Mathematical Statistics II

Muzammil Tanveer

mtanveer8689@gmail.com

0316-7017457

Dedicated

To

My Honorable Teacher

Sir Haidar Ali

&

My Parents

Lecture # 01

Recommended Book:

Introduction to Statistical theory part-II by Sher M. Chaudhary

Some Basic Definitions:

Population:

The population is the totality of the observation in which we are concerned. A population can either be finite or infinite.

Size of population:

The number of observations in a population is said to be the size of population denoted by N.

Sample:

The subset of population is called sample.

Sampling:

It is a statistically technique which is used to collect information and on the basic of this information. We form inferences (results) about the characteristics of the population.

Examples:

(i) The number of cards in a deck.
(ii) The height of residence in a certain city.
(iii) The number of students in mathematics department.
(iv) The population of all points on a line.
(v) The number of germs on the body of sick patient.

Remark:

(i), (ii) and (iii) are examples of finite population while (iv) and (v) are examples of infinite population.

Sampling unit:

An individual member of the population is called sampling unit or simple unit.

A sampling unit from which information is required, may be a college student, and animal, or tree, a business etc.

A set of 'n' sampling units selected from a given population is called a sample of size 'n' and process of selecting a sample is called sampling.

Random sample:

It is defined as a subset of the statistical population in which each of the member of the subset has equal probability when it is selected.

Parameter:

A numerical value such as mean, median and standard deviation calculated from the population is said to be parameter of the population.

Statistics:

A numerical value such as mean, median and standard deviation calculated from a sample is called statistics.

Note:

The parameter has fixed value i.e. it is constant and it is denoted by a Greek letter μ ,δ for the population mean and standard deviation of the population, while on the other hand the statistics varies from sample to sample of the same population and denoted by $\mu_{\overline{X}}, \delta_{\overline{X}}$ for sample mean and standard deviation of the sample.

Sampling distribution:

The sampling distribution of the statistics depend on the size of the sample and sampling.

Sampling with replacement and without replacement:

If we select a sample from a population, observed and then returned again to the population before selecting the next sample. In this case the size of population remains same while on the other hand when we select a sample, observed it and not returned to the population before select the other sample. In this case the size of the population step by step decrease.

Theorem:

The mean of the sampling distribution of $\overline{X}$ is denoted by $\mu_{\overline{X}}$ is equal to the population mean μ. **OR** Show that $\mu_{\overline{X}} = \mu$

Proof: Let $X_1, X_2, ...X_n$ be a random sample of size 'n'. It is selected from the population with mean μ

We know that $$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Taking expected value on both side

$$E(\overline{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right)$$

$$\mu_{\overline{X}} = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right)$$

$$\mu_{\overline{X}} = \frac{1}{n} E(X_1 + X_2 + ... + X_n)$$

$$\mu_{\overline{X}} = \frac{1}{n}\left[E(X_1) + E(X_2) + ... + E(X_n)\right] \quad \_\_\_ (i)$$

As the sample which is selected as random sample. So, the random variables are independent.

$$E(X_1) = E(X_2) = ... = E(X_n) = \mu$$

Put in (i) $\Rightarrow$ $$\mu_{\overline{X}} = \frac{1}{n}[\mu + \mu + ... + \mu]$$

$$\mu_{\overline{X}} = \frac{1}{n}[n\mu] = \mu$$ Hence proved.

Theorem:

Let a random sample of size 'n' is drawn from an infinite population or with replacement from a finite population.

The standard deviation of the sampling distribution of $\overline{X}$ is equal to the population standard deviation divided by positive square root of the sample size i.e. $\delta_{\overline{X}} = \frac{\delta}{\sqrt{n}}$.

Proof:

Let $X_1, X_2, ...X_n$ be a random sample selected from the population with standard deviation δ.

We know that $$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Taking variance on both side

$$Var\left(\overline{X}\right) = Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right)$$

$$\delta_{\overline{X}}^2 = \frac{1}{n^2}\left[Var\left(\sum_{i=1}^{n} X_i\right)\right]$$

$$\delta_{\overline{X}}^2 = \frac{1}{n^2}\left[Var\left(X_1 + X_2 + ... + X_n\right)\right]$$

$$\delta_{\overline{X}}^2 = \frac{1}{n^2}\left[Var\left(X_1\right) + Var\left(X_2\right) + ... + Var\left(X_n\right)\right] \_\_\_ (i)$$

As the sample which is selected as random sample. So, the random variables are independent.

$$Var\left(X_1\right) = Var\left(X_2\right) = ... = Var\left(X_n\right) = \delta^2$$

Put in (i) $\Rightarrow$ $$\delta_{\overline{X}}^2 = \frac{1}{n^2}\left[\delta^2 + \delta^2 + ... + \delta^2\right]$$

$$\delta_{\overline{X}}^2 = \frac{1}{n^2}\left[n\delta^2\right]$$

$$\delta_{\overline{X}}^2 = \frac{1}{n}\delta^2$$

$$\delta_{\overline{X}} = \frac{\delta}{\sqrt{n}} \quad \text{proved}$$

Note:

If the size of the population is N and size of sample is n then N^n is used to form samples with replacement and combination C_n^N used to form samples without replacement.

Question:

Suppose that the population consist of five numbers 1,2,3,4,5. Draw all possible sample with replacement of size 2 and also verify that $\mu_{\overline{X}} = \mu$ and $\delta_{\overline{X}} = \frac{\delta}{\sqrt{n}}$.

Solution: As given N = 5 , n = 2

$sample = N^n = 5^2 = 25$ and these are

(1,1) (1,2) (1,3) (1,4) (1,5)

(2,1) (2,2) (2,3) (2,4) (2,5)

(3,1) (3,2) (3,3) (3,4) (3,5)

(4,1) (4,2) (4,3) (4,4) (4,5)

(5,1) (5,2) (5,3) (5,4) (5,5)

The corresponding means are 1 , 1.5 , 2 , 2.5 , 3

1.5 , 2 , 2.5 , 3 , 3.5

2 , 2.5 , 3 , 3.5 , 4

2.5 , 3 , 3.5 , 4 , 4.5

3 , 3.5 , 4 , 4.5 , 5

$$\text{Population mean} = \mu = \frac{\sum X}{N}$$

$$\text{Sample mean} = \mu_{\overline{X}} = \sum \overline{X} \, . f\left(\overline{X}\right)$$

$$\text{Population standard deviation} = \delta = \sqrt{\frac{\sum (X-\mu)^2}{N}}$$

$$\text{Sample standard deviation} = \delta_{\overline{X}} = \sqrt{\sum \overline{X^2} \, . f\left(\overline{X}\right) - \left(\sum \overline{X} \, . f\left(\overline{X}\right)\right)^2}$$

$\overline{X}$	$f\left(\overline{X}\right)$	$\overline{X} f\left(\overline{X}\right)$	$\overline{X}^2$	$\overline{X}^2 f\left(\overline{X}\right)$
1	1/25	1/25	1	1/25
1.5	2/25	3/25	2.25	4.5/25
2	3/25	6/25	4	12/25
2.5	4/25	10/25	6.25	25/25
3	5/25	15/25	9	45/25
3.5	4/25	14/25	12.25	49/25
4	3/25	12/25	16	48/25
4.5	2/25	9/25	20.25	40.5/25
5	1/25	5/25	25	25/25
		75/25=3		250/25=10

$$\mu = \frac{\sum X}{N} = \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$$

$$\mu_{\overline{X}} = \sum \overline{X} \, . f\left(\overline{X}\right) = 3$$

$$\Rightarrow \quad \mu = \mu_{\overline{X}}$$

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

$$\delta = \sqrt{\frac{10}{5}} = \sqrt{2}$$

$$\delta_{\overline{X}} = \sqrt{10-3^2} = \sqrt{10-9} = 1$$

$$\delta_{\overline{X}} = \frac{\delta}{\sqrt{n}}$$

$$1 = \frac{\sqrt{2}}{\sqrt{2}}$$

$$\Rightarrow 1 = 1$$

Question:

Suppose that a population consist of four numbers such as 3,7,9,15. Draw all possible sample with replacement of size 2 and verify that $\mu_{\overline{X}} = \mu$ and $\delta_{\overline{X}} = \dfrac{\delta}{\sqrt{n}}$.

Solution:

As given N = 4 , n = 2

$sample = N^n = 4^2 = 16$ and these are

(3,3) , (3,7) ,(3,9) , (3,15)

(7,3) , (7,7) , (7,9) , (7,15)

(9,3) , (9,7),(9,9) , (9,15)

(15,3) (15,7) , (15,9) , (15,15)

There corresponding means are

3 , 5 , 6 , 9

5 ,7 , 8 , 11

6 , 8 , 9 , 12

9 , 11 , 12 ,15

Population mean = $\mu = \dfrac{\sum X}{N}$

Sample mean = $\mu_{\overline{X}} = \sum \overline{X} \, . f\left(\overline{X}\right)$

$$\text{Population standard deviation} = \delta = \sqrt{\frac{\sum(X-\mu)^2}{N}}$$

$$\text{Sample standard deviation} = \delta_{\overline{X}} = \sqrt{\sum \overline{X^2}\,.f\left(\overline{X}\right) - \left(\sum \overline{X}\,.f\left(\overline{X}\right)\right)^2}$$

$\overline{X}$	$f\left(\overline{X}\right)$	$\overline{X}f\left(\overline{X}\right)$	$\overline{X}^2$	$\overline{X}^2 f\left(\overline{X}\right)$
3	1/16	3/16	9	9/16
5	2/16	10/16	25	50/16
6	2/16	12/16	36	72/16
7	1/16	7/16	49	49/16
8	2/16	16/16	64	128/16
9	3/16	27/16	81	243/16
11	2/16	22/16	121	242/16
12	2/16	24/16	144	288/16
15	1/16	15/16	225	225/16
		136/16 = 8.5		1306/16 = 81.625

$$\mu = \frac{\sum X}{N} = \frac{3+7+9+15}{4} = \frac{34}{4} = 8.5$$

$$\mu_{\overline{X}} = \sum \overline{X}\,.f\left(\overline{X}\right) = 8.5$$

$$\Rightarrow \quad \mu = \mu_{\overline{X}}$$

$$\delta = \sqrt{\frac{(3-8.5)^2+(7-8.5)^2+(9-8.5)^2+(15-8.5)^2}{4}} = \sqrt{\frac{75}{4}} = 4.33$$

$$\delta_{\overline{X}} = \sqrt{81.625-(8.5)^2} = \sqrt{81.625-72.25} = 3.062$$

$$\delta_{\overline{X}} = \frac{\delta}{\sqrt{n}}$$

$$3.062 = \frac{4.33}{\sqrt{2}}$$

$$\Rightarrow 3.062 = 3.062$$

Lecture # 02

Question: Draw all possible sample of size 2 with replacement from a population consisting of 3,6,9,12,15 from the sampling distribution of sample means and verify the results

(i) $\mu_{\overline{X}} = \mu$

(ii) $\delta_{\overline{X}} = \dfrac{\delta}{\sqrt{n}}$

Solution: Here N = 5 , n = 2

$sample = N^n = 5^2 = 25$

$$\mu = \frac{\sum X_i}{N} = \frac{3+6+9+12+15}{5} = 9$$

$$\delta = \sqrt{\frac{\sum (X_i - \mu)^2}{N}} = \sqrt{\frac{(3-9)^2 + (6-9)^2 + (9-9)^2 + (12-9)^2 + (15-9)^2}{5}}$$

$$\delta = \sqrt{18} = 3\sqrt{2}$$

(3,3) , (3,6) ,(3,9) ,(3,12), (3,15)

(6,3) , (6,6) ,(6,9) ,(6,12), (6,15)

(9,3) , (9,6) ,(9,9) ,(9,12), (9,15)

(12,3) , (12,6) ,(12,9) ,(12,12), (12,15)

(15,3) , (15,6) ,(15,9) ,(15,12), (15,15)

There corresponding means

3,4.5,6,7.5,9

4.5,6,7.5,9,10.5

6,7.5,9,10.5,12

7.5,9,10.5,12,13.5

9,10.5,12,12.5,15

$\overline{X}$	$f\left(\overline{X}\right)$	$\overline{X}f\left(\overline{X}\right)$	$\overline{X}^2$	$\overline{X}^2 f\left(\overline{X}\right)$
3	1/25	3/25	9	9/25
4.5	2/25	9/25	20.25	40.5/25
6	3/25	18/25	36	108/25
7.5	4/25	30/25	56.25	225/25
9	5/25	45/25	81	405/25
10.5	4/25	42/25	110.25	441/25
12	3/25	36/25	144	432/25
13.5	2/25	27/25	182.25	364.5/25
15	1/25	15/25	225	225/25
		225/25=9		2250/25=90

$$\mu_{\overline{X}} = \sum \overline{X}\,.f\left(\overline{X}\right) = 9$$

$$\Rightarrow \quad \mu = \mu_{\overline{X}}$$

$$\delta_{\overline{X}} = \sqrt{\sum \overline{X^2}\,.f\left(\overline{X}\right) - \left(\sum \overline{X}\,.f\left(\overline{X}\right)\right)^2} = \sqrt{90-(9)^2} = \sqrt{90-81} = 3$$

$$\delta_{\overline{X}} = \frac{\delta}{\sqrt{n}}$$

$$3 = \frac{3\sqrt{2}}{\sqrt{2}} = 3$$

$$\delta_{\overline{X}} = \delta$$

Theorem:

If a random sample of size ‘n’ is taken from a finite population without replacement of size ‘N’ then the standard deviation of sample distribution is $\overline{X}$ is

$$\delta_{\overline{X}} = \frac{\delta}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Solution: Let $X_1, X_2, ...X_n$ be a random variable of a finite population with size N and standard deviation δ.

We know that

$$\delta_{\overline{X}}^2 = E\left(\overline{X} - E\left(\overline{X}\right)\right)^2$$

$$\delta_{\overline{X}}^2 = E\left(\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right)^2 \qquad \because \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \; E\left(\overline{X}\right) = \mu$$

$$\delta_{\overline{X}}^2 = E\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{n}\right)^2$$

$$\delta_{\overline{X}}^2 = \frac{1}{n^2} E\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu\right)^2 \qquad \because \sum_{i=1}^{n} = n$$

$$\delta_{\overline{X}}^2 = \frac{1}{n^2} E\left(\sum_{i=1}^{n} (X_i - \mu)\right)^2$$

$$\delta_{\overline{X}}^2 = \frac{1}{n^2} E\left(\sum_{i=1}^{n} (X_i - \mu)^2\right) \qquad \because E\left(\sum_{i=1}^{n} (X_i - \mu)\right)^2 = E\left(\sum_{i=1}^{n} (X_i - \mu)^2\right)$$

$$\delta_{\overline{X}}^2 = \frac{1}{n^2} E\left(\sum_{i=1}^{n} (X_i - \mu)^2 + 0\right)$$

$$\delta_{\overline{X}}^2 = \frac{1}{n^2} E\left(\sum_{i=1}^{n} (X_i - \mu)^2 + \sum_{i \neq j}^{N} (X_i - \mu)(X_j - \mu)\right)$$

$$\delta_{\overline{X}}^2 = \frac{1}{n^2} \left(\sum_{i=1}^{n} E(X_i - \mu)^2 + \sum_{i \neq j}^{N} E(X_i - \mu)(X_j - \mu)\right) \qquad \_\_\_(i)$$

$$\because \; E(X_i - \mu)^2 = \delta^2 \; and \; E(X_i - \mu)(X_j - \mu) = \frac{-\delta^2}{N-1}$$

Put in (i)

In general

$$\delta_{\overline{X}}^2 = E\left(\overline{X^2}\right) - \left(E\left(\overline{X}\right)\right)^2$$

Now

$$\delta_{\overline{X}}^2 = E\left(\overline{X} - E\left(\overline{X}\right)\right)^2$$

$$= E\left(\overline{X^2} - 2\overline{X}E\left(\overline{X}\right) + \left(E\left(\overline{X}\right)\right)^2\right)$$

$$= E\left(\overline{X^2}\right) - E\left(2\overline{X}E\left(\overline{X}\right)\right) + \left(E\left(\overline{X}\right)\right)^2$$

$$= E\left(\overline{X^2}\right) - 2E\left(\overline{X}\right)E\left(\overline{X}\right) + \left(E\left(\overline{X}\right)\right)^2$$

$$= E\left(\overline{X^2}\right) - 2\left(E\left(\overline{X}\right)\right)^2 + \left(E\left(\overline{X}\right)\right)^2$$

$$\delta_{\overline{X}}^2 = E\left(\overline{X^2}\right) - \left(E\left(\overline{X}\right)\right)^2$$

$$\because \sum_{i=1}^{n} = n \quad , \quad \because \sum_{i \neq j}^{N} = n(n-1)$$

In general

$$\left(\sum_{i=1}^{n} (X_i - \mu)\right)^2 \neq \left(\sum_{i=1}^{n} (X_i - \mu)^2\right)$$

But

$$E\left(\sum_{i=1}^{n} (X_i - \mu)\right)^2 = E\left(\sum_{i=1}^{n} (X_i - \mu)^2\right)$$

In tensor $\delta_{ij} = \begin{cases} 1 & if \;\; i = j \\ 0 & if \;\; i \neq j \end{cases}$

$$\delta_{\overline{X}}^2=\frac{1}{n^2}\left(\sum_{i=1}^{n}\delta^2+\sum_{i\neq j}^{N}\frac{-\delta^2}{N-1}\right)$$

$$\delta_{\overline{X}}^2=\frac{1}{n^2}\left(n\delta^2+n(n-1)\left(\frac{-\delta^2}{N-1}\right)\right)\quad \because \sum_{i=1}^{n}=n\ ,\ \sum_{i\neq j}^{N}=n(n-1)$$

$$\delta_{\overline{X}}^2=\frac{n\delta^2}{n^2}\left(1-\frac{(n-1)}{N-1}\right)$$

$$\delta_{\overline{X}}^2=\frac{\delta^2}{n}\left(\frac{N-1-n+1}{N-1}\right)$$

$$\delta_{\overline{X}}^2=\frac{\delta^2}{n}\left(\frac{N-n}{N-1}\right)$$

$$\delta_{\overline{X}}=\frac{\delta}{\sqrt{n}}\sqrt{\left(\frac{N-n}{N-1}\right)}$$

Question: Draw all possible sample of size 2 without replacement from a population consisting 3,6,9,12,15 from a sampling distribution of sample mean and verify the result.

