-1)S^2$ and vice –versa

Problem1:

In one sample of 8 observations the sum of the squares of deviations of the sample values from the sample mean was 84.4 and in the other sample of 10 observations it is 102.6. Test whether this difference is significant at 5% level.

Solution

$n_1 = 8$, $n_2 = 10$

$$S_1^{\ 2} = \frac{\sum (x - \bar{x})^2}{n_1 - 1} = \frac{84.4}{7} = 12.057$$

$$S_2^{\ 2} = \frac{\sum (y - \bar{y})^2}{n_2 - 1} = \frac{102.6}{9} = 11.4$$

Null Hypothesis H_0 : $\sigma_1^{\ 2} = \sigma_2^{\ 2}$ (i.e.,population variance are same)

Test statistic $F = \dfrac{S_1^{\ 2}}{S_2^{\ 2}} = \dfrac{12.057}{11.4} = 1.057$ (calculated value)

Tabulated value of F at 5% level for (7,9) degrees of freedom is 3.29

i.e., $F_{0.05}(7,9)=3.29$

Since calculated F < tabulated F , we accept the null Hypothesis.

Problem:2

The time taken by workers in performing a job by method 1 and method 2 is given below:

Method 1	20	16	26	27	23	22	-
Method 2	27	33	42	35	32	34	38

Do the data show that the variances of time distribution from population from which these samples are drawn do not differ significant?

Solution

$n_1 = 6 \quad , \quad n_2 = 7$

$$\bar{x} = \frac{\sum x}{n_1} = \frac{134}{6} = 22.3$$

$$\bar{y} = \frac{\sum y}{n_2} = \frac{241}{7} = 34.4$$

Calculation of sample variances

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
20	-2.3	5.29	27	-7.4	54.76
16	-6.3	39.69	33	-1.4	1.96
26	3.7	13.69	42	7.6	57.76
27	4.7	22.09	35	0.6	0.36
23	0.7	0.49	32	-2.4	5.76
22	-0.3	0.09	34	-0.4	0.16
			38	3.6	12.96
134		81.34	241		133.72

$$S_1{}^2 = \frac{\sum(x-\bar{x})^2}{n_1 - 1} = \frac{81.34}{5} = 16.26$$

$$S_2{}^2 = \frac{\sum(y-\bar{y})^2}{n_2 - 1} = \frac{133.72}{6} = 22.29$$

Null Hypothesis H_0 : $\sigma_1{}^2 = \sigma_2{}^2$ (i.e.,population variance are same)

Since $S_2{}^2 > S_1{}^2$

Test statistic F = $\frac{S_2{}^2}{S_1{}^2} = \frac{22.29}{16.268} = 1.3699$ (calculated value)

Tabulated value of F at 5% level for (6,5) degrees of freedom is 3.29

i.e., $F_{0.05}(6,5)=4.95$

Since calculated F < tabulated F , we accept the null Hypothesis.

EXERCISE

1. In one sample of 10 observations from a normal population, the sum of the squares of the deviations of the sample values from the sample mean is 102.4 and in another sample of 12 observations from another normal population , the sum of squares of the deviations of the sample values from the sample mean is 120.5 . Examine whether the two normal populations have the same variance.
2. Two random samples gave the following results:

Sample	Size	Sample mean	Sum of Squares of Deviations from the mean
1	10	15	90
2	12	14	108

MAXIMUM LIKELIHOOD ESTIMATION

Maximum Likelihood Estimation is a systematic technique for estimating parameters in a probability model from a data sample. Suppose a sample $x_1, ..., x_n$ has been obtained from a probability model specified by mass or density function $f_X(x;\theta)$ depending on parameter(s) θ lying in parameter space Θ. The **maximum likelihood estimate** or **m.l.e.** is produced as follows;

STEP 1 Write down the **likelihood function**, $L(\theta)$, where

$$L(\theta) = \prod_{i=1}^{n} f_X(x_i;\theta)$$

that is, the product of the n mass/density function terms (where the ith term is the mass/density function evaluated at x_i) viewed as a function of θ.

STEP 2 Take the natural log of the likelihood, collect terms involving θ.