(i) $\mu_{\overline{X}}=\mu$

(ii) $\delta_{\overline{X}}=\frac{\delta}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

Solution: Here N = 5 , n = 2

$sample\ {}^{N}C_n={}^{5}C_2=10$

$$\mu=\frac{\sum X_i}{N}=\frac{3+6+9+12+15}{5}=9$$

$$\delta=\sqrt{\frac{\sum(X_i-\mu)^2}{N}}=\sqrt{\frac{(3-9)^2+(6-9)^2+(9-9)^2+(12-9)^2+(15-9)^2}{5}}$$

$$\Rightarrow\ \delta=\sqrt{18}=3\sqrt{2}$$

Samples: (3,6),(3,9),(3,12),(3,15),(6,9),(6,12),(6,15),(9,12),(9,15),(12,15)

There corresponding means

4.5, 6, 7.5 ,9 ,7.5 ,9 ,10.5 ,10.5 ,12 ,13.5

$\overline{X}$	$f\left(\overline{X}\right)$	$\overline{X}f\left(\overline{X}\right)$	$\overline{X}^2$	$\overline{X}^2 f\left(\overline{X}\right)$
4.5	1/10	4.5/10	20.25	20.25/10
6	1/10	6/10	36	36/10
7.5	2/10	15/10	56.25	112.5/10
9	2/10	18/10	81	162/10
10.5	2/10	21/10	110.25	220.5/10
12	1/10	12/10	144	144/10
		90/10 = 9		877.75/10 = 87.775

$$\mu_{\overline{X}} = \sum \overline{X} \,.\, f\left(\overline{X}\right) = 9$$

$$\Rightarrow \quad \mu = \mu_{\overline{X}}$$

$$\delta_{\overline{X}} = \sqrt{\sum \overline{X^2} \,.\, f\left(\overline{X}\right) - \left(\sum \overline{X} \,.\, f\left(\overline{X}\right)\right)^2} = \sqrt{87.775 - 81} = 2.60288$$

$$\delta_{\overline{X}} = \frac{\delta}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$2.60288 = \frac{3\sqrt{2}}{\sqrt{2}}\sqrt{\frac{5-2}{5-1}}$$

$$2.60288 = \sqrt{\frac{3}{4}}$$

$$2.60288 = 2.598$$

$$2.6 = 2.6$$

$$\Rightarrow \quad \delta_{\overline{X}} = \frac{\delta}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Question: Suppose that a population consist of 5 number i.e. 4,8,12,16,20. Draw all possible sample of size 3

(a) With replacement
(b) Without replacement

And verify the results

(i) $\mu_{\overline{X}} = \mu \quad , \; \delta_{\overline{X}} = \dfrac{\delta}{\sqrt{n}}$

(ii) $\mu_{\overline{X}} = \mu \quad , \; \delta_{\overline{X}} = \dfrac{\delta}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}}$

Solution: Here N = 5 , n = 3

$sample = N^n = 5^3 = 125$

$$\mu = \frac{\sum X_i}{N} = \frac{4+8+12+16+20}{5} = 12$$

$$\delta = \sqrt{\frac{\sum (X_i - \mu)^2}{N}} = \sqrt{\frac{(4-12)^2 + (8-12)^2 + (12-12)^2 + (16-12)^2 + (20-12)^2}{5}} = 5.65685$$

$(4,4,4),(4,4,8),(4,4,12),(4,4,16),(4,4,20),(4,8,4),(4,8,8),(4,8,12),(4,8,16),(4,8,20),$

$(4,12,4),(4,12,8),(4,12,12),(4,12,16),(4,12,20),(4,16,4),(4,16,8),(4,16,12),(4,16,16),$

$(4,16,20),(4,20,4),(4,20,8),(4,20,12),(4,20,16),(4,20,20),$

$(8,4,4),(8,4,8),(8,4,12),(8,4,16),(8,4,20),(8,8,4),(8,8,8),(8,8,12),(8,8,16),(8,8,20),$

$(8,12,4),(8,12,8),(8,12,12),(8,12,16),(8,12,20),(8,16,4),(8,16,8),(8,16,12),(8,16,16),$

$(8,16,20),(8,20,4),(8,20,8),(8,20,12),(8,20,16),(8,20,20),$

$(12,4,4),(12,4,8),(12,4,12),(12,4,16),(12,4,20),(8,8,4),(12,8,8),(12,8,12),(12,8,16),(12,8,20),$

$(12,12,4),(12,12,8),(12,12,12),(12,12,16),(12,12,20),(8,16,4),(12,16,8),(12,16,12),(12,16,16),$

$(12,16,20),(12,20,4),(12,20,8),(12,20,12),(12,20,16),(12,20,20),$

$(16,4,4),(16,4,8),(16,4,12),(16,4,16),(16,4,20),(16,8,4),(16,8,8),(16,8,12),(16,8,16),(16,8,20),$

$(16,12,4),(16,12,8),(16,12,12),(16,12,16),(16,12,20),(16,16,4),(16,16,8),(16,16,12),(16,16,16),$

$(16,16,20),(16,20,4),(16,20,8),(16,20,12),(16,20,16),(16,20,20),$

$(20,4,4),(20,4,8),(20,4,12),(20,4,16),(20,4,20),(20,8,4),(20,8,8),(20,8,12),(20,8,16),(20,8,20),$

$(20,12,4),(20,12,8),(20,12,12),(20,12,16),(20,12,20),(20,16,4),(20,16,8),(20,16,12),(20,16,16),$

$(20,16,20),(20,20,4),(20,20,8),(20,20,12),(20,20,16),(20,20,20)$

These corresponding means

4, 5.3, 6.67, 8, 9.33, 5.3, 6.67, 8, 9.33, 10.67,

6.67, 8, 9.33, 10.67, 12, 8, 9.33, 10.67, 12, 13.33,

9.33, 10.67, 12, 13.33, 14.67

5.3, 6.67, 8, 9.33, 10.67, 6.67, 8, 9.33, 10.67, 12,

8, 933, 10.67, 12, 13.33, 9.33, 10.67, 12, 13.33, 14.67,

10.67, 12, 13.33, 14.67, 16,

6.67, 8, 9.33, 10.67, 12, 8, 9.33, 10.67, 12, 13.33

9.33, 10.67, 12, 13.33, 14.67, 10.67, 12, 13.33, 14.67, 16,

12, 13.33, 14.67, 16, 17.33

8, 9.33, 10.67, 12, 13.33, 9.33, 10.67, 12, 13.33, 14.67,

10.67, 12, 13.33, 14.67, 16, 12, 13.33, 14.67, 16,17.33,

13.33, 14.67, 16, 17.33, 18.67

9.33, 10.67, 12, 13.33, 14.67, 10.67, 12, 13.33, 14.67, 16,

12, 13.33, 14.67, 16, 17.33, 13.33, 14.67, 16, 17.33, 18.67,

14.67, 16, 17.33, 18.67, 20

$\overline{X}$	$f\left(\overline{X}\right)$	$\overline{X}f\left(\overline{X}\right)$	$\overline{X}^2$	$\overline{X}^2 f\left(\overline{X}\right)$
4	1/125	4/125	16	16/125
5.3	3/125	15.9/125	28.09	84.27/125
6.67	6/125	40.02/125	44.4889	266.93/125
8	10/125	80/125	64	640/125
9.33	15/125	139.95/125	87.0489	1305.73/125
10.67	18/125	192.06/125	113.8489	2049.28/125
12	19/125	228/125	144	2736/125
13.33	18/125	239.94/125	177.6889	3198.4/125
14.67	15/125	220.05/125	215.2089	3228.13/125
16	10/125	160/125	256	2560/125
17.33	6/125	130.98/125	300.3289	1801.97/125
18.67	3/125	56.01/125	348.5689	1045.71/125
20	1/125	20/125	400	400/125
		1499.91/125=11.999		19332.42/125=154.66

$$\mu_{\overline{X}} = \sum \overline{X} \,.\, f\left(\overline{X}\right) = 11.999 \approx 12$$

$$\Rightarrow \mu_{\overline{X}} = \mu$$

$$\delta_{\overline{X}} = \sqrt{\sum \overline{X^2} \,.\, f\left(\overline{X}\right) - \left(\sum \overline{X} \,.\, f\left(\overline{X}\right)\right)^2} = \sqrt{154.66 - 144} = 3.26497$$

$$\delta_{\overline{X}} = \frac{\delta}{\sqrt{n}}$$

$$3.26497 = \frac{5.65685}{\sqrt{3}} = 3.2659$$

$$\Rightarrow \delta_{\overline{X}} = \frac{\delta}{\sqrt{n}}$$

(c) Without replacement

$sample\ {}^{N}C_n = {}^{5}C_3 = 10$

$(4,8,12),(4,8,16),(4,8,20),(4,12,16),(4,12,20),(4,16,20),(8,12,16),(8,12,20),(8,16,20),(12,16,20)$

There corresponding means 8, 9.33, 10.67, 10.67, 12, 13.33, 12, 13.33, 14.67, 16

$\overline{X}$	$f\left(\overline{X}\right)$	$\overline{X}f\left(\overline{X}\right)$	$\overline{X}^2$	$\overline{X}^2 f\left(\overline{X}\right)$
8	1/10	8/10	64	64/10
9.33	1/10	9.33/10	87.0489	87.0489/10
10.67	2/10	21.34/10	113.85	227.698/10
12	2/10	24/10	144	288/10
13.33	2/10	26.66/10	177.69	355.38/10
14.67	1/10	14.67/10	215.21	215.21/10
16	1/10	16/10	256	256/10
		120/10 = 12		1493.777/10 = 149.3777

$$\mu_{\overline{X}} = \sum \overline{X}\ .f\left(\overline{X}\right) = 12$$

$$\Rightarrow\ \mu_{\overline{X}} = \mu$$

$$\delta_{\overline{X}} = \sqrt{\sum \overline{X^2}\ .f\left(\overline{X}\right) - \left(\sum \overline{X}\ .f\left(\overline{X}\right)\right)^2} = \sqrt{149.3777 - 144} = 2.319$$

$$\delta_{\overline{X}} = \frac{\delta}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$2.319 = \frac{5.65685}{\sqrt{3}}\sqrt{\frac{5-3}{5-1}} = 2.31$$

Lecture # 03

Statistical Inference:

The process of drawing the inferences (result of conclusion) about a population on the basis of information contained in a sample taken from a population is called Statistical Inference.

The major part statistical inference is divided into two areas

(i) Estimation
(ii) Testing of hypothesis

The process of making judgement about a population parametric is called statistical estimation or simply estimation. There are two types of estimation

(i) Point estimation
(ii) Interval estimation

OR A rough calculation of the value, number, quantity or extent of something.

Point Estimation:

An estimation of a population parameter given by a single number is called point estimation.

Interval estimation:

An estimation of a population parameter given by a two number between which the parameter may be considered to lie is called an interval estimation.

Example: If we say a distance in measured as 5.28m, we are giving a point estimate.

If on the other hand we say that the distance is 5.28 ± 0.03 m (i.e. the distance lies between 5.25 and 5.31m), we are giving an interval estimation.

Point Estimator:

A statistic which is used to estimate a parameter is called point estimator.
i.e. $E\left(\overline{X}\right)=\mu$. Here, $\overline{X}$ is a point estimator.

Good point estimator:

If $E\left(\overline{X}\right)=\mu$ and Let $\overline{X}$ = 3 and μ = 3 then $\overline{X}$ is called good point estimator.

Criteria for Good point estimator:

For a good point estimator, the following condition will be satisfied

(i) Unbiasedness
(ii) Consistency
(iii) Efficiency
(iv) Sufficiency

Unbiasedness:

An estimator is called unbiasedness is its expected value is equal to the corresponding population parameter otherwise it is called biased.

Suppose that θ is an arbitrary estimator and for population parameter. It is denoted by $\hat{\theta}$, then according to the definition of unbiasedness.

We can write $E\left(\hat{\theta}\right)=\theta$

Here θ = parameter & $\hat{\theta}$ = statistics

There arise three cases

(i) $E\left(\hat{\theta}\right)-\theta=0 \quad , \textit{ the unbiased}$

(ii) $E\left(\hat{\theta}\right)-\theta>0 \quad , +ve \textit{ unbiased}$

(iii) $E\left(\hat{\theta}\right)-\theta<0 \quad , -ve \textit{ unbiased}$

Examples:

1. $\overline{X}$ is unbiased estimator of a population mean μ i.e. $E\left(\overline{X}\right)=\mu$
2. $\overline{X}$ is unbiased estimator of a Bernoulli distribution parameter P i.e. $E\left(\overline{X}\right)=P$
3. $\overline{X}$ is unbiased estimator of a normal distribution parameter (μ)
4. $\overline{X}$ is unbiased estimator of a Poisson distribution λ.

Theorem:

Show that the sample mean $\overline{X}$ is an unbiased estimator of a population mean μ.

Proof:

Let $X_1, X_2, ...X_n$ be a random sample of size n from the population with mean μ

We know that $$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Taking expected value on both side

$$E\left(\overline{X}\right) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right)$$

$$E\left(\overline{X}\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right)$$

$$E\left(\overline{X}\right) = \frac{1}{n} E\left(X_1 + X_2 + ... + X_n\right)$$

$$E\left(\overline{X}\right) = \frac{1}{n}\left[E\left(X_1\right) + E\left(X_2\right) + ... + E\left(X_n\right)\right] \quad \_\_\_(i)$$

As the random variables $X_1 + X_2 + ... + X_n$ are independent.

$$E\left(X_1\right) = E\left(X_2\right) = ... = E\left(X_n\right) = \mu$$

Put in (i) $\Rightarrow$ $$E\left(\overline{X}\right) = \frac{1}{n}\left[\mu + \mu + ... + \mu\right]$$

$$E\left(\overline{X}\right) = \frac{1}{n}\left[n\mu\right]$$

$E\left(\overline{X}\right) = \mu$ Hence proved.

Theorem:

Show that $E\left(S^2\right)=\left(\frac{n-1}{n}\right)\delta^2$ where S^2 is a sample variance of a random sample of size n and δ^2 is a variance of population.

Proof:

The sample variance S^2 can be written as

$$S^2=\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n}$$

$$S^2=\frac{1}{n}\left[\left(X_1-\overline{X}\right)^2+\left(X_2-\overline{X}\right)^2+....+\left(X_n-\overline{X}\right)^2\right]$$

$$E\left(S^2\right)=\frac{1}{n}\left[E\left(X_1-\overline{X}\right)^2+E\left(X_2-\overline{X}\right)^2+....+E\left(X_n-\overline{X}\right)^2\right] \_\_\_\_ (i)$$

Consider $$X_1-\overline{X}=X_1-\frac{\sum_{i=1}^{n}X_i}{n}$$

$$X_1-\overline{X}=\frac{nX_1-X_1-X_2-....-X_n}{n}$$

$$X_1-\overline{X}=\frac{1}{n}\left[(n-1)X_1-X_2-....-X_n\right]$$

Adding and subtracting (n–1)μ

$$X_1-\overline{X}=\frac{1}{n}\left[(n-1)X_1-X_2-....-X_n+(n-1)\mu-(n-1)\mu\right]$$

$$X_1-\overline{X}=\frac{1}{n}\left[(n-1)(X_1-\mu)-X_2-....+(n-1)\mu\right]$$

$$X_1-\overline{X}=\frac{1}{n}\left[(n-1)(X_1-\mu)-(X_2-\mu)-(X_3-\mu)-....-(X_n-\mu)\right]$$

$$\because (n-1)\mu=\mu+\mu+...+\mu(n-1)times$$

Taking square on both sides

$$\left(X_1-\overline{X}\right)^2=\frac{1}{n^2}\left[(n-1)(X_1-\mu)-(X_2-\mu)-(X_3-\mu)-....-(X_n-\mu)\right]^2$$

$$\left(X_1-\overline{X}\right)^2=\frac{1}{n^2}\left[(n-1)^2(X_1-\mu)^2+(X_2-\mu)^2+(X_3-\mu)^2+....+(X_n-\mu)^2-2\,product\ ter\min g\right]$$

Since the variables $X_1,X_2,.....X_n$ are independent. So, the expected values of that product term will be zero and we have the result as under

$$E\left(X_1-\overline{X}\right)^2=\frac{1}{n^2}\left[(n-1)^2E(X_1-\mu)^2+E(X_2-\mu)^2+E(X_3-\mu)^2+....+\mathrm{E}(X_n-\mu)^2\right]$$

$$E\left(X_1-\overline{X}\right)^2=\frac{1}{n^2}\left[(n-1)^2\delta^2+\delta^2+\delta^2+....+\delta^2\right]$$

$$E\left(X_1-\overline{X}\right)^2=\frac{1}{n^2}\left[(n-1)^2\delta^2+(n-1)\delta^2\right]$$

$$E\left(X_1-\overline{X}\right)^2=\frac{1}{n^2}\left[(n-1)(n-1+1)\delta^2\right]$$

$$E\left(X_1-\overline{X}\right)^2=\frac{1}{n^2}\left[(n-1).n.\delta^2\right]=\frac{n-1}{n}\delta^2$$

Similarly, for k

$$E\left(X_1-\overline{X}\right)^2=\left(\frac{n-1}{n}\right)\delta^2\ ;\ k=1,2,...,n$$

Eq. (i) becomes

$$E\left(S^2\right)=\frac{1}{n}\left[\left(\frac{n-1}{n}\right)\delta^2+\left(\frac{n-1}{n}\right)\delta^2+....+\left(\frac{n-1}{n}\right)\delta^2\right]$$

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

$$E\left(S^2\right)=\frac{1}{n}\left[(n-1)\delta^2\right]\quad\Rightarrow\quad E\left(S^2\right)=\left(\frac{n-1}{n}\right)\delta^2\ \text{proved}$$

Lecture # 04

Theorem: If $\overline{X}$ *and* S^2 are sample mean and sample variance defined as

$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$, $S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$ from a population of variance δ^2 mean μ , then show that $E\left(S^2\right) \neq \delta^2$

Proof: Let $X_1, X_2, ...X_n$ be a random sample of size n.

As we know that $S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$

Taking expected value on both side of the above equation.

$$E\left(S^2\right) = \frac{1}{n} E\left(\sum_{i=1}^{n}(X_i - \overline{X})^2\right)$$

$$E\left(S^2\right) = \frac{1}{n} E\left(\sum_{i=1}^{n}(X_i^2 + \overline{X} - 2X_i\overline{X})\right)$$

$$E\left(S^2\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i^2 + \overline{X^2}\sum_{i=1}^{n} -2\overline{X}\sum_{i=1}^{n} X_i\right)$$

$$E\left(S^2\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i^2 + n\overline{X^2} - 2n\overline{X^2}\right) \quad \because \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$E\left(S^2\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i^2 - n\overline{X^2}\right)$$

$$E\left(S^2\right) = \frac{1}{n}\left[\sum_{i=1}^{n} E\left(X_i^2\right) - nE\left(\overline{X^2}\right)\right] \quad \_\_\_\_\_(i)$$

$$\text{Now } \delta_{\overline{X}}^2 = E\left(\overline{X^2}\right) - \left(E\left(\overline{X}\right)\right)^2$$

$$\delta_{\overline{X}}^2 = E\left(\overline{X^2}\right) - \mu_{\overline{X}}^2$$

$$\Rightarrow E\left(\overline{X^2}\right) = \delta_{\overline{X}}^2 + \mu_{\overline{X}}^2$$

$$Also \;\; \delta^2 = E\left(X^2\right) - \left(E\left(X\right)\right)^2$$

$$\delta^2 = E\left(X^2\right) - \mu^2$$

$$E\left(X^2\right) = \delta^2 + \mu^2$$

Similarly, $E\left(X_i^2\right) = \delta^2 + \mu^2$

Since $X_1, X_2, ...X_n$ be a random variable. Put these values in (i)

$$E\left(S^2\right) = \frac{1}{n} E\left(\sum_{i=1}^{n}\left(\delta^2 + \mu^2\right) - n\left(\delta_{\overline{X}}^2 + \mu_{\overline{X}}^2\right)\right)$$

$$E\left(S^2\right) = \frac{1}{n} E\left[n\left(\delta^2 + \mu^2\right) - n\left(\frac{\delta^2}{n} + \mu^2\right)\right] \quad \because \delta_{\overline{X}} = \frac{\delta}{\sqrt{n}} \; , \mu_{\overline{X}} = \mu$$

$$E\left(S^2\right) = \frac{1}{n} E\left[n\delta^2 + n\mu^2 - \delta^2 - n\mu^2\right]$$

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

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

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

$$\Rightarrow E\left(S^2\right) \neq \delta^2$$

Theorem: Show that s^2 is an unbiased estimator of S^2 i.e. $\Rightarrow E\left(s^2\right)=\delta^2$

Proof: Let $X_1, X_2, \ldots\ldots X_n$ be a random sample of size n.

As we know that $s^2=\dfrac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$

Taking expected value on both side of the above equation.

$$E\left(S^2\right)=\frac{1}{n-1}E\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right)$$

$$(n-1)E\left(S^2\right)=E\left[\sum_{i=1}^{n}(X_i-\bar{X}+\mu-\mu)^2\right]$$

$$(n-1)E\left(S^2\right)=E\left[\sum_{i=1}^{n}\left((X_i-\mu)-(\bar{X}-\mu)\right)^2\right]$$

$$(n-1)E\left(S^2\right)=E\left[\sum_{i=1}^{n}\left\{(X_i-\mu)^2+(\bar{X}-\mu)^2-2(X_i-\mu)(\bar{X}-\mu)\right\}\right]$$

$$(n-1)E\left(S^2\right)=E\left[\sum_{i=1}^{n}(X_i-\mu)^2+(\bar{X}-\mu)^2\sum_{i=1}^{n}-2(\bar{X}-\mu)\sum_{i=1}^{n}(X_i-\mu)\right]$$

$$(n-1)E\left(S^2\right)=E\left[\sum_{i=1}^{n}(X_i-\mu)^2+n(\bar{X}-\mu)^2-2n(\bar{X}-\mu)^2\right]$$

$$(n-1)E\left(S^2\right)=E\left[\sum_{i=1}^{n}(X_i-\mu)^2-n(\bar{X}-\mu)^2\right]$$

$$(n-1)E\left(S^2\right)=E\left[(X_1-\mu)^2+(X_2-\mu)^2+\ldots+(X_n-\mu)^2-n(\bar{X}-\mu)^2\right]$$

As $X_1, X_2, \ldots\ldots X_n$ be a random variable. So,

$$(n-1)E\left(s^2\right)=E(X_1-\mu)^2+E(X_2-\mu)^2+\ldots+E(X_n-\mu)^2-nE(\bar{X}-\mu)^2$$

$$(n-1)E\left(s^2\right)=\delta^2+\delta^2+\delta^2+\ldots+\delta^2-\delta_{\bar{X}}^2$$

$$(n-1)E\left(s^2\right)=\delta^2+\delta^2+\delta^2+\ldots+\delta^2-\delta_{\overline{X}}^2$$

$$(n-1)E\left(s^2\right)=n\delta^2-n\frac{\delta^2}{n} \qquad \because \delta_{\overline{X}}=\frac{\delta}{\sqrt{n}}$$

$$(n-1)E\left(s^2\right)=n\delta^2-\delta^2$$

$$(n-1)E\left(s^2\right)=\delta^2(n-1)$$

$$\Rightarrow E\left(s^2\right)=\delta^2$$

Theorem: If s^2 is the variance of random sample size 'n' then we can also write as

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

Proof: As we know that $s^2=\dfrac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}$

$$s^2=\frac{\sum_{i=1}^{n}(X_i+\overline{X}^2-2\mathrm{X}_i\overline{X})}{n-1}$$

$$s^2=\frac{\sum_{i=1}^{n}X_i^2+\overline{X^2}\sum_{i=1}^{n}-2\overline{X}\sum_{i=1}^{n}X_i}{n-1}$$

$$s^2=\frac{\sum_{i=1}^{n}X_i^2+n\overline{X}^2-2n\overline{X}^2}{n-1} \qquad \because \frac{\sum_{i=1}^{n}X_i}{n}=\overline{X}$$

$$s^2=\frac{\sum_{i=1}^{n}X_i^2-n\overline{X}^2}{n-1}$$

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

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

Question: Suppose that a population consist five numbers such that 2,4,6,8,10. Draw all possible samples of size 2 with replacement then verify the following results.

(i) $E\left(\overline{X}\right) = \mu$

(ii) $E\left(S^2\right) \neq \delta^2$

(iii) $E\left(s^2\right) = \delta^2$

Solution: Here N = 5 , n = 5

Sample = $5^2 = 25$ $\qquad \mu = \frac{\sum_{i=1}^{5} X_i}{N} = \frac{2+4+6+8+10}{5} = 6$

$$\delta = \sqrt{\frac{\sum_{i=1}^{5}(X_i-\mu)^2}{N}} = \sqrt{\frac{(2-6)^2+(4-6)^2+(6-6)^2+(8-6)^2+(10-6)^2}{5}} = \sqrt{8}$$

$$\Rightarrow \quad \delta^2 = 8$$

Now samples

$(2,2),(2,4),(2,6),(2,8),(2,10),$
$(4,2),(4,4),(4,6),(4,8),(4,10),$
$(6,2),(6,4),(6,6),(6,8),(6,10),$
$(8,2),(8,4),(8,6),(8,8),(8,10),$
$(10,2),(10,4),(10,6),(10,8),(10,10)$

There corresponding means are $2,3,4,5,6$

$3,4,5,6,7$

$4,5,6,7,8$

$5,6,7,8,9$

$6,7,8,9,10$

$\overline{X}$	$f(\overline{X})$	$\overline{X}f(\overline{X})$	$\overline{X}^2$	$\overline{X}^2 f(\overline{X})$
2	1/25	2/25	4	4/25
3	2/25	6/25	9	18/25
4	3/25	12/25	16	48/25
5	4/25	20/25	25	100/25
6	5/25	30/25	36	180/25
7	4/25	28/25	49	196/25
8	3/25	24/25	64	192/25
9	2/25	18/25	81	162/25
10	1/25	10/25	100	100/25
		$\sum \overline{X}f(\overline{X}) = \frac{150}{25} = 6$		$\sum \overline{X^2}f(\overline{X}) = \frac{1000}{25} = 40$

$$\sum \overline{X}f(\overline{X}) = 6 = E(\overline{X})$$

$$\Rightarrow E(\overline{X}) = \mu$$

$$Also \quad S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$$

$$And \quad s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

Sample	$\overline{X}$	$S^2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$	$s^2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$
(2,2)	2	$\dfrac{(2-2)^2+(2-2)^2}{2}=0$	$\dfrac{(2-2)^2+(2-2)^2}{2-1}=0$
(2,4)	3	$\dfrac{(2-3)^2+(4-3)^2}{2}=1$	$\dfrac{(2-3)^2+(4-3)^2}{2-1}=2$
(2,6)	4	$\dfrac{(2-4)^2+(6-4)^2}{2}=4$	$\dfrac{(2-4)^2+(6-4)^2}{2-1}=8$
(2,8)	5	9	18
(2,10)	6	16	32
(4,2)	3	1	2
(4,4)	4	0	0
(4,6)	5	1	2
(4,8)	6	4	8
(4,10)	7	9	18
(6,2)	4	4	8
(6,4)	5	1	2
(6,6)	6	0	0
(6,8)	7	1	2
(6,10)	8	4	8
(8,2)	5	9	18
(8,4)	6	4	8
(8,6)	7	1	2
(8,8)	8	0	0
(8,10)	9	1	2
(10,2)	6	16	32
(10,4)	7	9	18
(10,6)	8	4	8
(10,8)	9	1	2
(10,10)	10	0	0

S^2	f	$S^2 f$	s^2	f	$s^2 f$
0	5	0	0	5	0
1	8	8	2	8	16
4	6	24	8	6	48
9	4	36	18	4	72
16	2	32	32	2	64
	$\sum f = 25$	$\sum S^2 f = 100$		$\sum f = 25$	$\sum s^2 f = 200$

As we know that

$$E\left(S^2\right) = \frac{\sum S^2 f}{\sum f} = \frac{100}{25} = 4$$

$$\Rightarrow E\left(S^2\right) \neq \delta^2$$

$$\text{Also } E\left(s^2\right) = \frac{\sum s^2 f}{\sum f} = \frac{200}{25} = 8$$

$$\Rightarrow E\left(s^2\right) = \delta^2$$

Question: Suppose that a population consist five numbers such that 2,4,6,8,10. Draw all possible samples of size 3 with replacement then verify the following results.

(i) $E\left(\overline{X}\right) = \mu$

(ii) $E\left(S^2\right) \neq \delta^2$

(iii) $E\left(s^2\right) = \delta^2$

Solution: Here N = 5 , n = 5

Samples = $5^3 = 125$

$$\mu = \frac{\sum_{i=1}^{5} X_i}{N} = \frac{2+4+6+8+10}{5} = 6$$

$$\delta=\sqrt{\frac{\sum_{i=1}^{5}(X_i-\mu)^2}{N}}=\sqrt{\frac{(2-6)^2+(4-6)^2+(6-6)^2+(8-6)^2+(10-6)^2}{5}}=\sqrt{8}$$

$$\Rightarrow \quad \delta^2=8$$

Samples

$(2,2,2),(2,2,4),(2,2,6),(2,2,8),(2,2,10),(2,4,2),(2,4,4),(2,4,6),(2,4,8),(2,4,10)$

$(2,6,2),(2,6,4),(2,6,6),(2,6,8),(2,6,10),(2,8,2),(2,8,4),(2,8,6),(2,8,8),(2,8,10)$

$(2,10,2),(2,10,4),(2,10,6),(2,10,8),(2,10,10)$

$(4,2,2),(4,2,4),(4,2,6),(4,2,8),(4,2,10),(4,4,2),(4,4,4),(4,4,6),(4,4,8),(4,4,10)$

$(4,6,2),(4,6,4),(4,6,6),(4,6,8),(4,6,10),(4,8,2),(4,8,4),(4,8,6),(4,8,8),(4,8,10)$

$(4,10,2),(4,10,4),(4,10,6),(4,10,8),(4,10,10)$

$(6,2,2),(6,2,4),(6,2,6),(6,2,8),(6,2,10),(6,4,2),(6,4,4),(6,4,6),(6,4,8),(6,4,10)$

$(6,6,2),(6,6,4),(6,6,6),(6,6,8),(6,6,10),(6,8,2),(6,8,4),(6,8,6),(6,8,8),(6,8,10)$

$(6,10,2),(6,10,4),(6,10,6),(6,10,8),(6,10,10)$

$(8,2,2),(8,2,4),(8,2,6),(8,2,8),(8,2,10),(8,4,2),(8,4,4),(8,4,6),(8,4,8),(8,4,10)$

$(8,6,2),(8,6,4),(8,6,6),(8,6,8),(8,6,10),(8,8,2),(8,8,4),(8,8,6),(8,8,8),(8,8,10)$

$(8,10,2),(8,10,4),(8,10,6),(8,10,8),(8,10,10)$

$(10,2,2),(10,2,4),(10,2,6),(10,2,8),(10,2,10),(10,4,2),(10,4,4),(10,4,6),(10,4,8),(10,4,10)$

$(10,6,2),(10,6,4),(10,6,6),(10,6,8),(10,6,10),(10,8,2),(10,8,4),(10,8,6),(10,8,8),(10,8,10)$

$(10,10,2),(10,10,4),(10,10,6),(10,10,8),(10,10,10)$

Their corresponding means are $2, 2.67, 3.33, 4, 4.67, 2.67, 3.33, 4, 4.67, 5.33$

$3.33, 4, 4.67, 5.33, 6, 4, 4.67, 5.33, 6, 6.67$

$4.67, 5.33, 6, 6.67, 7.33$

$2.67, 3.33, 4, 4.67, 5.33, 3.33, 4, 4.67, 5.33, 6$

$4, 4.67, 5.33, 6, 6.67, 4.67, 5.33, 6, 6.67, 7.33$

$5.33, 6, 6.67, 7.33, 8$

$3.33, 4, 4.67, 5.33, 6, 4, 4.67, 5.33, 6, 6.67$

$4.67, 5.33, 6, 6.67, 7.33, 5.33, 6, 6.67, 7.33, 8$

$6, 6.67, 7.33, 8, 8.67$

$4, 4.67, 5.33, 6, 6.67, 4.67, 5.33, 6, 6.67, 7.33$

$5.33, 6, 6.67, 7.33, 8, 6, 6.67, 7.33, 8, 8.67$

$6.67, 7.33, 8, 8.67, 9.33$

$4.67, 5.33, 6, 6.67, 7.33, 5.33, 6, 6.67, 7.33, 8$

$6, 6.67, 7.33, 8, 8.67, 6.67, 7.33, 8, 8.67, 9.33$

$7.33, 8, 8.67, 9.33, 10$

$\overline{X}$	$f\left(\overline{X}\right)$	$\overline{X}f\left(\overline{X}\right)$
2	1/125	2/125
2.67	3/125	8.01/125
3.33	6/125	19.98/125
4	10/125	40/125
4.67	15/125	70.05/125
5.33	18/125	95.94/125
6	19/125	114/125
6.67	18/125	120.06/125
7.33	15/125	109.95/125
8	10/125	80/125
8.67	6/125	52.02/125
9.33	3/125	27.99/125
10	1/125	10/125
		$\sum \overline{X}f\left(\overline{X}\right) = \frac{750}{125} = 6$

$$\sum \overline{X}f\left(\overline{X}\right) = 6 = E\left(\overline{X}\right) \qquad \Rightarrow E\left(\overline{X}\right) = \mu$$

Sample	$\overline{X}$	$S^2=\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n}$	$s^2=\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}$
(2,2,2)	2	$\frac{(2-2)^2+(2-2)^2+(2-2)^2}{3}=0$	$\frac{(2-2)^2+(2-2)^2+(2-2)^2}{3-1}=0$
(2,2,4)	2.67	0.8889	1.33
(2,2,6)	3.33	3.56	5.33
(2,2,8)	4	8	12
(2,2,10)	4.67	14.22	21.33
(2,4,2)	2.67	0.8889	1.33
(2,4,4)	3.33	0.8889	1.33
(2,4,6)	4	2.67	4
(2,4,8)	4.67	6.22	9.33
(2,4,10)	5.33	11.56	17.33
(2,6,2)	3.33	3.56	5.33
(2,6,4)	4	2.67	4
(2,6,6)	4.67	3.56	5.33
(2,6,8)	5.33	6.22	9.33
(2,6,10)	6	10.67	16
(2,8,2)	4	8	12
(2,8,4)	4.67	6.22	9.33
(2,8,6)	5.33	6.22	9.33
(2,8,8)	6	8	12
(2,8,10)	6.67	11.56	17.33
(2,10,2)	4.67	14.22	21.33
(2,10,4)	5.33	11.56	17.33
(2,10,6)	6	10.67	16
(2,10,8)	6.67	11.56	17.33
(2,10,10)	7.33	14.22	21.33

Sample	$\overline{X}$	$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$	$s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$
(4,2,2)	2.67	0.8889	1.33
(4,2,4)	3.33	0.8889	1.33
(4,2,6)	4	2.67	4
(4,2,8)	4.67	6.22	9.33
(4,2,10)	5.33	11.56	17.33
(4,4,2)	3.33	0.8889	1.33
(4,4,4)	4	0	0
(4,4,6)	4.67	0.8889	1.33
(4,4,8)	5.33	3.56	5.33
(4,4,10)	6	8	12
(4,6,2)	4	2.67	4
(4,6,4)	4.67	0.8889	1.33
(4,6,6)	5.33	0.8889	1.33
(4,6,8)	6	2.67	4
(4,6,10)	6.67	6.22	9.33
(4,8,2)	4.67	6.22	9.33
(4,8,4)	5.33	3.56	5.33
(4,8,6)	6	2.67	4
(4,8,8)	6.67	3.56	5.33
(4,8,10)	7.33	6.22	9.33
(4,10,2)	5.33	11.56	17.33
(4,10,4)	6	8	12
(4,10,6)	6.67	6.22	9.33
(4,10,8)	7.33	6.22	9.33
(4,10,10)	8	8	12