STEP 3 Find the value of $\theta \in \Theta$, $\widehat{\theta}$, for which $\log L(\theta)$ is maximized, for example by differentiation. If θ is a single parameter, find $\widehat{\theta}$ by solving

$$\frac{d}{d\theta}\{\log L(\theta)\} = 0$$

in the parameter space Θ. If θ is vector-valued, say $\theta = (\theta_1, ..., \theta_k)$, then find $\widehat{\theta} = (\widehat{\theta}_1, ..., \widehat{\theta}_k)$ by simultaneously solving the k equations given by

$$\frac{\partial}{\partial\theta_j}\{\log L(\theta)\} = 0 \qquad j = 1, ..., k$$

in parameter space Θ. Note that, if parameter space Θ is a bounded interval, then the maximum likelihood estimate may lie on the boundary of Θ.

STEP 4 Check that the estimate $\widehat{\theta}$ obtained in STEP 3 truly corresponds to a maximum in the (log) likelihood function by inspecting the second derivative of $\log L(\theta)$ with respect to θ. In the single parameter case, if the second derivative of the log-likelihood is negative at $\theta = \widehat{\theta}$, then $\widehat{\theta}$ is confirmed as the m.l.e. of θ (other techniques may be used to verify that the likelihood is maximized at $\widehat{\theta}$).

EXAMPLE Suppose a sample $x_1, ..., x_n$ is modelled by a Poisson distribution with parameter denoted λ, so that

$$f_X(x;\theta) \equiv f_X(x;\lambda) = \frac{\lambda^x}{x!}e^{-\lambda} \qquad x = 0, 1, 2, ...$$

for some $\lambda > 0$. To estimate λ by maximum likelihood, proceed as follows.

STEP 1 Calculate the likelihood function $L(\lambda)$.

$$L(\lambda) = \prod_{i=1}^{n} f_X(x_i;\lambda) = \prod_{i=1}^{n} \left\{ \frac{\lambda^{x_i}}{x_i!}e^{-\lambda} \right\} = \frac{\lambda^{x_1+...+x_n}}{x_1!....x_n!}e^{-n\lambda}$$

for $\lambda \in \Theta = R^+$.

STEP 2 Calculate the log-likelihood $\log L(\lambda)$.

$$\log L(\lambda) = \sum_{i=1}^{n} x_i \log \lambda - n\lambda - \sum_{i=1}^{n} \log(x_i!)$$

STEP 3 Differentiate $\log L(\lambda)$ with respect to λ, and equate the derivative to zero to find the m.l.e..

$$\frac{d}{d\lambda}\{\log L(\lambda)\} = \sum_{i=1}^{n} \frac{x_i}{\lambda} - n = 0 \Rightarrow \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

Thus the maximum likelihood estimate of λ is $\hat{\lambda} = \bar{x}$

STEP 4 Check that the second derivative of $\log L(\lambda)$ with respect to λ is negative at $\lambda = \hat{\lambda}$.

$$\frac{d^2}{d\lambda^2}\{\log L(\lambda)\} = -\frac{1}{\lambda^2}\sum_{i=1}^{n} x_i < 0 \quad \text{at } \lambda = \hat{\lambda}$$