Sample	$\overline{X}$	$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$	$s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$
(6,2,2)	3.33	3.56	5.33
(6,2,4)	4	2.67	4
(6,2,6)	4.67	3.56	5.33
(6,2,8)	5.33	6.22	9.33
(6,2,10)	6	10.67	16
(6,4,2)	4	2.67	4
(6,4,4)	4.67	0.8889	1.33
(6,4,6)	5.33	0.8889	1.33
(6,4,8)	6	2.67	4
(6,4,10)	6.67	6.22	9.33
(6,6,2)	4.67	3.56	5.33
(6,6,4)	5.33	0.8889	1.33
(6,6,6)	6	0	0
(6,6,8)	6.67	0.8889	1.33
(6,6,10)	7.33	3.56	5.33
(6,8,2)	5.33	6.22	9.33
(6,8,4)	6	2.67	4
(6,8,6)	6.67	0.8889	1.33
(6,8,8)	7.33	0.8889	1.33
(6,8,10)	8	2.67	4
(6,10,2)	6	10.67	16
(6,10,4)	6.67	6.22	9.33
(6,10,6)	7.33	3.56	5.33
(6,10,8)	8	2.67	4
(6,10,10)	8.67	3.56	5.33

Sample	$\overline{X}$	$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$	$s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$
(8,2,2)	4	8	12
(8,2,4)	4.67	6.22	9.33
(8,2,6)	5.33	6.22	9.33
(8,2,8)	6	8	12
(8,2,10)	6.67	11.56	17.33
(8,4,2)	4.67	6.22	9.33
(8,4,4)	5.33	3.56	5.33
(8,4,6)	6	2.67	4
(8,4,8)	6.67	3.56	5.33
(8,4,10)	7.33	6.22	9.33
(8,6,2)	5.33	6.22	9.33
(8,6,4)	6	2.67	4
(8,6,6)	6.67	0.8889	1.33
(8,6,8)	7.33	0.8889	1.33
(8,6,10)	8	2.67	4
(8,8,2)	6	8	12
(8,8,4)	6.67	3.56	5.33
(8,8,6)	7.33	0.8889	1.33
(8,8,8)	8	0	0
(8,8,10)	8.67	0.8889	1.33
(8,10,2)	6.67	11.56	17.33
(8,10,4)	7.33	6.22	9.33
(8,10,6)	8	2.67	4
(8,10,8)	8.67	0.8889	1.33
(8,10,10)	9.33	0.8889	1.33

Sample	$\overline{X}$	$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$	$s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$
(10,2,2)	4.67	14.22	21.33
(10,2,4)	5.33	11.56	17.33
(10,2,6)	6	10.67	16
(10,2,8)	6.67	11.56	17.33
(10,2,10)	7.33	14.22	21.33
(10,4,2)	5.33	11.56	17.33
(10,4,4)	6	8	12
(10,4,6)	6.67	6.22	9.33
(10,4,8)	7.33	6.22	9.33
(10,4,10)	8	8	12
(10,6,2)	6	10.67	16
(10,6,4)	6.67	6.22	9.33
(10,6,6)	7.33	3.56	5.33
(10,6,8)	8	2.67	4
(10,6,10)	8.67	3.56	5.33
(10,8,2)	6.67	11.56	17.33
(10,8,4)	7.33	6.22	9.33
(10,8,6)	8	2.67	4
(10,8,8)	8.67	0.8889	1.33
(10,8,10)	9.33	0.8889	1.33
(10,10,2)	7.33	14.22	21.33
(10,10,4)	8	8	12
(10,10,6)	8.67	3.56	5.33
(10,10,8)	9.33	0.8889	1.33
(10,10,10)	10	0	0

S^2	f	$S^2 f$	s^2	f	$s^2 f$
0	5	0	0	5	0
0.8889	24	21.3336	1.33	24	31.92
2.67	18	48.06	4	18	72
3.56	18	64.08	5.33	18	95.94
6.22	24	149.28	9.33	24	223.94
8	12	96	12	12	144
10.67	6	64.02	16	6	96
11.56	12	138.72	17.33	12	207.96
14.22	6	85.35	21.33	6	127.98
	$\sum f = 125$	$\sum S^2 f = 666.8136$			$\sum s^2 f = 999.72$

As we know that

$$E\left(S^2\right) = \frac{\sum S^2 f}{\sum f} = \frac{666.8136}{125} = 5.33$$

$$\Rightarrow E\left(S^2\right) \neq \delta^2$$

Also $E\left(s^2\right) = \frac{\sum s^2 f}{\sum f} = \frac{999.72}{125} = 7.99 = 8$

$$\Rightarrow E\left(s^2\right) = \delta^2$$

Lecture # 05

Question: If X is a random variable has binomial distribution, then show that the proportional $\frac{X}{n}$ is an unbiased estimator of parameter ‘p’.

Solution: Here we have to show that

$$E\left(\frac{X}{n}\right) = p$$

As we know that binomial distribution

$$E(X) = np$$

$$\Rightarrow E\left(\frac{X}{n}\right) = \frac{1}{n}E(X)$$

$$E\left(\frac{X}{n}\right) = \frac{1}{n}np$$

$$\Rightarrow E\left(\frac{X}{n}\right) = p$$

Question: Suppose that sample mean $\overline{X}$ of a sample from population is an unbiased estimator of ‘θ’ if $\overline{X}$ has density function

$$f(x,\theta) = \begin{cases} \frac{1}{\theta}e^{-\frac{x}{\theta}} & ; 0 < x < \infty \\ 0 & ;\ otherwise \end{cases}$$

Solution: Here we have to show that

$$E(\overline{X}) = \theta$$

As we know that

$$E(\overline{X}) = \mu_{\overline{X}}$$

$$Also\ \ E(X) = \mu$$

$$\Rightarrow \; E(\overline{X}) = E(X)$$

$$\Rightarrow \; We\ show\ E(X) = \theta$$

Now by definition of continuous random variable

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

$$E(X) = \int_{-\infty}^{\infty} x \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx$$

$$E(X) = \int_{-\infty}^{0} x \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx + \int_{0}^{\infty} x \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx$$

$$E(X) = 0 + \int_{0}^{\infty} x \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx$$

$$E(X) = \frac{1}{\theta}\left[x \frac{e^{-\frac{x}{\theta}}}{-\frac{1}{\theta}} \Bigg|_{0}^{\infty} - \int_{0}^{\infty} \frac{e^{-\frac{x}{\theta}}}{-\frac{1}{\theta}} .1\, dx \right]$$

$$E(X) = \frac{1}{\theta}\left[0 + \theta \int_{0}^{\infty} e^{-\frac{x}{\theta}} dx \right]$$

$$E(X) = \frac{1}{\theta}\left[\theta . \frac{e^{-\frac{x}{\theta}}}{-\frac{1}{\theta}} \Bigg|_{0}^{\infty} \right]$$

$$E(X) = -\theta\left[e^{-\frac{\infty}{\theta}} - e^{0} \right] = -\theta[0-1] = \theta$$

$$E(X) = \theta = E(\overline{X}) \quad proved$$

Question: Suppose that a population of five numbers such that 1,3,5,7,9. Draw all possible sample of size 2 with replacement and without replacement. Then verify the following results

(i) $E\left(\overline{X}\right)=\mu$

(ii) $E\left(S^2\right)\neq\delta^2$

(iii) $E\left(s^2\right)=\delta^2$

Solution: Here N = 5 , n = 2

$$\text{Samples} = 5^2 = 25 \quad , \quad \mu=\frac{\sum_{i=1}^{5}X_i}{N}=\frac{1+3+5+7+9}{5}=5$$

$$\delta=\sqrt{\frac{\sum_{i=1}^{5}(X_i-\mu)^2}{N}}=\sqrt{\frac{(1-5)^2+(3-5)^2+(5-5)^2+(7-5)^2+(9-5)^2}{5}}=\sqrt{8}$$

$$\Rightarrow \delta^2=8$$

Samples

$$(1,1),(1,3),(1,5),(1,7),(1,9)$$

$$(3,1),(3,3),(3,5),(3,7),(3,9)$$

$$(5,1),(5,3),(5,5),(5,7),(5,9)$$

$$(7,1),(7,3),(7,5),(7,7),(7,9)$$

$$(9,1),(9,3),(9,5),(9,7),(9,9)$$

There corresponding means are 1,2,3,4,5

2,3,4,5,6

3,4,5,6,7

4,5,6,7,8

5,6,7,8,9

Sample	$\overline{X}$	$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$	$s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$
(1,1)	1	0	0
(1,3)	2	1	2
(1,5)	3	4	8
(1,7)	4	9	18
(1,9)	5	16	32
(3,1)	2	1	2
(3,3)	3	0	0
(3,5)	4	1	2
(3,7)	5	4	8
(3,9)	6	9	18
(5,1)	3	4	8
(5,3)	4	1	2
(5,5)	5	0	0
(5,7)	6	1	2
(5,9)	7	4	8
(7,1)	4	9	18
(7,3)	5	4	8
(7,5)	6	1	2
(7,7)	7	0	0
(7,9)	8	1	2
(9,1)	5	16	32
(9,3)	6	9	18
(9,5)	7	4	8
(9,7)	8	1	2
(9,9)	9	0	0

S^2	f	$S^2 f$	s^2	f	$s^2 f$
0	5	0	0	5	0
1	8	8	2	8	16
4	6	24	8	6	48
9	4	36	18	4	72
16	2	32	32	2	64
	$\sum f = 25$	$\sum S^2 f = 100$			$\sum s^2 f = 200$

$\overline{X}$	$f(\overline{X})$	$\overline{X}f(\overline{X})$
1	1/25	1/25
2	2/25	4/25
3	3/25	9/25
4	4/25	16/25
5	5/25	25/25
6	4/25	24/25
7	3/25	21/25
8	2/25	16/25
9	1/25	9/25
		$\sum \overline{X}f(\overline{X}) = \frac{125}{25} = 5$

$$\sum \overline{X}f(\overline{X}) = 5 = E(\overline{X})$$

$$\Rightarrow E(\overline{X}) = \mu$$

$$E(S^2) = \frac{\sum S^2 f}{\sum f} = \frac{100}{25} = 4 \neq 8$$

$$\Rightarrow E(S^2) \neq \delta^2$$

$$\text{Also } E(s^2) = \frac{\sum s^2 f}{\sum f} = \frac{200}{25} = 8$$

$$\Rightarrow E(s^2) = \delta^2$$

Now without replacement:

$Sample\ {}^{N}C_{n} = {}^{5}C_{2} = 10$

Samples: (1,3),(1,5),(1,7),(1,9),(3,5),(3,7),(3,9),(5,7),(5,9),(7,9)

Their corresponding means are

2,3,4,5,4,5,6,6,7,8

$\overline{X}$	$f(\overline{X})$	$\overline{X}f(\overline{X})$
2	1/10	2/10
3	1/10	3/10
4	2/10	8/10
5	2/10	10/10
6	2/10	12/10
7	1/10	7/10
8	1/10	8/10
		$\sum \overline{X}f(\overline{X}) = \frac{50}{10} = 5$

$$\sum \overline{X}f(\overline{X}) = 5 = E(\overline{X})$$

$$\Rightarrow E(\overline{X}) = \mu$$

Sample	$\overline{X}$	$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$	$s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$
(1,3)	2	1	1
(1,5)	3	4	8
(1,7)	4	9	18
(1,9)	5	16	32
(3,5)	4	1	2
(3,7)	5	4	8
(3,9)	6	9	18
(5,7)	6	1	2
(5,9)	7	4	8
(7,9)	8	1	2

S^2	f	$S^2 f$	s^2	f	$s^2 f$
1	4	4	2	4	8
4	3	12	8	3	24
4	2	18	18	2	36
16	1	16	32	1	32
	$\sum f = 10$	$\sum S^2 f = 50$			$\sum s^2 f = 100$

$$E\left(S^2\right)=\frac{\sum S^2 f}{\sum f}=\frac{50}{10}=5\neq 8$$

$$\Rightarrow\ E\left(S^2\right)\neq \delta^2$$

Also $E\left(s^2\right)=\frac{\sum s^2 f}{\sum f}=\frac{100}{10}=10$

For without replacement

$$E\left(s^2\right)=\frac{N}{N-1}\left(\delta^2\right)$$

$$10=\frac{5}{5-1}(8)$$

$$10=\frac{5}{4}(8)$$

$$10=10$$

$$\Rightarrow\ E\left(s^2\right)=\delta^2$$

Lecture # 06

Normal Distribution density function N(μ,δ²):

$$f\left(x,\delta^2\right)=\frac{1}{\sqrt{2\pi}\delta}e^{-\frac{1}{2}\left(\frac{x-\mu}{\delta}\right)^2}$$

Range $-\infty < x < \infty$

Now for N(0,θ)

$$f\left(x,\theta\right)=\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{1}{2}\left(\frac{x}{\sqrt{\theta}}\right)^2} \qquad \because\ \delta^2=\theta \ \Rightarrow \delta=\sqrt{\theta}$$

$$f\left(x,\theta\right)=\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{1}{2}\left(\frac{x^2}{\theta}\right)}$$

$$f\left(x,\theta\right)=\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{x^2}{2\theta}}$$

Theorem: Let $X_1, X_2, X_3,...X_n$ be random sample of size 'n' form a normal distribution N(0,θ) then show that $\frac{\sum X_i^2}{n}$ is an unbiased estimator of parameter θ.

Proof: As we have μ = 0 and δ² = θ . So, normal distribution density function can be defined as

$$f\left(x,\theta\right)=\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{x^2}{2\theta}}$$

Here we have to show that

$$E\left(\frac{\sum_{i=1}^{n} X_i^2}{n}\right)=\theta$$

$$\text{Now } E\left(X^2\right)=\int_{-\infty}^{\infty} x^2\ f\left(x\right)dx$$

$$E\left(X^2\right)=\int_{-\infty}^{\infty} x^2\ \frac{1}{\sqrt{2\pi\theta}}e^{-\frac{x^2}{2\theta}}dx$$

$$E\left(X^2\right)=\int_{-\infty}^{0} x^2\ \frac{1}{\sqrt{2\pi\theta}}e^{-\frac{x^2}{2\theta}}dx+\int_{0}^{\infty} x^2\ \frac{1}{\sqrt{2\pi\theta}}e^{-\frac{x^2}{2\theta}}dx$$

$$E\left(X^2\right)=\frac{1}{\sqrt{2\pi\theta}}\left[\int_{-\infty}^{0} x^2\ e^{-\frac{x^2}{2\theta}}dx+\int_{0}^{\infty} x^2\ e^{-\frac{x^2}{2\theta}}dx\right]\_\_\_\_\left(A\right)$$

$$Now\int_{-\infty}^{0} x^2\ e^{-\frac{x^2}{2\theta}}dx$$

Here it is along the negative region. We will put $x = -u \Rightarrow x^2 = u^2$

$$dx = -du$$

$$u\to\infty\ as\,x\to-\infty\ \ and\ u\to 0\ as\,x\to 0$$

$$\int_{-\infty}^{0} x^2\ e^{-\frac{x^2}{2\theta}}dx=\int_{\infty}^{0} u^2\ e^{-\frac{u^2}{2\theta}}\left(-du\right)$$

$$=\int_{0}^{\infty} u^2\ e^{-\frac{u^2}{2\theta}}du$$

Replace u by x

$$=\int_{0}^{\infty} x^2\ e^{-\frac{x^2}{2\theta}}dx$$

Put in (A)

$$E\left(X^2\right)=\frac{1}{\sqrt{2\pi\theta}}\left[\int_{0}^{\infty} x^2\ e^{-\frac{x^2}{2\theta}}dx+\int_{0}^{\infty} x^2\ e^{-\frac{x^2}{2\theta}}dx\right]$$

$$E\left(X^2\right)=\frac{1}{\sqrt{2\pi\theta}}\left[2\int_0^\infty x^2\, e^{-\frac{x^2}{2\theta}}dx\right]$$

$$E\left(X^2\right)=\frac{2}{\sqrt{2\pi\theta}}\int_0^\infty x^2\, e^{-\frac{x^2}{2\theta}}dx$$

$$E\left(X^2\right)=\frac{(-2\theta)}{\sqrt{2\pi\theta}}\int_0^\infty x^2\, e^{-\frac{x^2}{2\theta}}\left(\frac{-x}{\theta}\right)dx$$

$$E\left(X^2\right)=\frac{(-2\theta)}{\sqrt{2\pi\theta}}\left[x.e^{-\frac{x^2}{2\theta}}\Big|_0^\infty-\int_0^\infty e^{-\frac{x^2}{2\theta}}dx\right]$$

$$E\left(X^2\right)=\frac{(-2\theta)}{\sqrt{2\pi\theta}}\left[0-\int_0^\infty e^{-\frac{x^2}{2\theta}}dx\right]$$

$$E\left(X^2\right)=\theta\left[2\int_0^\infty \frac{1}{\sqrt{2\pi\theta}}e^{-\frac{x^2}{2\theta}}dx\right]$$

Here in the square brackets it is the area under the normal curve with $\mu = 0$ and $\delta^2 = \theta$ so its area is always unity.

$$\Rightarrow 2\int_0^\infty \frac{1}{\sqrt{2\pi\theta}}e^{-\frac{x^2}{2\theta}}dx=1$$

$$\Rightarrow E\left(X^2\right)=\theta$$

$$\Rightarrow E\left(X_i^2\right)=\theta \qquad \because \textit{No change}$$

Because R.H.S is independent w.r.t index 'i'

Now $$\Rightarrow E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)=\frac{1}{n}E\left(\sum_{i=1}^{n}X_i^2\right)$$

$$E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)=\frac{1}{n}E\left(X_1^2+X_2^2+\ldots+X_n^2\right)$$

$$E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)=\frac{1}{n}E\left(X_1^2+X_2^2+\ldots+X_n^2\right)$$

$$E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)=\frac{1}{n}\left[E\left(X_1^2\right)+E\left(X_2^2\right)+E\left(X_3^2\right)+\ldots+E\left(X_n^2\right)\right]$$

$$E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)=\frac{1}{n}\left[\theta+\theta+\theta+\ldots+\theta\right]$$

$$E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)=\frac{1}{n}\left[n\theta\right]$$

$$E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)=\theta$$

Efficiency:

It is possible that a parameter has more than one estimator. So, from these estimators only one estimator will be efficient as compare to the others.

Consider $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are the two estimators of a same parameter θ, then if their variance $\widehat{\theta}_1$ i.e. Var($\widehat{\theta}_1$) is less than Var($\widehat{\theta}_2$) then $\widehat{\theta}_1$ is more efficient than $\widehat{\theta}_2$.

It means that efficiency is a comparison of the variances of the estimators. It can also be written as

$$Efficiency=E=\frac{Var\left(\widehat{\theta}_2\right)}{Var\left(\widehat{\theta}_1\right)}$$

is the ratio of the measure of a relative efficiency of $\widehat{\theta}_1$ w.r.t $\widehat{\theta}_2$.

If E < 1 then $\widehat{\theta}_2$ is more efficient than $\widehat{\theta}_1$

$$\frac{Var\left(\widehat{\theta}_2\right)}{Var\left(\widehat{\theta}_1\right)}<1$$

$$\Rightarrow \; Var\left(\widehat{\theta}_2\right) < Var\left(\widehat{\theta}_1\right)$$

If E > 1 then $\widehat{\theta}_1$ is more efficient than $\widehat{\theta}_2$.

$$\frac{Var\left(\widehat{\theta}_2\right)}{Var\left(\widehat{\theta}_1\right)} > 1$$

$$\Rightarrow \; Var\left(\widehat{\theta}_1\right) < Var\left(\widehat{\theta}_2\right)$$

If $\hat{\theta}$ is the biased estimator of the parameter θ then to check the efficiency of biased estimator, we make the efficiency comparison on the basis of mean square error instead of variance written as

$$\text{Mean Square Error } \hat{\theta} = \text{MSE}(\hat{\theta}) = \text{E}\left[\hat{\theta} - \theta\right]^2$$

$$= E\left[\hat{\theta} - E\left(\hat{\theta}\right) + E\left(\hat{\theta}\right) - \theta\right]^2$$

$$= E\left[\left(\hat{\theta} - E\left(\hat{\theta}\right)\right) + \left(E\left(\hat{\theta}\right) - \theta\right)\right]^2$$

$$= E\left[\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)^2 + \left(E\left(\hat{\theta}\right) - \theta\right)^2 + 2\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)\left(E\left(\hat{\theta}\right) - \theta\right)\right]$$

$$= E\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)^2 + E\left(E\left(\hat{\theta}\right) - \theta\right)^2 + 2E\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)\left(E\left(\hat{\theta}\right) - \theta\right)$$

$$= E\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)^2 + E\left(E\left(\hat{\theta}\right) - \theta\right)^2 + 0$$

$$= Var\left(\hat{\theta}\right) + \left(Biased\right)^2$$

i.e. Mean Square Error of $\hat{\theta}$ is equal to the variance of the estimator plus squared biased.

Note: An estimator which has less error then it will be more efficient as compared to the other.

Example: Suppose that X_1, X_2, X_3 are the random sample of size 3 from a population with mean μ and variance δ^2. Also consider the following two estimators of the mean μ

$$T_1 = \frac{X_1 + X_2 + X_3}{3} \quad , \; T_2 = \frac{X_1 + 2X_2 + X_3}{4}$$

Find which estimator is more efficient.

Solution: First of all, we check the unbiasedness of T_1 and T_2.

$$E(T_1) = E\left(\frac{X_1 + X_2 + X_3}{3}\right) = \frac{1}{3}\left[E(X_1) + E(X_2) + E(X_3)\right]$$

$$E(T_1) = \frac{1}{3}[\mu + \mu + \mu] = \frac{1}{3}(3\mu)$$

$$E(T_1) = \mu$$

$$E(T_2) = E\left(\frac{X_1 + 2X_2 + X_3}{4}\right) = \frac{1}{4}\left[E(X_1) + 2E(X_2) + E(X_3)\right]$$

$$E(T_2) = \frac{1}{4}[\mu + 2\mu + \mu] = \frac{1}{4}(4\mu)$$

$$E(T_2) = \mu$$

So T_1 and T_2 are the unbiased estimator of mean μ.

Now we have to find their variance.

$$Var(T_1) = Var\left(\frac{X_1 + X_2 + X_3}{3}\right) = \frac{1}{9}\left[Var(X_1) + Var(X_2) + Var(X_3)\right]$$

$$Var(T_1) = \frac{1}{9}\left[\delta^2 + \delta^2 + \delta^2\right] = \frac{1}{9}\left(3\delta^2\right)$$

$$Var(T_1) = \frac{\delta^2}{3}$$

$$Var(T_2) = Var\left(\frac{X_1 + 2X_2 + X_3}{4}\right) = \frac{1}{16}\left[Var(X_1) + 4Var(X_2) + Var(X_3)\right]$$

$$Var(T_2) = \frac{1}{16}\left[\delta^2 + 4\delta^2 + \delta^2\right] = \frac{1}{16}\left(6\delta^2\right)$$

$$Var(T_2) = \frac{3}{8}\delta^2$$

$$E = \frac{Var(T_2)}{Var(T_1)} = \frac{\frac{3}{8}\delta^2}{\frac{\delta^2}{3}} = \frac{9}{8} > 1$$

T_1 is more efficient than T_2.

Example: Suppose that X_1, X_2, X_3, X_4 be a random sample of size 4 from a $N\left(\mu, \delta^2\right)$. A person wishes to estimate the mean by using either of the following two estimators of mean μ.

$$T_1 = \frac{X_1 + X_2 + X_3 + X_4}{4}\ ,\ T_2 = \frac{X_1 + 3X_2 + 2X_3 + X_4}{7}$$

First of all, we check the unbiasedness of T_1 and T_2.

$$E(T_1) = E\left(\frac{X_1 + X_2 + X_3 + X_4}{4}\right) = \frac{1}{4}\left[E(X_1) + E(X_2) + E(X_3) + E(X_4)\right]$$

$$E(T_1) = \frac{1}{4}[\mu + \mu + \mu + \mu] = \frac{1}{4}(4\mu)$$

$$E(T_1) = \mu$$

$$E(T_2) = E\left(\frac{X_1 + 3X_2 + 2X_3 + X_4}{7}\right) = \frac{1}{7}\left[E(X_1) + 3E(X_2) + 2E(X_3) + E(X_4)\right]$$

$$E(T_2) = \frac{1}{7}[\mu + 3\mu + 2\mu + \mu] = \frac{1}{7}(7\mu)$$

$$E(T_2) = \mu$$

So T_1 and T_2 are the unbiased estimator of mean μ.

Now we have to find their variance.

$$Var(T_1) = Var\left(\frac{X_1 + X_2 + X_3 + X_4}{4}\right) = \frac{1}{16}\left[Var(X_1) + Var(X_2) + Var(X_3) + Var(X_4)\right]$$

$$Var(T_1) = \frac{1}{16}\left[\delta^2 + \delta^2 + \delta^2 + \delta^2\right] = \frac{1}{16}\left(4\delta^2\right)$$

$$Var(T_1) = \frac{\delta^2}{4}$$

$$Var(T_2) = Var\left(\frac{X_1 + 3X_2 + 2X_3 + X_4}{7}\right) = \frac{1}{49}\left[Var(X_1) + 9Var(X_2) + 4Var(X_3) + Var(X_4)\right]$$

$$Var(T_2) = \frac{1}{49}\left[\delta^2 + 9\delta^2 + 4\delta^2 + \delta^2\right] = \frac{1}{49}\left(15\delta^2\right)$$

$$Var(T_2) = \frac{15\delta^2}{49}$$

$$E = \frac{Var(T_2)}{Var(T_1)} = \frac{\frac{15}{49}\delta^2}{\frac{\delta^2}{4}} = \frac{60}{49} > 1$$

T_1 is more efficient than T_2.

$\Rightarrow$ estimator T_1 is preferred.

Lecture # 07

Consistency:

An estimator is said to be consistence if the sample statistics to be used as estimator becomes closer and closer to the population being estimated as the sample size ‘n’ increase.

Notes: A consistent estimator may or may not be unbiased.

Criteria for consistency:

If $\hat{\theta}$ be a sample statistic and θ be population parameter then $\hat{\theta}$ is a consistent estimator of θ if the following condition holds

$$Variance\left(\hat{\theta}\right) \to 0 \quad when \quad n \to \infty$$

Question: Show that $\overline{X}$ is a consistent estimator of mean μ.

Solution: We know that the variance of sample mean

$$\delta_{\overline{X}} = \frac{\delta}{\sqrt{n}}$$

$$\delta_{\overline{X}}^2 = \frac{\delta^2}{n}$$

$$\Rightarrow Var\left(\overline{X}\right) = \frac{\delta^2}{n}$$

Taking Limit n → ∞ on the both side of above expression

$$\lim_{n\to\infty} Var\left(\overline{X}\right) = \lim_{n\to\infty} \frac{\delta^2}{n}$$

$$\lim_{n\to\infty} Var\left(\overline{X}\right) = \delta^2 . \lim_{n\to\infty} \left(\frac{1}{n}\right) = \delta^2 (0)$$

$$\lim_{n\to\infty} Var\left(\overline{X}\right) = 0$$

Hence the $\overline{X}$ is a consistent estimator of mean μ.

Question: If a random variable X is a binomial distribution b(x ; x ,p) then show that the sample proportion $\frac{X}{n}$ is unbiased and consistent estimator of parameter p.

Solution: We know that $E(X) = np$

We have to show that $E\left(\frac{X}{n}\right) = p$

$$Now \quad E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n}(np) = p$$

$$Also \ \ Var(X) = npq$$

$$Now \quad Var\left(\frac{X}{n}\right) = \frac{1}{n^2}Var(X) = \frac{1}{n^2}(npq) = \frac{pq}{n}$$

Taking Limit $n \to \infty$ on the both side of above expression

$$\lim_{n\to\infty} Var\left(\frac{X}{n}\right) = \lim_{n\to\infty} \frac{pq}{n}$$

$$\lim_{n\to\infty} Var\left(\frac{X}{n}\right) = pq \lim_{n\to\infty} \left(\frac{1}{n}\right)$$

$$\lim_{n\to\infty} Var\left(\frac{X}{n}\right) = pq(0) = 0$$

Hence, $\frac{X}{n}$ is unbiased and consistent estimator of parameter p.

Question: Show that the sample mean $\overline{X}$ of random sample of size n from a density function

$$f(x,\theta) = \begin{cases} \frac{1}{\theta}e^{\frac{-x}{\theta}} & ; if \ \ 0 < x < \infty \\ 0 & ; \ otherwise \end{cases}$$

is an unbiased and consistent estimator of parameter θ.

Solution: First we show that $E(\overline{X}) = \theta$

As we know that $E(\overline{X}) = \mu_{\overline{X}}$

$$E(X) = \mu$$

$$E(\overline{X}) = E(X)$$

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

$$E(X) = \int_{-\infty}^{\infty} x \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx$$

$$E(X) = \int_{-\infty}^{0} x \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx + \int_{0}^{\infty} x \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx$$

$$E(X) = 0 + \int_{0}^{\infty} x \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx$$

$$E(X) = \frac{1}{\theta}\left[x \frac{e^{-\frac{x}{\theta}}}{-\frac{1}{\theta}} \Bigg|_{0}^{\infty} - \int_{0}^{\infty} \frac{e^{-\frac{x}{\theta}}}{-\frac{1}{\theta}} .1 dx \right]$$

$$E(X) = \frac{1}{\theta}\left[0 + \theta \int_{0}^{\infty} e^{-\frac{x}{\theta}} dx \right]$$

$$E(X) = \frac{1}{\theta}\left[\theta . \frac{e^{-\frac{x}{\theta}}}{-\frac{1}{\theta}} \Bigg|_{0}^{\infty} \right]$$

$$E(X) = -\theta\left[e^{-\frac{\infty}{\theta}} - e^{0} \right] = -\theta[0 - 1] = \theta$$

$$Now \quad Var(X) = E(X^2) - (E(X))^2$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \frac{1}{\theta} e^{-\frac{x}{\theta}} dx$$

$$E(X^2) = \int_{-\infty}^{0} x^2 \frac{1}{\theta} e^{-\frac{x}{\theta}} dx + \int_{0}^{\infty} x^2 \frac{1}{\theta} e^{-\frac{x}{\theta}} dx$$

$$E(X^2) = 0 + \int_{0}^{\infty} x^2 \frac{1}{\theta} e^{-\frac{x}{\theta}} dx$$

$$E(X^2) = -\int_{0}^{\infty} x^2 e^{-\frac{x}{\theta}} \left(-\frac{1}{\theta}\right) dx$$

$$E(X^2) = \left[x^2 e^{-\frac{x}{\theta}} \Big|_0^{\infty} - \int_0^{\infty} e^{-\frac{x}{\theta}} .2x\, dx \right]$$

$$E(X^2) = -\left[0 - 2\int_0^{\infty} x e^{-\frac{x}{\theta}} dx \right]$$

$$E(X^2) = 2\int_0^{\infty} x e^{-\frac{x}{\theta}} dx$$

$$E(X^2) = -2\theta \int_0^{\infty} x e^{-\frac{x}{\theta}} \left(-\frac{1}{\theta}\right) dx$$

$$E(X^2) = -2\theta \left[x e^{-\frac{x}{\theta}} \Big|_0^{\infty} - \int_0^{\infty} e^{-\frac{x}{\theta}} dx \right]$$

$$E\left(X^2\right)=-2\theta\left[0-\left.\frac{e^{-\frac{x}{\theta}}}{\frac{-1}{\theta}}\right|_0^\infty\right]=-2\theta\left[\theta\left(e^{-\infty}-e^0\right)\right]$$

$$E\left(X^2\right)=-2\theta\left[\theta\left(0-1\right)\right]=2\theta^2$$

$$\Rightarrow Var\left(X\right)=2\theta^2-\left(\theta\right)^2=2\theta^2-\theta^2$$

$$\Rightarrow Var\left(X\right)=\theta^2$$

$$\Rightarrow Var\left(X\right)=\theta^2=\delta^2$$

$$\Rightarrow Var\left(X\right)=\theta^2=\delta_{\bar{X}}^2=\frac{\delta^2}{n}$$

Taking Limit n $\rightarrow \infty$ on the both side of above expression

$$\lim_{n\to\infty} Var\left(X\right)=\lim_{n\to\infty}\frac{\delta^2}{n}$$

$$\lim_{n\to\infty} Var\left(X\right)=\delta^2\lim_{n\to\infty}\left(\frac{1}{n}\right)$$

$$\lim_{n\to\infty} Var\left(X\right)=\delta^2\left(0\right)=0$$

Hence the density function is an unbiased and consistent estimator of parameter θ.

Lecture # 08

Sufficiency:

An estimator $\hat{\theta}$ is said to be a sufficient estimator of θ if $\hat{\theta}$ has all information relevant to parameter θ.

OR

An estimator is said to be a sufficient, if the statistics used an estimator uses all the information i.e. contain in the sample. Any statistics i.e. not computed from all values in the sample is not a sufficient estimator.

Example 1:

The sample mean $\overline{X}$ is a sufficient estimator of μ. This implies that $\overline{X}$ contains all the information in the sample relative to the estimation of the population parameter μ and no other estimator such as the sample median, mode etc. calculated from same sample can add any information concerning μ.

Example 2:

The sample proportion $\hat{P}$ is also a sufficient estimator of the population proportion P.

Neyman-Fisher Factorization Criterion for sufficiency:

If $X_1, X_2, X_3, \ldots X_n$ be a random sample of a random variable X, whose distribution depends on the unknown values of the parameter θ then $\hat{\theta}$ is said to be the sufficient estimator of θ iff

$$f\left(x_1, x_2, x_3, \ldots x_n, \theta\right) = f\left(x_1, \theta\right).f\left(x_2, \theta\right)\ldots f\left(x_n, \theta\right)$$

$$f\left(x_1, x_2, \ldots x_n, \theta\right) = g\left(\hat{\theta}, \theta\right).h\left(x_1, x_2, \ldots x_n\right)$$

Where $g\left(\hat{\theta}, \theta\right)$ is a function depends on the estimators $\hat{\theta}$ and θ. Where $h\left(x_1, x_2, \ldots x_n\right)$ does not depends on θ.

Question: Let $X_1, X_2, ... X_n$ be a random sample from a density function $f(x,p) = px^{p-1}$; 0 < x < 1 , p > 0. Then show that the product $x_1.x_2...x_n$ be a sufficient estimator of parameter p.

Solution: As the joint probability function is defined by

$$f(x_1,x_2,...x_n,p) = f(x_1,p).f(x_2,p)...f(x_n,\text{p})$$

$$f(x_1,x_2,...x_n,p) = px_1^{p-1}.px_2^{p-2}...px_n^{p-n}$$

$$f(x_1,x_2,...x_n,p) = p^n(x_1.x_2...x_n)^{p-1}$$

$$f(x_1,x_2,...x_n,p) = p^n(x_1.x_2...x_n)^p.(x_1.x_2...x_n)^{-1}$$

$$f(x_1,x_2,...x_n,p) = g(x_n,p)h(x_1.x_2...x_n)$$

$$\textit{Where } g(x_n,p) = p^n(x_1.x_2...x_n)^p$$

$$h(x_1.x_2...x_n) = (x_1.x_2...x_n)^{-1} = \frac{1}{(x_1.x_2...x_n)}$$

As the given joint probability function is factorized into two function therefore Neyman-Fisher Factorization criteria is satisfied which implies $x_1.x_2...x_n$ is a sufficient estimator of parameter p.

Question: Let $X_1, X_2, ... X_n$ be a random sample of a binomial density function $f(x,p) = p^x(1-p)^{1-x}; x = 0,1$ Then show that $\sum_{i=1}^{n} x_i$ be a sufficient estimator of parameter p.

Solution: As the joint probability function is defined by

$$f(x_1,x_2,...x_n,p) = f(x_1,p).f(x_2,p)...f(x_n,\text{p})$$

$$f(x_1,x_2,...x_n,p) = p^{x_1}(1-p)^{1-x_1}.p^{x_2}(1-p)^{1-x_2}...p^{x_n}(1-p)^{1-x_n}$$

$$f(x_1,x_2,...x_n,p) = p^{\sum_{i=1}^{n} x_i}(1-p)^{1-\sum_{i=1}^{n} x_i}.1$$

$$Where\ \ g\left(\sum x_i, p\right) = p^{\sum_{i=1}^{n} x_i}\left(1-p\right)^{1-\sum_{i=1}^{n} x_i}$$

$$h\left(x_1.x_2...x_n\right) = 1$$

As the given joint probability function is factorized into two function therefore Neyman-Fisher factorization criteria is satisfied which implies $\sum_{i=1}^{n} x_i$ is a sufficient estimator of parameter p.