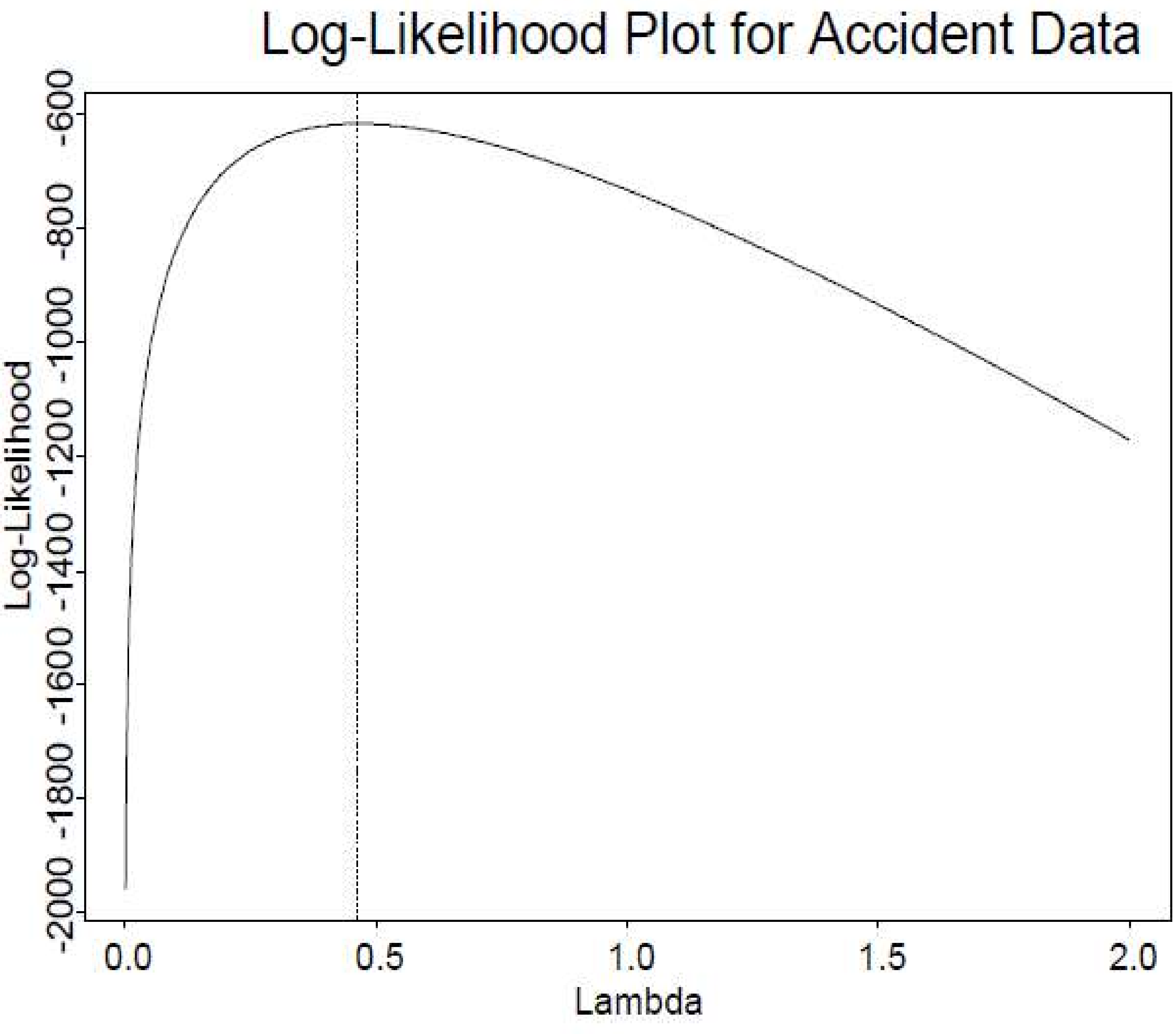
EXAMPLE: The following data are the observed frequencies of occurrence of domestic accidents: we have $n = 647$ data as follows

Number of accidents	Frequency
0	447
1	132
2	42
3	21
4	3
5	2

.

The estimate of λ if a Poisson model is assumed is

$$\widehat{\lambda}_{ML} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{(447 \times 0) + (132 \times 1) + (42 \times 2) + (21 \times 3) + (3 \times 4) + (2 \times 5)}{647} = 0.465$$



Log-Likelihood Plot for Accident Data
Log-Likelihood
-2000
-1800
-1600
-1400
-1200
-1000
-800
-600
0.0
0.5
1.0
1.5
2.0
Lambda

TESTING OF HYPOTHESIS OF SINGLE PROPORTION

Prob1). A manufacturer clamed 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 sample size n = 200

Number of pieces confirming to specification = 200 – 18 = 182

P = Proportion of pieces confirming to specifications = $\frac{182}{200} = 0.91$

P = population proportion = $\frac{95}{100} = 0.95$

Null Hypothesis H_0 : The proportion of pieces confirming to specifications .

i.e P = 0.95

Alternative Hypothesis H_1 : P < 0.95 (left one tail test)

The test statistic Z = $\frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.91-0.95}{\sqrt{\frac{0.95\times0.05}{200}}} = \frac{-0.04}{0.0154} = -2.59$

Since alternative hypothesis is left tailed, the tabulated value of Z at 5% level of significance is 1.645.