Question: Let $X_1, X_2, ... X_n$ be a random sample from a density function $f\left(x,\mu\right) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}(x-\mu)^2}$ $-\infty < x < \infty$ Then show that $\sum_{i=1}^{n} x_i$ be a sufficient estimator of parameter μ.

Solution: As the joint probability function is defined as

$$f\left(x_1, x_2, ... x_n, \mu\right) = f\left(x_1, \mu\right).f\left(x_2, \mu\right)...f\left(x_n, \mu\right)$$

$$f\left(x_1, x_2, ... x_n, \mu\right) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}(x_1-\mu)^2} . \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}(x_2-\mu)^2} ... \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}(x_n-\mu)^2}$$

$$f\left(x_1, x_2, ... x_n, \mu\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{\frac{-1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$f\left(x_1, x_2, ... x_n, \mu\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{\frac{-1}{2}\sum_{i=1}^{n}\left(x_i^2+\mu^2-2\mu x_i\right)}$$

$$f\left(x_1, x_2, ... x_n, \mu\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{\frac{-1}{2}\left(\sum_{i=1}^{n} x_i^2 + n\mu^2 - 2\mu\sum_{i=1}^{n} x_i\right)}$$

$$f\left(x_1, x_2, ... x_n, \mu\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{\frac{-1}{2}\left(n\mu^2 - 2\mu\sum_{i=1}^{n} x_i\right)}.e^{\frac{-1}{2}\left(\sum_{i=1}^{n} x_i^2\right)} \quad \because \sum_{i=1} = n$$

$$f\left(x_1, x_2, ... x_n, \mu\right) = g\left(\sum x_i, \mu\right) h\left(x_1.x_2...x_n\right)$$

$$where\ g\left(\sum x_i, \mu\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{\frac{-1}{2}\left(n\mu^2 - 2\mu\sum_{i=1}^{n} x_i\right)}$$

$$h(x_1.x_2 \ldots x_n) = e^{\frac{-1}{2}\left(\sum_{i=1}^{n} x_i^2\right)}$$

As the given joint probability function is factorized into two function. Therefore, Neyman-Fisher factorization criteria is satisfied.

which implies $\sum_{i=1}^{n} x_i$ is a sufficient estimator of parameter μ.

Question: Let $X_1, X_2, \ldots X_n$ be a random sample from a Poisson density function $f(x,\lambda) = e^{-\lambda}\lambda^x \ ; x = 0,1$ Then show that $\sum_{i=1}^{n} x_i$ be a sufficient estimator of parameter λ.

Solution: As the joint probability function is defined by

$$f(x_1, x_2, \ldots x_n, \lambda) = f(x_1, \lambda).f(x_2, \lambda) \ldots f(x_n, \lambda)$$

$$f(x_1, x_2, \ldots x_n, \lambda) = e^{-\lambda}\lambda^{x_1}.e^{-\lambda}\lambda^{x_2} \ldots e^{-\lambda}\lambda^{x_n}$$

$$f(x_1, x_2, \ldots x_n, \lambda) = e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}.1$$

$$f(x_1, x_2, \ldots x_n, \lambda) = g\left(\sum x_i, \lambda\right)h(x_1.x_2 \ldots x_n)$$

$$Where\ \ g\left(\sum x_i, \lambda\right) = e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}$$

$$h(x_1.x_2 \ldots x_n) = 1$$

As the given joint probability function is factorized into two function therefore Neyman-Fisher factorization criteria is satisfied which implies $\sum_{i=1}^{n} x_i$ is a sufficient estimator of parameter λ.

Question: Let $X_1, X_2, \ldots X_n$ be a random sample from a density function $f(x,\theta) = \frac{1}{\theta} e^{\frac{-x}{\theta}}$; $0 < x < \infty$ Then show that $\sum_{i=1}^{n} x_i$ be a sufficient estimator of parameter θ.

Solution: As the joint probability function is defined by

$$f(x_1, x_2, \ldots x_n, \theta) = f(x_1, \theta).f(x_2, \theta)\ldots f(x_n, \theta)$$

$$f(x_1, x_2, \ldots x_n, \theta) = \frac{1}{\theta} e^{\frac{-x_1}{\theta}} . \frac{1}{\theta} e^{\frac{-x_2}{\theta}} \ldots . \frac{1}{\theta} e^{\frac{-x_n}{\theta}}$$

$$f(x_1, x_2, \ldots x_n, \theta) = \left(\frac{1}{\theta}\right)^n e^{\frac{-1}{\theta}\sum_{i=1}^{n} x_i} .1$$

$$f(x_1, x_2, \ldots x_n, \theta) = g\left(\sum x_i, \theta\right) h(x_1.x_2 \ldots x_n)$$

$$Where \;\; g\left(\sum x_i, \theta\right) = \left(\frac{1}{\theta}\right)^n e^{\frac{-1}{\theta}\sum_{i=1}^{n} x_i}$$

$$h(x_1.x_2 \ldots x_n) = 1$$

As the given joint probability function is factorized into two function therefore Neyman-Fisher factorization criteria is satisfied which implies $\sum_{i=1}^{n} x_i$ is a sufficient estimator of parameter θ.

Lecture 09

Methods of point estimator:

A point estimator of a parameter can be obtained by several methods by we shall consider the following three methods only

(i) The method of Maximum likelihood estimator.
(ii) The method of moments (introduced in 18^{th} century, it is oldest method)
(iii) The method of least squares

The method of Maximum likelihood estimator:

In general, it was introduced in the early 20^{th} century and it was given by Ronald A. Fisher (1890-1962). The method is very useful in the early age of life.

Likelihood function:

Let $X_1, X_2, ... X_n$ be a random sample from a distribution having probability function. Probability density function of $X_1, X_2, ... X_n$ is

$$L(\theta, x_1, x_2, ... x_n) = L(\theta) = f(\theta, x_1).f(\theta, x_2)...f(\theta, x_n)$$

Probability density function of unknown parameter 'θ' is called a likelihood function of the sample and its mathematical form is given above.

Maximum likelihood Estimator:

The value of the parameter 'θ' that maximize the likelihood function $L(\theta, x_1, x_2, ... x_n) = L(\theta)$ is called a maximum likelihood estimator of the parameter 'θ' and it is denoted by $\hat{\theta}$. As to find the maximum likelihood estimator we take a first derivative of likelihood function and setting it against zero. As a result, we obtain a single value and that value we replace in the second derivative of likelihood function. That value is called stationary value of the function and if its value is less than zero then likelihood function at that value is maximum.

Question: For a binomial population the sample proportion is the Maximum likelihood estimator (MLE) of the population parameter p.

Solution: As the likelihood function of given sample is

$$L(p)=\binom{n}{x}p^x(1-p)^{n-x}$$

Taking ln on both sides

$$\ln(L(p))=\ln\left[\binom{n}{x}p^x(1-p)^{n-x}\right]$$

$$\ln(L(p))=\ln\binom{n}{x}+\ln p^x+\ln(1-p)^{n-x}$$

$$\ln(L(p))=\ln\binom{n}{x}+x\ln p+(n-x)\ln(1-p)$$

Diff. w.r.t 'p'

$$\frac{L'(p)}{L(p)}=0+x.\frac{1}{p}+(n-x).\frac{1}{(1-p)}(-1)$$

$$\frac{L'(p)}{L(p)}=\frac{x}{p}-\left(\frac{n-x}{1-p}\right) \quad \_\_\_(i)$$

$$\text{Put } \frac{d}{dp}L(p)=0$$

$$\frac{x}{p}-\left(\frac{n-x}{1-p}\right)=0$$

$$\frac{x(1-p)-p(n-x)}{p(1-p)}=0$$

$$x-xp-np+xp=0$$

$$x=np$$

$$p=\frac{x}{n}$$

Again diff. equation (i) w.r.t 'p'

$$\frac{L''(p)}{L(p)}-\frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2}=-\frac{x}{p^2}-(n-x)\left[-(1-p)^{-2}(-1)\right]$$

$$\frac{L''(p)}{L(p)}-\frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2}=-\frac{x}{p^2}-\frac{n-x}{(1-p)^2}$$

Put $p=\frac{x}{n}$

$$\frac{L''(p)}{L(p)}-\frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2}=-\frac{x}{\left(\frac{x}{n}\right)^2}-\frac{n-x}{\left(1-\frac{x}{n}\right)^2}$$

$$\frac{L''(p)}{L(p)}-\frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2}=-\frac{n^2}{x}-\frac{n-x}{\frac{(n-x)^2}{n^2}}$$

$$\frac{L''(p)}{L(p)}-\frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2}=-\frac{n^2}{x}-\frac{n^2}{n-x}=\frac{-n^2(n-x)-n^2x}{(n-x)x}$$

$$\frac{L''(p)}{L(p)}-\frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2}=\frac{-n^3+n^2x-n^2x}{(n-x)x}$$

$$\frac{L''(p)}{L(p)}-\frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2}=\frac{-n^3}{(n-x)x}<0$$

$$\Rightarrow\frac{d^2}{dp^2}\left(L(p)\right)<0$$

The likelihood function attains its maximum value at p = x/n. Therefore, the maximum likelihood estimator (MLE) for parameter p is $p = \frac{X}{n}$ or $\hat{p} = \frac{x}{n}$

Question: Suppose that X is a Bernoulli's random variable with parameter 'p' given a random sample of X. Then find the Maximum likelihood estimator (MLE) of parameter p.

Solution: As the likelihood function of given sample is

$$L(p) = p^x (1-p)^{1-x}$$

Taking ln on both sides

$$\ln(L(p)) = \ln\left[p^x (1-p)^{1-x}\right]$$

$$\ln(L(p)) = \ln p^x + \ln(1-p)^{1-x}$$

$$\ln(L(p)) = x\ln p + (1-x)\ln(1-p)$$

Diff. w.r.t 'p'

$$\frac{L'(p)}{L(p)} = x.\frac{1}{p} + (1-x).\frac{1}{(1-p)}(-1)$$

$$\frac{L'(p)}{L(p)} = \frac{x}{p} - \left(\frac{1-x}{1-p}\right) \quad \_\_\_(i)$$

Put $\frac{d}{dp}L(p) = 0$

$$\frac{x}{p} - \left(\frac{1-x}{1-p}\right) = 0$$

$$\frac{x(1-p) - p(1-x)}{p(1-p)} = 0$$

$$x - xp - p + xp = 0$$

$$p = x$$

Again diff. equation (i) w.r.t 'p'

$$\frac{L''(p)}{L(p)}-\frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2}=-\frac{x}{p^2}-(1-x)\left[-(1-p)^{-2}(-1)\right]$$

$$\frac{L''(p)}{L(p)}-\frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2}=-\frac{x}{p^2}-\frac{1-x}{(1-p)^2}$$

Put $p=x$

$$\frac{L''(p)}{L(p)}-\frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2}=-\frac{x}{(x)^2}-\frac{1-x}{(1-x)^2}$$

$$\frac{L''(p)}{L(p)}-\frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2}=-\frac{1}{x}-\frac{1}{1-x}$$

$$\frac{L''(p)}{L(p)}-\frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2}=\frac{-(1-x)-x}{(1-x)x}$$

$$\frac{L''(p)}{L(p)}-\frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2}=\frac{-1+x-x}{(1-x)x}$$

$$\frac{L''(p)}{L(p)}-\frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2}=\frac{-1}{(1-x)x}<0$$

$$\Rightarrow\frac{d^2}{dp^2}\left(L(p)\right)<0$$

The likelihood function attains its maximum value at p = x. Therefore, the maximum likelihood estimator (MLE) for parameter p is $p=X \ or \ \hat{p}=x$

Lecture # 10

Question: Suppose that X is a Bernoulli random variable with parameter p given a random sample of n observation of X then find the Maximum Likelihood Estimator (MLE) of parameter p.

Solution: As the likelihood function of given sample is

$$L(p) = p^{x}(1-p)^{n-x}$$

$$L(p, x_1, x_2, \ldots x_n) = L(p) = L(x_1, \mathrm{p}).L(x_2, \mathrm{p})\ldots L(x_n, \mathrm{p})$$

$$L(p) = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

Taking ln on both sides

$$\ln(L(p)) = \ln p^{\sum_{i=1}^{n} x_i} + \ln(1-p)^{nN-\sum_{i=1}^{n} x_i}$$

$$\ln(L(p)) = \sum_{i=1}^{n} x_i \ln p + \left(n - \sum_{i=1}^{n} x_i\right)\ln(1-p)$$

Diff. w.r.t 'p'

$$\frac{L'(p)}{L(p)} = \sum_{i=1}^{n} x_i . \frac{1}{p} + \left(n - \sum_{i=1}^{n} x_i\right)\frac{1}{(1-p)}(-1)$$

$$\frac{L'(p)}{L(p)} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{(1-p)} \quad \_\_\_(i)$$

Put $\frac{d}{dp}L(p) = 0$

$$\frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{(1-p)} = 0$$

$$\frac{(1-p)\sum_{i=1}^{n}x_i - p\left(n-\sum_{i=1}^{n}x_i\right)}{p(1-p)} = 0$$

$$\sum_{i=1}^{n}x_i - p\sum_{i=1}^{n}x_i - pn + p\sum_{i=1}^{n}x_i = 0$$

$$\sum_{i=1}^{n}x_i - pn = 0 \Rightarrow \quad p = \frac{\sum_{i=1}^{n}x_i}{n}$$

Again diff. equation (i) w.r.t 'p'

$$\frac{L''(p)}{L(p)} - \frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2} = \sum_{i=1}^{n}x_i.\left(-\frac{1}{p^2}\right) - \left(n-\sum_{i=1}^{n}x_i\right)\left[-(1-p)^{-2}(-1)\right]$$

$$\frac{L''(p)}{L(p)} - \frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2} = \sum_{i=1}^{n}x_i.\left(-\frac{1}{p^2}\right) - \left(n-\sum_{i=1}^{n}x_i\right)\frac{1}{(1-p)^2}$$

Put $p = \frac{\sum_{i=1}^{n}x_i}{n}$

$$\frac{L''(p)}{L(p)} - \frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2} = \sum_{i=1}^{n}x_i.\left(-\frac{n^2}{\left(\sum_{i=1}^{n}x_i\right)^2}\right) - \left(n-\sum_{i=1}^{n}x_i\right)\frac{1}{\left(1-\frac{\sum_{i=1}^{n}x_i}{n}\right)^2}$$

$$\frac{L''(p)}{L(p)} - \frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2} = \left(-\frac{n^2}{\sum_{i=1}^{n}x_i}\right) - \left(n-\sum_{i=1}^{n}x_i\right)\left(\frac{n^2}{\left(n-\sum_{i=1}^{n}x_i\right)^2}\right)$$

$$\frac{L''(p)}{L(p)}-\frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2}=-\frac{n^2}{\sum_{i=1}^{n}x_i}-\frac{n^2}{\left(n-\sum_{i=1}^{n}x_i\right)}$$

$$\frac{L''(p)}{L(p)}-\frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2}=\frac{-n^2\left(n-\sum_{i=1}^{n}x_i\right)-n^2\left(\sum_{i=1}^{n}x_i\right)}{\sum_{i=1}^{n}x_i\left(n-\sum_{i=1}^{n}x_i\right)}$$

$$\frac{L''(p)}{L(p)}-\frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2}=\frac{-n^3+n^2\sum_{i=1}^{n}x_i-n^2\sum_{i=1}^{n}x_i}{\sum_{i=1}^{n}x_i\left(n-\sum_{i=1}^{n}x_i\right)}$$

$$\frac{L''(p)}{L(p)}-\frac{\left(L'(p)\right)^2}{\left(L(p)\right)^2}=\frac{-n^3}{\sum_{i=1}^{n}x_i\left(n-\sum_{i=1}^{n}x_i\right)}<0$$

$$\Rightarrow \frac{d^2}{dp^2}L(p)<0$$

He likelihood function attains its maximum value at $p=\frac{\sum_{i=1}^{n}x_i}{n}$ .Therefore, the MLE for parameter p is $\hat{p}=\frac{\sum_{i=1}^{n}x_i}{n}$.

Question: Let $X_1, X_2, \ldots X_n$ be a random sample from the Poisson distribution $f(x,\lambda)=\frac{e^{-\lambda}\lambda^x}{x!}$ Find the MLE of λ

Solution: As the likelihood function of given sample is

$$L(\lambda,x_1,x_2,\ldots x_n)=L(\lambda)=f(\lambda,x_1).f(\lambda,x_2)\ldots f(\lambda,x_n)$$

$$L(\lambda)=\frac{e^{-\lambda}\lambda^{x_1}}{x_1!}.\frac{e^{-\lambda}\lambda^{x_2}}{x_2!}\ldots\frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$

$$L(\lambda)=\frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n}x_i}}{x_1!.\,x_2!.x_3!\ldots x_n!}$$

Taking ln on both side

$$\ln\,L(\lambda)=\ln\left[\frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n}x_i}}{x_1!.\,x_2!.x_3!\ldots x_n!}\right]$$

$$\ln\,L(\lambda)=\ln\left(e^{-n\lambda}\lambda^{\sum_{i=1}^{n}x_i}\right)-\ln\left[x_1!.\,x_2!.x_3!\ldots x_n!\right]$$

$$\ln\,L(\lambda)=\ln e^{-n\lambda}+\ln\lambda^{\sum_{i=1}^{n}x_i}-\ln\left[x_1!.\,x_2!.x_3!\ldots x_n!\right]$$

$$\ln\,L(\lambda)=-n\lambda+\sum_{i=1}^{n}x_i\ln\lambda-\ln\left[x_1!.\,x_2!.x_3!\ldots x_n!\right]$$

Diff. w.r.t 'λ'

$$\frac{L'(\lambda)}{L(\lambda)}=-n+\sum_{i=1}^{n}x_i.\frac{1}{\lambda}\qquad \_\_\_\_\_\_(i)$$

$$Put\ \frac{d}{d\lambda}L(\lambda)=0\Rightarrow\quad -n+\sum_{i=1}^{n}x_i.\frac{1}{\lambda}=0$$

$$\frac{-n\lambda+\sum_{i=1}^{n}x_i}{\lambda}=0$$

$$\Rightarrow\quad -n\lambda+\sum_{i=1}^{n}x_i=0$$

$$\Rightarrow \quad n\lambda = \sum_{i=1}^{n} x_i \qquad \Rightarrow \quad \lambda = \frac{\sum_{i=1}^{n} x_i}{n}$$

Diff. Eq (i) w.r.t 'λ'

$$\frac{L''(\lambda)}{L(\lambda)} - \frac{\left(L'(\lambda)\right)^2}{\left(L(\lambda)\right)^2} = \sum_{i=1}^{n} x_i . \left(\frac{-1}{\lambda^2}\right)$$

Put $\lambda = \frac{\sum_{i=1}^{n} x_i}{n}$

$$\Rightarrow \frac{L''(\lambda)}{L(\lambda)} - \frac{\left(L'(\lambda)\right)^2}{\left(L(\lambda)\right)^2} = \sum_{i=1}^{n} x_i . \left(\frac{-1}{\left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2}\right)$$

$$\frac{L''(\lambda)}{L(\lambda)} - \frac{\left(L'(\lambda)\right)^2}{\left(L(\lambda)\right)^2} = \sum_{i=1}^{n} x_i . \left(\frac{-n^2}{\left(\sum_{i=1}^{n} x_i\right)^2}\right)$$

$$\frac{L''(\lambda)}{L(\lambda)} - \frac{\left(L'(\lambda)\right)^2}{\left(L(\lambda)\right)^2} = \frac{-n^2}{\sum_{i=1}^{n} x_i} < 0$$

$\Rightarrow \frac{d^2}{d\lambda^2} L(\lambda) < 0$ The likelihood function has maximum value at $\lambda = \frac{\sum_{i=1}^{n} x_i}{n}$.