Since calculated value of IzI = 2.6 is greater than 1.645, we reject the null hypothesis **H_0 at 5% level of significance. Hence the manufactures claim is rejected.**

Pb2) In a sample of 1000 people in Karnataka 540 are rice eaters and the rest are wheat eaters. Can we assume that both rice and wheat are equally popular in this state at 1% level of significance

Sol. Give n = 1000

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

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

Q = 0.5

Null Hypothesis $\mathbf{H_0}$: Both rice and wheat are equally popular in the state.i.e P = 0.5

Alternative Hypothesi $\mathbf{H_1}$: P $\neq 0.5$ Test is Two tailed test

Test statistics Z = $\frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.54-0.5}{\sqrt{\frac{0.5\times0.5}{1000}}} = 2.532$

The calculated value of z = 2.532

The tabulated value of Z at 1% level of significance for two tailed test is 2.58.

Since calculated Z < tabulated Z we accept null hypothesis. i.e. both rice and wheat are equally popular in the state at 1% level of significance.

Pb3) 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. Given n = 600

Number of smokers = 325

P = sample proportion of smokers = $\frac{325}{600} = 0.5417$

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

Q = 0.5

Null Hypothesis $\mathbf{H_0}$: P = 0.5. I.e . the number of smokers and non smokers are equal in the city.

Alternative Hypothesis $\mathbf{H_1}$: P > 0.5 .test is right one tailed test.

Test statistics Z = $\frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.5417-0.5}{\sqrt{\frac{0.5\times0.5}{1000}}} = 2.04$

The calculated value of z = 2.04

The tabulated value of Z at 5% level of significance for right one tailed test is 1.645.

Since calculated Z > tabulated Z we reject null hypothesis. i.e.the majority of men in the city are smokers.

Pb4) 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?

Sol. Given n = 9000

P = proportion of successes of getting 3 or 4 in 9000 throws = $\frac{3220}{9000} = 0.3578$

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

Q = 1 – P = 1- 0.3333 = 0.6667

Null Hypothesis $\mathbf{H_0}$: the die is unbiased.

Alternative Hypothesis $\mathbf{H_1}$: P $\neq \frac{1}{3}$ test is two tailed.

Test statistics Z = $\frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.3578-0.3333}{\sqrt{\frac{0.3333 \times 0.6667}{9000}}} = 4.94$

Calculated Z = 4.94

Since Z > 3, the null hypothesis $\mathbf{H_0}$ is **rejected .we conclude that the die is biased.**

Pb5) In a random sample of 125 cola drinkers,68 said they prefer thumsup to pepsi.Test the null hypothesis $P = 0.5$ against the alternative hypothesis $P > 0.5$.

sol. We have n = 125,x = 68 and p = $\frac{x}{n} = \frac{68}{125} = 0.544$

Null Hypothesis $\mathbf{H_0}$: P = 0.5.

Alternative Hypothesis $\mathbf{H_1}$: P > 0.5.

Level of significance $\propto$ =0.05

Test statistics Z = $\frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.544-0.5}{\sqrt{\frac{0.5\times0.5}{125}}} = 0.9839$

Since calculated value of IzI is less than 1.645, we accept the null hypothesis $\mathbf{H_0}$ at5% level of significance.

Pb6) Experience had shown that 20% of a manufactured product is of the top quality .In one days production of 400 articles only 50 are of top quality. Test the hypothesis at 0.05 level

Sol. We have n=400, x = 50 and p = $\frac{x}{n} = \frac{50}{400} = 0.125$

Null Hypothesis $\mathbf{H_0}$: P = 0.2.

Alternative Hypothesis $\mathbf{H_1}$: P $\neq$ 0.2.

Level of significance $\propto$ =0.05

Test statistics Z = $\frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.125-0.2}{\sqrt{\frac{0.2\times0.8}{400}}} = -3.75$

Since IzI = 3.75 > 1.96, we reject the null hypothesis $\mathbf{H_0}$ at 5% level of significance

Pb7) A social worker belives that fewer than 25% of the couple in a certain area have ever used any form of birth control .A random sample of 120 couples was contacted twenty of them said that they have used.Test the belief of the social worker at 0.05 level.