Therefore, the MLE for parameter λ is $\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}$

Question: Suppose that X is a random variable with density function $f(x,\theta)=\theta e^{-\theta x}$ where θ > 0 , x > 0. What is the MLE for θ based the sample variable on n observation.

Solution: $$L(\theta,x_1,x_2,...x_n)=L(\theta)=f(\theta,x_1).f(\theta,x_2)...f(\theta,x_n)$$

$$L(\theta)=\theta e^{-\theta x_1}.\theta e^{-\theta x_2}...\theta e^{-\theta x_n}$$

$$L(\theta)=\theta^n e^{-\theta\sum_{i=1}^{n}x_i}$$

Taking ln on both side

$$\ln L(\theta)=\ln\left[\theta^n e^{-\theta\sum_{i=1}^{n}x_i}\right]$$

$$\ln L(\theta)=\ln\theta^n+\ln e^{-\theta\sum_{i=1}^{n}x_i}$$

$$\ln L(\theta)=n\ln\theta+\left(-\theta\sum_{i=1}^{n}x_i\right)\ln e$$

$$\ln L(\theta)=n\ln\theta-\theta\sum_{i=1}^{n}x_i$$

Diff. w.r.t 'θ'

$$\frac{L'(\theta)}{L(\theta)}=\frac{n}{\theta}-\sum_{i=1}^{n}x_i \quad \_\_\_\_\_\_(i)$$

$$Put\ \frac{d}{d\theta}L(\theta)=0 \Rightarrow \frac{n}{\theta}-\sum_{i=1}^{n}x_i=0$$

$$\frac{n-\theta\sum_{i=1}^{n}x_i}{\theta}=0$$

$$\Rightarrow \quad n=\theta\sum_{i=1}^{n}x_i$$

$$\Rightarrow \quad \theta = \frac{n}{\sum_{i=1}^{n} x_i}$$

Diff. Eq (i) w.r.t 'θ'

$$\frac{L''(\theta)}{L(\theta)} - \frac{\left(L'(\theta)\right)^2}{\left(L(\theta)\right)^2} = \frac{-n}{\theta^2}$$

Put $\theta = \frac{n}{\sum_{i=1}^{n} x_i}$

$$\frac{L''(\theta)}{L(\theta)} - \frac{\left(L'(\theta)\right)^2}{\left(L(\theta)\right)^2} = \frac{-n}{\left(\frac{n}{\sum_{i=1}^{n} x_i}\right)^2}$$

$$\frac{L''(\theta)}{L(\theta)} - \frac{\left(L'(\theta)\right)^2}{\left(L(\theta)\right)^2} = \frac{-\left(\sum_{i=1}^{n} x_i\right)^2}{n} < 0$$

$$\Rightarrow \frac{d^2}{d\theta^2} L(\theta) < 0$$

The likelihood function has maximum value at $\theta = \frac{n}{\sum_{i=1}^{n} x_i}$. Therefore, the MLE for parameter θ is $\hat{\theta} = \frac{n}{\sum_{i=1}^{n} x_i}$

Lecture # 11

Normal distribution:

A continuous random variable having bell-shaped curve is called a Normal random variable. A normal random variable X with mean μ and variance δ^2 has the density function written as

$$N\left(x,\mu,\delta^2\right)=\frac{1}{\sqrt{2\pi}\,\delta}e^{\frac{-1}{2}\left(\frac{x-\mu}{\delta}\right)^2}\quad ;-\infty<x<\infty$$

Theorem:

Show that area under the normal curve and above the X-axis is always one. **OR** If f(x) is a density function from a normal distribution then show that

$$\int_{-\infty}^{\infty} f(x)\,dx=1$$

Proof: As given f(x) is a density function from a $N\left(x,\mu,\delta^2\right)$. So,

$$\int_{-\infty}^{\infty} f(x)\,dx=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\,\delta}e^{\frac{-1}{2}\left(\frac{x-\mu}{\delta}\right)^2}dx$$

$$\int_{-\infty}^{\infty} f(x)\,dx=\frac{1}{\sqrt{2\pi}\,\delta}\int_{-\infty}^{\infty}e^{\frac{-1}{2}\left(\frac{x-\mu}{\delta}\right)^2}dx \qquad \_\_\_\_\_\_(i)$$

$$\text{Put } \frac{x-\mu}{\delta}=t \quad \Rightarrow x-\mu=\delta t$$

$$dx=\delta dt \quad \& \quad t\to\pm\infty \text{ as } x\to\pm\infty$$

Put in (i) ⇒

$$\int_{-\infty}^{\infty} f(x)\,dx=\frac{1}{\sqrt{2\pi}\,\delta}\int_{-\infty}^{\infty}e^{\frac{-1}{2}t^2}\delta dt$$

$$\int_{-\infty}^{\infty} f(x)\,dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{\frac{-1}{2}t^2}dt$$

As $\int\limits_{-\infty}^{\infty} e^{\frac{-1}{2}t^2}\,dt$ is an even function. So,

$$\int\limits_{-\infty}^{\infty} e^{\frac{-1}{2}t^2}\,dt = 2\int\limits_{0}^{\infty} e^{\frac{-1}{2}t^2}\,dt$$

$$\int\limits_{-\infty}^{\infty} f(x)\,dx = \frac{2}{\sqrt{2\pi}}\int\limits_{0}^{\infty} e^{\frac{-1}{2}t^2}\,dt \qquad \_\_\_\_\_(ii)$$

Again put $\frac{1}{2}t^2 = z \Rightarrow t^2 = 2z$

$$\Rightarrow \sqrt{t} = \sqrt{2z}$$

$$\Rightarrow dz = t\,dt$$

$$\Rightarrow dt = \frac{dz}{t} = \frac{dz}{\sqrt{2z}}$$

$$z \to 0 \text{ as } t \to 0$$
$$z \to \infty \text{ as } t \to \infty$$

Put in (ii) $\Rightarrow$ $$\int\limits_{-\infty}^{\infty} f(x)\,dx = \frac{2}{\sqrt{2\pi}}\int\limits_{0}^{\infty} e^{-z}\frac{dz}{\sqrt{2z}}$$

$$\int\limits_{-\infty}^{\infty} f(x)\,dx = \frac{2}{\sqrt{2\pi}.\sqrt{2}}\int\limits_{0}^{\infty} e^{-z}.z^{\frac{-1}{2}}\,dz$$

$$\int\limits_{-\infty}^{\infty} f(x)\,dx = \frac{2}{2\sqrt{\pi}}\left|\overline{\frac{1}{2}}\right. \qquad \because \int\limits_{0}^{\infty} e^{-z}.z^{\frac{-1}{2}}\,dz = \left|\overline{\frac{1}{2}}\right.$$

$$\int\limits_{-\infty}^{\infty} f(x)\,dx = \frac{1}{\sqrt{\pi}}.\sqrt{\pi} \qquad \because \left|\overline{\frac{1}{2}}\right. = \sqrt{\pi}$$

$$\int\limits_{-\infty}^{\infty} f(x)\,dx = 1$$ Hence proved.

Theorem:

Show that the parameter μ and δ^2 are the mean and variance of a normal distribution.

Proof: First we prove E(X) = μ and then Var(x) = δ^2

(i) As we know that

$$Mean = E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

$$E(X) = \int_{-\infty}^{\infty} x.\frac{1}{\sqrt{2\pi}\,\delta}.e^{\frac{-1}{2}\left(\frac{x-\mu}{\delta}\right)^2} dx$$

$$E(X) = \frac{1}{\sqrt{2\pi}\,\delta}\int_{-\infty}^{\infty} x.e^{\frac{-1}{2}\left(\frac{x-\mu}{\delta}\right)^2} dx \quad \_\_\_\_(i)$$

$$Put\ \frac{x-\mu}{\delta} = t \quad \Rightarrow\ x-\mu = \delta t \quad \Rightarrow\ x = \mu + \delta t$$

$$d\,x = \delta dt \quad \&\ \text{t} \to \pm\infty \text{ as } x \to \pm\infty$$

Put in (i) $\Rightarrow$ $$E(X) = \frac{1}{\sqrt{2\pi}\,\delta}\int_{-\infty}^{\infty} (\mu + \delta t) e^{\frac{-1}{2}t^2}\delta dt$$

$$E(X) = \frac{1}{\sqrt{2\pi}}\left[\mu\int_{-\infty}^{\infty} e^{\frac{-1}{2}t^2} dt + \delta\int_{-\infty}^{\infty} t\, e^{\frac{-1}{2}t^2} dt\right]$$

$$E(X) = \mu.\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\frac{-1}{2}t^2} dt + \frac{\delta}{\sqrt{2\pi}}\int_{-\infty}^{\infty} t\, e^{\frac{-1}{2}t^2} dt$$

$$\because \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\frac{-1}{2}t^2} dt = 1$$

$$E(X) = \mu - \frac{\delta}{\sqrt{2\pi}}\left[\int_{-\infty}^{\infty} e^{\frac{-1}{2}t^2}(-t)\, dt\right]$$

$$E(X) = \mu - \frac{\delta}{\sqrt{2\pi}}\left| e^{\frac{-1}{2}t^2}\right|_{-\infty}^{\infty} = \mu - \frac{\delta}{\sqrt{2\pi}}\left[e^{-\infty} - e^{-\infty}\right]$$

$$E(X) = \mu - 0 = \mu$$

$$\Rightarrow\ E(X) = \mu$$

(ii) $$Var(X)=\int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx$$

$$Var(X)=\int_{-\infty}^{\infty}(x-\mu)^2\frac{1}{\sqrt{2\pi}\,\delta}.e^{\frac{-1}{2}\left(\frac{x-\mu}{\delta}\right)^2}dx$$

$$Var(X)=\frac{1}{\sqrt{2\pi}\,\delta}\int_{-\infty}^{\infty}(x-\mu)^2.e^{\frac{-1}{2}\left(\frac{x-\mu}{\delta}\right)^2}dx \qquad \_\_\_\_(i)$$

$$Put\ \frac{x-\mu}{\delta}=t \quad \Rightarrow\ x-\mu=\delta t$$

$$dx=\delta dt \quad \&\ \ t\to\pm\infty\ \text{ as }\ x\to\pm\infty$$

Put in (i) $\Rightarrow$ $$Var(X)=\frac{1}{\sqrt{2\pi}\,\delta}\int_{-\infty}^{\infty}(\delta t)^2 e^{\frac{-1}{2}t^2}\delta dt$$

$$Var(X)=\frac{\delta^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}t^2 e^{\frac{-1}{2}t^2}dt$$

$$Var(X)=\frac{\delta^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}t.e^{\frac{-1}{2}t^2}.tdt$$

$$Var(X)=\frac{-\delta^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}t.e^{\frac{-1}{2}t^2}.(-t)dt$$

$$Var(X)=\frac{-\delta^2}{\sqrt{2\pi}}\left[t.e^{\frac{-1}{2}t^2}\Big|_{-\infty}^{\infty}-\int_{-\infty}^{\infty}e^{\frac{-1}{2}t^2}dt\right]$$

$$Var(X)=\frac{-\delta^2}{\sqrt{2\pi}}\left[0-\int_{-\infty}^{\infty}e^{\frac{-1}{2}t^2}dt\right]$$

$$Var(X)=\frac{\delta^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{\frac{-1}{2}t^2}dt$$

$$Var(X)=\delta^2.1=\delta^2$$

Moment generating function: (m.g.f)

The m.g.f of X with respect to origin is

$$M_0(t)=E\left(e^{tX}\right)=\frac{1}{\sqrt{2\pi}\,\delta}\int_{-\infty}^{\infty}e^{tX}.e^{\frac{-1}{2}\left(\frac{x-\mu}{\delta}\right)^2}dx$$

The m.g.f of X with respect to mean μ is

$$M_\mu(t)=E\left(e^{t(X-\mu)}\right)$$

Question: Find the moment generating function (m.g.f) of a normal distribution about a mean ‘μ’.

Solution: As the m.g.f about mean μ and normal distribution is

$$M_\mu(t)=E\left(e^{t(x-\mu)}\right)$$

$$M_\mu(t)=\int_{-\infty}^{\infty}e^{t(x-\mu)}.\frac{1}{\sqrt{2\pi}\,\delta}e^{\frac{-1}{2}\left(\frac{x-\mu}{\delta}\right)^2}dx$$

$$M_\mu(t)=\frac{1}{\sqrt{2\pi}\,\delta}\int_{-\infty}^{\infty}e^{t(x-\mu)-\frac{1}{2}\left(\frac{x-\mu}{\delta}\right)^2}dx$$

$$M_\mu(t)=\frac{1}{\sqrt{2\pi}\,\delta}\int_{-\infty}^{\infty}e^{\frac{1}{2\delta^2}\left[2\delta^2t(x-\mu)-(x-\mu)^2\right]}dx$$

$$M_\mu(t)=\frac{1}{\sqrt{2\pi}\,\delta}\int_{-\infty}^{\infty}e^{\frac{-1}{2\delta^2}\left[(x-\mu)^2-2\delta^2t(x-\mu)\right]}dx$$

$$M_\mu(t)=\frac{1}{\sqrt{2\pi}\,\delta}\int_{-\infty}^{\infty}e^{\frac{-1}{2\delta^2}\left[(x-\mu)^2-2\delta^2t(x-\mu)+\left(\delta^2t\right)^2-\left(\delta^2t\right)^2\right]}dx$$

$$M_\mu(t)=\frac{1}{\sqrt{2\pi}\,\delta}\int_{-\infty}^{\infty}e^{\frac{-1}{2\delta^2}\left[\left((x-\mu)^2-\delta^2t\right)^2-\left(\delta^2t\right)^2\right]}dx$$

$$M_{\mu}(t) = \frac{1}{\sqrt{2\pi}\,\delta} \int_{-\infty}^{\infty} e^{\frac{-1}{2\delta^2}\left[\left((x-\mu)^2 - \delta^2 t\right)^2\right]} . e^{\frac{-1}{2\delta^2}\left(-\delta^4 t^2\right)} \, dx$$

$$M_{\mu}(t) = \frac{1}{\sqrt{2\pi}\,\delta} . e^{\frac{\delta^2 t^2}{2}} \int_{-\infty}^{\infty} e^{\frac{-1}{2\delta^2}\left(\frac{(x-\mu)-\delta^2 t}{\delta}\right)^2} \, dx$$

$$M_{\mu}(t) = \frac{(x-\mu)-\delta^2}{\delta} = z \quad \Rightarrow (x-\mu) - \delta^2 = \delta z$$

$$\Rightarrow dx = \delta dz$$

$$\Rightarrow z \to \pm\infty \;\; as \;\; x \to \pm\infty$$

$$M_{\mu}(t) = \frac{1}{\sqrt{2\pi}\,\delta} . e^{\frac{\delta^2 t^2}{2}} \int_{-\infty}^{\infty} e^{\frac{-1}{2} z^2} \, \delta dz$$

$$M_{\mu}(t) = \frac{1}{\sqrt{2\pi}} . e^{\frac{\delta^2 t^2}{2}} \int_{-\infty}^{\infty} e^{\frac{-1}{2} z^2} \, dz$$

$$M_{\mu}(t) = e^{\frac{\delta^2 t^2}{2}} .1$$

$$M_{\mu}(t) = e^{\frac{\delta^2 t^2}{2}}$$

Properties of Normal distribution:

The Normal distribution has the following properties

(i) The curve is symmetric about the vertical axis through the mean μ.
(ii) The mode is a point on horizontal axis where the curve is maximum at x = μ.
(iii) The total area under the normal curve and above the horizontal axis is always one.
(iv) The normal curve approaches the horizontal axis when we proceed in either side of mean μ.

Lecture # 12

The Chi-Square χ^2 Distribution:

- **Degree of freedom:**

The difference between the number of independent observations in a sample and number of populations to be estimated from a sample is called a degree of freedom.

- **Chi-Square random variable:**

Let $z_1, z_2, ..., z_n$ are independent normally distribution random variables with mean μ and variance δ^2. Then

$$\chi^2 = \sum_{i=1}^{n} z_i^2 = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\delta}\right)^2$$

is called a Chi-square χ^2 random variable with n−1 degree of freedom.