Sol. We have n=120, x = 20 and p = $\frac{x}{n} = \frac{20}{120} = \frac{1}{6}$ and

P = 0.25,Q = 1- P =0.75

Null Hypothesis $\mathbf{H_0}$: P = 0.25.

Alternative Hypothesis $\mathbf{H_1}$: P $<$ 0.25 (let one tailed test).

Level of significance $\propto$ =0.05

Test statistics Z = $\frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{\frac{1}{6}-0.25}{\sqrt{\frac{0.25\times0.75}{120}}} = -2.107$

Since IzI = 2.107 < 2.33= Z table value. We accept the null hypothesis $\mathbf{H_0}$. that is the claim or belief of social worker is true.

Pb8) A manufacturer claims that only 4% of his products are deective.A random sample of 500 were taken among which 100 were defective. Test the hypothesis at 0.05 level.

Sol. We have n=500, x = 100 and p = $\frac{x}{n} = 0.2$ and

P = 0.04,Q = 1- P =0.96.

Null Hypothesis $\mathbf{H_0}$: P = 0.04..

Alternative Hypothesis $\mathbf{H_1}$: P $>$ 0.04 (right one tailed test).

Level of significance $\propto$ =0.05

Test statistics Z = $\frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.2 - 0.04}{\sqrt{\frac{0.04\times 0.96}{500}}} = -18.26$

Since calculated I Z I= 18.26 > 1.645 = table Z.

We reject the Null Hypothesis $\mathbf{H_0}$.

Pb9) 20 people were attaced 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 that is more at 5% level.

Sol. Sample size n = 20

Number of survived people x = 18

Proportion of survived people p $= \frac{x}{n} = \frac{18}{20} = 0.9$, P = 0.85 and Q = 0.15

Null Hypothesis $\mathbf{H_0}$: P = 0.85.

Alternative Hypothesis $\mathbf{H_1}$: P $>$ 0.85 (right one tailed test).

Level of significance $\propto$ =0.05

Test statistics Z = $\frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.9 - 0.85}{\sqrt{\frac{0.85 \times 0.15}{20}}} = 0.625$

Since calculated I Z I= 0.625

Tabulated z at 5% level of significance is 1.645

Since calculated z < tabulated z, We accept the Null Hypothesis $\mathbf{H_0}$.

i.e, the proportion of the survived people is 0.85.

TESTING OF HYPOTHESIS OF DIFFERENCE OF PROPORTION

Pb1) Random sample of 400 men and 600 women were asked whether they would like to have a flyover near their residence. 200 men and 325 women were in favour of the proposal . Test the hypothesis that proportions of men and women in favour of the proposal are same at 5% level.

Sol. given sample size n_1=400,n_2 = 600.

Proportion of men $p_1 = \frac{x_1}{n_1} = \frac{200}{400} = 0.5$

Proportion of men $p_2 = \frac{x_2}{n_2} = \frac{325}{600} = 0.541$

$P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{200+325}{400+600} = 0.525$, q = 0.475

Null Hypothesis $\mathbf{H_0}$: $P_1 = P_2$. i.e there is no significant difference between the opinion of men and women as far as proposal of flyover is concerned.

Alternative Hypothesis $\mathbf{H_1}$: $P_1 \neq P_2$ (two tailed test).

Level of significance $\propto$ =0.05

Test statistics $Z = \frac{\mathrm{P1} - \mathrm{P2}}{\sqrt{pq\,(\frac{1}{\mathrm{n\,1}}+\frac{1}{\mathrm{n\,2}})}} = \frac{0.5 - 0.541}{\sqrt{0.525 \times 0.425\,(\frac{1}{400}+\frac{1}{600})}} = -1.28$

Since calculated I Z I= 1.28

Tabulated z at 5% level of significance is 1.96

Since calculated z < tabulated z, We accept the Null Hypothesis $\mathbf{H_0}$.

i.e, there is no difference of opinion between men and women as far as proposal of flyover is concerned.

Pb2) A manufacturer of electronic equipment subjects samples of two completing brands of transistors to an accelerated performance test. If 45 of 180 transistors of the first kind and 34 of 120 transistors of the second kind fail the test, what can he conclude at the level of significance $\propto = 0.05$ about the difference between the corresponding sample proportions?

Sol. given sample size n_1=180,n_2 = 120,x_1 = 45 and x_2 = 34.

$p_1 = \frac{x_1}{n_1} = \frac{45}{180} = 0.25$, $p_2 = \frac{x_2}{n_2} = \frac{34}{120} = 0.283$

$P = \frac{x_1+x_2}{n_1+n_2} = \frac{200+325}{400+600} = 0.263$, q = 0.737.

Null Hypothesis $\mathbf{H_0}$: $P_1 = P_2$. i.e there is no significant difference.

Alternative Hypothesis $\mathbf{H_1}$: $P_1 \neq P_2$, i.e There is a difference.(two tailed test).

Level of significance $\propto$ =0.05

Test statistics $Z = \frac{P1-P2}{\sqrt{pq\left(\frac{1}{n1}+\frac{1}{n2}\right)}} = \frac{0.25-0.283}{\sqrt{0.263\times0.737\left(\frac{1}{180}+\frac{1}{120}\right)}} = -0.647$

Since calculated I Z I= 0.647

Tabulated z at 5% level of significance is 1.96

Since calculated z < tabulated z, We accept the Null Hypothesis $\mathbf{H_0}$

i.e. the difference between the proportions is not significant.

Pb3) On the basis of their total scores , 200 candidates of a civil service examination are divided into two groups,the upper 30% and the remaining 70% . consider the firstquestion of the examination . Among the first group ,40 had the correct answer, whereas among the second group,80 had the correct answer . On the basis of these results.can one conclude that the first question is not good at discriminating ability of the type being examined here?

Sol. We have $n_1=60, n_2 = 140, x_1 = 40$ and $x_2 = 80$.

$p_1 = \frac{x_1}{n_1} = \frac{40}{60} = 0.667$, $p_2 = \frac{x_2}{n_2} = \frac{80}{140} = 0.571$

$P = \frac{x_1+x_2}{n_1+n_2} = \frac{40+80}{60+140} = 0.6$, $q = 0.4$.

Null Hypothesis $\mathbf{H_0}$: $P_1 = P_2$. i.e there is no significant difference.

Alternative Hypothesis $\mathbf{H_1}$: $P_1 \neq P_2$, i.e There is a difference.(two tailed test).

Level of significance $\propto = 0.05$

Test statistics $Z = \frac{P1 - P2}{\sqrt{pq\left(\frac{1}{n1}+\frac{1}{n2}\right)}} = \frac{0.667 - 0.571}{\sqrt{0.6\times0.4\left(\frac{1}{60}+\frac{1}{40}\right)}} = 1.27$

Since calculated I Z I= 1.27

Tabulated z at 5% level of significance is 1.96

Since calculated z < tabulated z, We accept the Null Hypothesis $\mathbf{H_0}$

i.e. the difference between the proportions is not significant.

Pb5) A cigarette manufacturing firm claims that its brand A line of cigarettes outsells its brand B by 8%. If it is found that 42 out of a samole of 200 smokers prefer brand A and 18 out of another sample of 100 smokers prefer brand B, test whether the 8% difference is a valid claim.

Sol. Here $n_1 = 200$, $n_2 = 100$, $x_1 = 42$ and $x_2 = 18$.

$p_1 = \frac{x_1}{n_1} = \frac{42}{200} = 0.21$, $p_2 = \frac{x_2}{n_2} = \frac{18}{100} = 0.18$. and $P_1 - P_2 = 8\% = 0.08$

$P = \frac{x_1+x_2}{n_1+n_2} = \frac{42+18}{200+100} = 0.2$, $q = 0.8$.

Null Hypothesis $\mathbf{H_0}$: $P_1 - P_2 = 8\% = 0.08$,

i.e, there is 8% difference in the sale of two brands of cigarettes is a valid claim..

Alternative Hypothesis $\mathbf{H_1}$: $P_1 - P_2 \neq 8\% = 0.08$ (two tailed test).

Level of significance $\propto = 0.05$

Test statistics $Z = \frac{(p_1 - p_2) - (P1 - P2)}{\sqrt{pq\left(\frac{1}{n1}+\frac{1}{n2}\right)}} = \frac{0.03 - 0.08}{\sqrt{0.2\times0.8\left(\frac{1}{200}+\frac{1}{100}\right)}} = -1.02$

Since calculated $|Z| = 1.02$

Tabulated z at 5% level of significance is 1.96

Since calculated z < tabulated z, We accept the Null Hypothesis $\mathbf{H_0}$

i.e. there is 8% difference in the sale of two brands of cigarettes is a valid claim.

Pb6) In two large populations, there are 30%,and 25% respectively of fair haired people.Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations.

Sol. Given $n_1 = 1200$, $n_2 = 900$, $x_1 = 30$ and $x_2 = 25$.

$p_1 = \frac{x_1}{n_1} = \frac{30}{100} = 0.3$, $p_2 = \frac{x_2}{n_2} = \frac{25}{100} = 0.25.$

Null Hypothesis $\mathbf{H_0}$: $P_1 = P_2$. i.e there is no significant difference.

Alternative Hypothesis $\mathbf{H_1}$: $P_1 \neq P_2$, i.e There is a difference.(two tailed test).

Level of significance $\propto$ =0.05

Test statistics $Z = \frac{P1 - P2}{\sqrt{(\frac{P1Q1}{n1} + \frac{P2Q2}{n2})}} = \frac{0.3 - 0.25}{\sqrt{(\frac{0.3 \times 0.7}{1200} + \frac{0.25 \times 0.75}{900})}} = 2.56$

Since calculated I Z I= 2.56

Tabulated z at 5% level of significance is 1.96

Since calculated z > tabulated z, We reject the Null Hypothesis $\mathbf{H_0}$

i.e. the difference between the proportions is significant.

CHI-SQUARE TEST FOR GOODNESS OF FIT $\left(\chi^2\right)$

A test for testing the significance of discrepancy between experimental values and the theoretical values obtained under some hypothesis.

Statistic

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

Where

O- observed frequency

E-Expected frequency

NOTE

If the data is given in a series of "n" numbers then degrees of freedom=n-1

In case of Binomial distribution , d.f. = n-1

In case of Poission distribution , d.f. = n-2

In case of Normal distribution , d.f. = n-3

PROBLEM:1

The number of automobile accidents per week in a certain community are as follows:

12,8,20,2,14,10,15,6,9,4. Are these frequencies in agreement with the belief that accident conditions were the same during this 10 week period.

SOLUTION

Expected frequency of accidents each week $= \dfrac{100}{10} = 10$

NULL HYPOTHESIS H_0: The accident conditions were the same during the 10 week period.

OBSERVED FREQUENCY(o)	EXPECTED FREQUENCY(E)	(O-E)	$\frac{(O-E)^2}{E}$
12	10	2	0.4
8	10	-2	0.4
20	10	10	10.0
2	10	-8	6.4
14	10	4	1.6
10	10	0	0.0
15	10	5	2.5
6	10	-4	1.6
9	10	-1	0.1
4	10	-6	3.6
Total	100		26.6

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 26.6 \quad \text{(calculated)}$$

Degrees of freedom = n-1= 10-1=9

Tabulated $\chi^2{}_{0.05} = $=16.9

SINCE Calculated χ^2 > Tabulated χ^2 , The null hypothesis is rejected

i.e., The accident conditions were not the same during the 10 week period

PROBLEM:2

A sample analysis of examination results of 500 students was made . It was found that 220 students had failed , 170 had secured a third class, 90 were placed in second class and 20 got a first class . Do these figures commensurate with the general examination result which is in the ratio of 4:3:2:1 for the various categories respectively

SOLUTION

Expected frequencies are in the ratio of 4:3:2:1

Total frequency=500

If we divide the total frequency 500 in the ratio 4:3:2:1, we get the expected frequencies as 200, 150, 100,50

Class	OBSERVED FREQUENCY(O)	EXPECTED FREQUENCY(E)	(O-E)	$\frac{(O-E)^2}{E}$
Failed	220	200	20	2.00
Third	170	150	20	2.667
Second	90	100	-10	1.000
First	20	50	-30	18.00
	500	500		23.667

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$ =23.667 (calculated)

Degrees of freedom = n-1= 4-1=3

Tabulated $\chi^2{}_{0.05}$ = =7.81

SINCE Calculated χ^2 > Tabulated χ^2 , The null hypothesis is rejected

The observed results are not commensurate with the general examination results.

EXERCISE

1.200digits were chosen at random from a set tables . The frequencies of the digits are shown below:

Digit	0	1	2	3	4	5	6	7	8	9
frequency	18	19	23	21	16	25	22	20	21	15

Use the chi square test to assess the correctness of the hypothesis that the digits were distributed in equal number in the tables from which these were chosen

2.A pair of dice are thrown 360 times and frequency of each sum is indicated below.

Sum	2	3	4	5	6	7	8	9	10	11	12
Frequency	8	24	35	37	44	65	51	42	26	14	14

Would you say that the dice are fair on the basis of the chi-square test at 0.05 level of significance

<u>CHI-SQUARE TEST FOR INDEPENDENCE OF ATTRIBUTES:</u>

In this Chi square test, we test if two attributes A and B under consideration are independent or not.

Null hypothesis H_0 : Attributes are independent.

Degrees of freedom : d.f= (r-1)(c-1)

R= number of rows, c= number of columns

PROBLEM:1

On the basis of information given below about the treatment of 200 patients suffering from a disease, state whether the new treatment is comparatively superior to the conventional treatment.

	Favourable	Not favourable	Total
NEW	60	30	90
Conventional	40	70	110

SOLUTION

NULL HYPOTHESIS $\mathbf{H_0}$**:** No difference between new and conventional treatment (new and conventional treatment are independent)

The number of degrees of freedom is (2-1)(2-1)=1

	FAVOURABLE	NOT FAVOURABLE	TOTAL
NEW	60	30	90
CONVENTIONAL	40	70	110
TOTAL	100	100	200

Expected frequencies are given in the table

$\frac{90(100)}{200} = 45$	$\frac{90(100)}{200} = 45$	90
$\frac{100(110)}{200} = 55$	$\frac{100(110)}{200} = 55$	110
100	100	200

Calculation of χ^2 :

Observed frequency	Expected frequency	$(O-E)^2$	$\frac{(O-E)^2}{E}$
60	45	225	5
30	45	225	5
40	55	225	4.09
70	55	225	4.09
			18.18

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad = 18.18$$

Tabulated $\chi^2{}_{0.05} = 3.841$(degrees of freedom=1)

SINCE Calculated χ^2 > Tabulated χ^2 , The null hypothesis is rejected

That is conventional and new treatment are not independent.

PROBLEM:2

Given the following contingency table for hair colour and eye colour. Find the value of χ^2 . Is there good association between the two.

Hair colour					
		Fair	Brown	Black	Total
Eye Colour	Blue	15	5	20	40
	gray	20	10	20	50
	brown	25	15	20	60
	Total	60	30	60	150

SOLUTION

Null Hypothesis H_0: The two attributes , hair and eye colour are independent

Table of expected frequencies:

$\frac{60\times 40}{150}=16$	$\frac{30\times 40}{150}=8$	$\frac{60\times 40}{150}=16$	40
$\frac{60\times 50}{150}=20$	$\frac{30\times 50}{150}=10$	$\frac{60\times 50}{150}=20$	50
$\frac{60\times 60}{150}=24$	$\frac{30\times 60}{150}=12$	$\frac{60\times 60}{150}=24$	60
Total= 60	30	60	150

Calculation of χ^2:

Observed frequency	Expected frequency	$(O-E)^2$	$\frac{(O-E)^2}{E}$
15	16	1	0.0625
5	8	9	1.125
20	16	16	1
20	20	0	0
10	10	0	0
20	20	0	0
25	24	1	0.042
15	12	9	0.75
20	24	16	0.666
			3.6458

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad = 3.6458$$

Tabulated $\chi^2{}_{0.05} = 9.488$(degrees of freedom=(3-1)(3-1)=4)

SINCE Calculated χ^2 < Tabulated χ^2 , The null hypothesis is accepted.

The hair colour and eye colour are independent.

EXERCISE

1. The following table gives the classification of 100 workers according to sex and nature of work. Test whether the nature of work is independent of the worker.

	stable	unstable	total
males	40	20	60
females	10	30	40
	50	50	100

2. From the following data, find whether there is any significant liking in the habit of taking soft drinks among the categories of employees.

	employees		
Soft drinks	clerks	teachers	officers
Pepsi	10	25	65
Thumsup	15	30	65
Fanta	50	60	30