- **Chi-Square distribution:**

The density function or distribution of Chi-square is defined as

$$f\left(\chi^2\right) = \frac{1}{2^{\frac{n}{2}}\left|\frac{n}{2}\right.}\left(\chi^2\right)^{\frac{n}{2}-1} e^{\frac{-\chi^2}{2}} \quad where\ 0 < \chi^2 < \infty$$

Theorem: Show that the density function of Chi-square is

$$f\left(\chi^2\right) = \frac{1}{2^{\frac{n}{2}}\left|\frac{n}{2}\right.}\left(\chi^2\right)^{\frac{n}{2}-1} e^{\frac{-\chi^2}{2}}$$

Proof: As we know that the Chi-square random variable is

$$\chi^2 = \sum_{i=1}^{n} z_i^2$$

And the moment generating function of Chi-Square is written as

$$M_0\left(t\right) = E\left(e^{t\chi^2}\right)$$

$$M_0(t)=E\left(e^{t\sum_{i=1}^{n}z_i^2}\right)$$

$$M_0(t)=E\left(e^{t\left(z_1^2+z_2^2+\ldots+z_n^2\right)}\right)$$

$$M_0(t)=E\left(e^{tz_1^2}.e^{tz_2^2}\ldots.e^{tz_n^2}\right)$$

$$M_0(t)=\prod_{i=1}^{n}E\left(e^{t\,z_i^2}\right) \qquad \_\_\_\_(A)$$

As we know that

$$E\left(e^{t\,z_i^2}\right)=\int_{-\infty}^{\infty}e^{t\,z_i^2}\;f(z)dz \qquad \_\_\_\_\_(B) \quad \because E(X)=\int_{-\infty}^{\infty}xf(x)dx$$

As we know that z_i are independent normally distributed random variable. So, if $\mu = 0$ and $\delta^2 = 1$ the normal distribution can be written as

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{\frac{-1}{2}z^2} \qquad \because f(x)=\frac{1}{\sqrt{2\pi}}e^{\frac{-1}{2}\left(\frac{x-\mu}{\delta}\right)^2}$$

Using this value in (B) we get

$$E\left(e^{t\,z_i^2}\right)=\int_{-\infty}^{\infty}e^{t\,z_i^2}\;\frac{1}{\sqrt{2\pi}}e^{\frac{-1}{2}z^2}\;dz$$

$$E\left(e^{t\,z_i^2}\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{t\,z^2-\frac{z^2}{2}}\;dz \quad \because z_i=z\ in\ general$$

$$E\left(e^{t\,z_i^2}\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{z^2}{2}(1-2t)}\;dz$$

$$E\left(e^{t z_i^2}\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\left(\sqrt{\frac{1-2t}{2}}.z\right)^2} \; dz \qquad \_\_\_\_\_(C)$$

$$Let \;\; \sqrt{\frac{1-2t}{2}}\,.z=v$$

$$\sqrt{\frac{1-2t}{2}}\;dz=dv$$

$$dz=\sqrt{\frac{2}{1-2t}}\,dv$$

$$v\to\pm\infty \;as\; z\to\pm\infty$$

Using these substitutions in (C) we get

$$E\left(e^{t z_i^2}\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-v^2}\sqrt{\frac{2}{1-2t}}\,dv$$

$$E\left(e^{t z_i^2}\right)=\frac{1}{\sqrt{\pi}\sqrt{1-2t}}\int_{-\infty}^{\infty} e^{-v^2}\;dv \qquad \_\_\_\_(D)$$

As the integral function is even. So,

$$\int_{-\infty}^{\infty} e^{-v^2}\;dv=2\int_{0}^{\infty} e^{-v^2}\;dv$$

Equation (D) $\Rightarrow$

$$E\left(e^{t z_i^2}\right)=\frac{1}{\sqrt{\pi}\sqrt{1-2t}}2\int_{0}^{\infty} e^{-v^2}\;dv \qquad \_\_\_\_\_(E)$$

$$Let \;\; v^2=x \;\Rightarrow v=\sqrt{x}$$

$$\Rightarrow\; dv=\frac{dx}{2\sqrt{x}}$$

$$x\to 0 \;\; as\; v\to 0 \quad and \;\; x\to\infty \;\; as\; v\to\infty$$

Equation (E) $\Rightarrow$ $$E\left(e^{t z_i^2}\right)=\frac{1}{\sqrt{\pi}\sqrt{1-2t}}2\int_0^{\infty} e^{-x}\frac{dx}{2\sqrt{x}}$$

$$E\left(e^{t z_i^2}\right)=\frac{1}{\sqrt{\pi}\sqrt{1-2t}}\int_0^{\infty} e^{-x}.\,x^{\frac{-1}{2}}\,dx$$

$$E\left(e^{t z_i^2}\right)=\frac{1}{\sqrt{\pi}\sqrt{1-2t}}.\Gamma\left(\frac{1}{2}\right) \qquad \because a^{-n}\Gamma(n)=\int_0^{\infty} e^{-ax}.\,x^{n-1}\,dx$$

$$E\left(e^{t z_i^2}\right)=\frac{1}{\sqrt{\pi}\sqrt{1-2t}}.\sqrt{\pi} \qquad \because \Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$$

$$E\left(e^{t z_i^2}\right)=\frac{1}{\sqrt{1-2t}} \qquad \_\_\_\_\_\_(F)$$

Using (F) in (A) , we get

$$M_0(t)=\prod_{i=1}^{n}\frac{1}{\sqrt{1-2t}}=\frac{1}{(1-2t)^{\frac{n}{2}}}$$

$$M_0(t)=\frac{1}{\left(1-\frac{t}{\frac{1}{2}}\right)^{\frac{n}{2}}} \qquad \_\_\_\_\_\_(G)$$

As we know that the distribution of Gamma function is

$$\phi_{\chi}=f\left(\chi^2\right)=\frac{a^p e^{-a\chi^2}\left(\chi^2\right)^{p-1}}{\Gamma p} \qquad \_\_\_\_(i)$$

And moment generating function of gamma function is

$$M_0(t)=\frac{1}{\left(1-\frac{t}{a}\right)^{p}} \qquad \_\_\_\_\_(ii)$$

Comparing (G) in (ii)

$\Rightarrow a = ½$, $p = n/2$

Now the distribution of χ^2 can be written as i.e. using the value of 'a' and 'p' in equation (i)

$$f\left(\chi^2\right)=\frac{\left(\frac{1}{2}\right)^{\frac{n}{2}} e^{\frac{-1}{2}\chi^2}\left(\chi^2\right)^{\frac{n}{2}-1}}{\Gamma\frac{n}{2}}$$

$$f\left(\chi^2\right)=\frac{1}{2^{\frac{n}{2}}\Gamma\frac{n}{2}}.e^{\frac{-\chi^2}{2}}\left(\chi^2\right)^{\frac{n}{2}-1}$$

Properties of Chi-square distribution:

(i) The mean of Chi-square distribution 'n' (degree of freedom).
(ii) The variance of Chi-square distribution is '2n'.
(iii) The mode of Chi-square distribution is 'n−2'.
(iv) The total area of χ^2 is always 1.

Question: If X and Y are any two Chi-square variances with n_1 and n_2 are the degree of freedom. Then show that X+Y is also a Chi-square random variable with degree of freedom 'n_1+n_2'.

Solution: As we know that the moment generating function is Chi-square is

$$M_X(t)=E\left(e^{tX}\right)=(1-2t)^{\frac{-n_1}{2}}$$

$$M_Y(t)=E\left(e^{tY}\right)=(1-2t)^{\frac{-n_2}{2}}$$

$$Now \quad M_{X+Y}(t)=E\left(e^{t(X+Y)}\right)=E\left(e^{tX+tY}\right)$$

$$Now \quad M_{X+Y}(t)=E\left(e^{tX}.e^{tY}\right)=E\left(e^{tX}\right).E\left(e^{tY}\right)$$

$$M_{X+Y}(t) = (1-2t)^{\frac{-n_1}{2}}.(1-2t)^{\frac{-n_2}{2}}$$

$$M_{X+Y}(t) = (1-2t)^{\frac{-n_1}{2}-\frac{n_2}{2}}$$

$$M_{X+Y}(t) = (1-2t)^{-\left(\frac{n_1+n_2}{2}\right)}$$

Lecture # 13

Question: Show that the mean and variance of χ^2 distribution is 'n' and '2n' respectively.

Solution: As we know that the density function of χ^2 is

$$f\left(\chi^2\right)=\frac{1}{2^{\frac{n}{2}}\Gamma\frac{n}{2}}.e^{\frac{-\chi^2}{2}}\left(\chi^2\right)^{\frac{n}{2}-1}$$

$$Put\ \chi^2=x \quad and \quad A=\frac{1}{2^{\frac{n}{2}}\Gamma\frac{n}{2}}$$

$$Then\ f(x)=Ae^{\frac{-x}{2}}(x)^{\frac{n}{2}-1}$$

$$Mean=E(X)=\int_{-\infty}^{\infty}x\,f(x)dx$$

$$Mean=E(X)=\int_{0}^{\infty}x\,f(x)dx \qquad for\ \chi^2$$

$$E(X)=\int_{0}^{\infty}x\,Ae^{\frac{-x}{2}}(x)^{\frac{n}{2}-1}dx$$

$$E(X)=A\int_{0}^{\infty}e^{\frac{-x}{2}}(x)^{\frac{n}{2}-1+1}dx$$

$$E(X)=A\int_{0}^{\infty}e^{\frac{-x}{2}}(x)^{\left(\frac{n}{2}+1\right)-1}dx$$

$$E(X)=A\left[\left(\frac{1}{2}\right)^{-\left(\frac{n}{2}+1\right)}.\Gamma\left(\frac{n}{2}+1\right)\right] \qquad \because \int_{0}^{\infty}e^{-ax}x^{n-1}dx=a^{-n}\Gamma n$$

$$E\left(X\right)=A\left(2\right)^{\frac{n}{2}+1}.\frac{n}{2}\Gamma\left(\frac{n}{2}\right) \qquad \because \Gamma\left(n+1\right)=n\Gamma\left(n\right)$$

$$E\left(X\right)=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\left(2\right)^{\frac{n}{2}}.2.\frac{n}{2}\Gamma\left(\frac{n}{2}\right)$$

$$E\left(X\right)=n$$

$$Now \quad Variance\left(X\right)=E\left(X^2\right)-\left(E\left(X\right)\right)^2 \qquad \_\_\_\_\_\left(i\right)$$

$$E\left(X^2\right)=\int_{-\infty}^{\infty}x^2 f\left(x\right)dx$$

$$E\left(X^2\right)=\int_{0}^{\infty}x^2 f\left(x\right)dx \qquad \because for\ \chi^2$$

$$E\left(X^2\right)=\int_{0}^{\infty}x^2 Ae^{\frac{-x}{2}}\left(x\right)^{\frac{n}{2}-1}dx$$

$$E\left(X^2\right)=A\int_{0}^{\infty}e^{\frac{-x}{2}}\left(x\right)^{\frac{n}{2}-1+2}dx$$

$$E\left(X^2\right)=A\int_{0}^{\infty}e^{\frac{-x}{2}}\left(x\right)^{\left(\frac{n}{2}+2\right)-1}dx$$

$$E\left(X^2\right)=A\left[\left(\frac{1}{2}\right)^{-\left(\frac{n}{2}+2\right)}.\Gamma\left(\frac{n}{2}+2\right)\right] \qquad \because \int_{0}^{\infty}e^{-ax}x^{n-1}dx=a^{-n}\Gamma\left(n\right)$$

$$E\left(X^2\right)=A\left(2\right)^{\frac{n}{2}+2}\Gamma\left(\frac{n}{2}+2\right)$$

$$E\left(X^2\right)=A\left(2\right)^{\frac{n}{2}+2}\Gamma\left(\frac{n}{2}+1+1\right)$$

$$E\left(X^2\right)=A(2)^{\frac{n}{2}}.(2)^2.\left(\frac{n}{2}+1\right)\Gamma\left(\frac{n}{2}+1\right)$$

$$E\left(X^2\right)=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}(2)^{\frac{n}{2}}.4.\left(\frac{n}{2}+1\right)\frac{n}{2}\Gamma\left(\frac{n}{2}\right) \qquad \because A=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}$$

$$E\left(X^2\right)=n(n+2)=n^2+2n$$

Put the value of $E\left(X^2\right)$ *and* $E(X)$ *in*(i)

$$Var(X)=n^2+2n-(n)^2$$

$$Var(X)=n^2+2n-n^2$$

$$Var(X)=2n$$

T-Distribution:

Let 'z' be a standard normal random variable and 'v' be a χ^2 random variable, then if 'n' is the degree of freedom and 'z' and 'v' are independent random variables we can define 'T' random variable as

$$T=\frac{z}{\sqrt{\frac{v}{n}}}$$

And its distribution (density function) can be written as

$$h(t)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}\left[1+\frac{t^2}{n}\right]^{-\left(\frac{n+1}{2}\right)}$$

$$h(t)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n}\sqrt{\pi}}\left[1+\frac{t^2}{n}\right]^{-\left(\frac{n+1}{2}\right)}$$

$$h(t)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right).\Gamma\left(\frac{1}{2}\right)\sqrt{n}}\left[1+\frac{t^2}{n}\right]^{-\left(\frac{n+1}{2}\right)} \qquad \because \Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$$

As Beta function is $\beta(m,n)=\frac{\Gamma m.\Gamma n}{\Gamma(m+n)}$

$$h(t)=\frac{1}{\beta\left(\frac{n}{2},\frac{1}{2}\right).\sqrt{n}}\left[1+\frac{t^2}{n}\right]^{-\left(\frac{n+1}{2}\right)} \; ; -\infty<t<\infty$$

Theorem: Show that the Area under the normal T-distribution is 1 (unity).

Proof: As we know that

$$h(t)=\frac{1}{\beta\left(\frac{n}{2},\frac{1}{2}\right).\sqrt{n}}\left[1+\frac{t^2}{n}\right]^{-\left(\frac{n+1}{2}\right)}$$

Here we have to prove that $\int_{-\infty}^{\infty} h(t)\,dt=1$

$$\int_{-\infty}^{\infty} h(t)\,dt=\int_{-\infty}^{\infty}\frac{1}{\beta\left(\frac{n}{2},\frac{1}{2}\right).\sqrt{n}}\left[1+\frac{t^2}{n}\right]^{-\left(\frac{n+1}{2}\right)}dt$$

$$Let \quad A=\frac{1}{\beta\left(\frac{n}{2},\frac{1}{2}\right).\sqrt{n}}$$

$$\int_{-\infty}^{\infty} h(t)dt=\int_{-\infty}^{\infty} A\left[1+\frac{t^2}{n}\right]^{-\left(\frac{n+1}{2}\right)} dt$$

$$\int_{-\infty}^{\infty} h(t)dt=A\int_{-\infty}^{\infty}\left[1+\left(\frac{t}{\sqrt{n}}\right)^2\right]^{-\left(\frac{n+1}{2}\right)} dt$$

$$Let \;\; x=\frac{t}{\sqrt{n}} \quad \Rightarrow \quad dt=\sqrt{n}\,dx$$

$$x\to\pm\infty \;\; as \;\; t\to\pm\infty$$

$$\int_{-\infty}^{\infty} h(t)dt=A\int_{-\infty}^{\infty}\left[1+x^2\right]^{-\left(\frac{n+1}{2}\right)}\sqrt{n}dx$$

$$\int_{-\infty}^{\infty} h(t)dt=A\sqrt{n}\int_{-\infty}^{\infty}\frac{1}{\left(1+x^2\right)^{\left(\frac{n+1}{2}\right)}}dx$$

As the given function is even function. So,

$$\int_{-\infty}^{\infty} h(t)dt=2A\sqrt{n}\int_{0}^{\infty}\frac{1}{\left(1+x^2\right)^{\left(\frac{n+1}{2}\right)}}dx$$

$$Let \; v=\frac{1}{1+x^2}$$

$$1+x^2=\frac{1}{v} \quad \Rightarrow \; x^2=\frac{1}{v}-1=\frac{1-v}{v} \;\Rightarrow x=\sqrt{\frac{1-v}{v}}$$

$$\frac{-1}{v^2}dv = 2xdx$$

$$\frac{-1}{v^2}dv = 2\sqrt{\frac{1-v}{v}}dx$$

$$\frac{-1}{v^2}\left(\frac{1}{2}\sqrt{\frac{v}{1-v}}\right)dv = dx$$

$$v \to 1 \;\; as \;\; x \to 0$$

$$v \to \infty \;\; as \;\; x \to \infty$$

$$\int_{-\infty}^{\infty} h(t)\,dt = 2A\sqrt{n}\int_{1}^{0}(v)^{\left(\frac{n+1}{2}\right)}\left(\frac{-1}{2v^2}\sqrt{\frac{v}{1-v}}\right)dv$$

$$\int_{-\infty}^{\infty} h(t)\,dt = A\sqrt{n}\int_{0}^{1}(v)^{\left(\frac{n+1}{2}\right)}\left(\frac{1}{v^2}.\frac{v^{\frac{1}{2}}}{\sqrt{1-v}}\right)dv$$

$$\int_{-\infty}^{\infty} h(t)\,dt = A\sqrt{n}\int_{0}^{1}(v)^{\frac{n+1}{2}-2+\frac{1}{2}}.\frac{1}{\sqrt{1-v}}dv$$

$$\int_{-\infty}^{\infty} h(t)\,dt = A\sqrt{n}\int_{0}^{1}(v)^{\frac{n-2}{2}}.(1-v)^{\frac{-1}{2}}dv$$

$$\int_{-\infty}^{\infty} h(t)\,dt = A\sqrt{n}\int_{0}^{1}(v)^{\frac{n}{2}-1}.(1-v)^{\frac{1}{2}-1}dv$$

$$\int_{-\infty}^{\infty} h(t)\,dt = A\sqrt{n}\,\beta\left(\frac{n}{2},\frac{1}{2}\right) \qquad \because \beta(m,n)=\int_{0}^{1}x^{m-1}.(1-x)^{n-1}dx$$

$$\int_{-\infty}^{\infty} h(t)\,dt = \frac{1}{\beta\left(\frac{n}{2},\frac{1}{2}\right).\sqrt{n}}\sqrt{n}\,\beta\left(\frac{n}{2},\frac{1}{2}\right) \quad \because A=\frac{1}{\beta\left(\frac{n}{2},\frac{1}{2}\right).\sqrt{n}}$$

$$\int_{-\infty}^{\infty} h(t)\,dt = 1 \qquad \text{Proved}$$

Properties of T-distribution:

(i) The T-distribution like a normal distribution has a bell-shaped; uni model and symmetric around the mean (μ=0).

(ii) In T-distribution, the number of degrees of freedom is the function of size. The shape of T-distribution curve is changed, when we changed the number of degrees of freedom. It means that we can obtain a family of T-distribution curve according to the degree of freedom.

(iii) For a very small number of degrees of freedom, the curve of T-distribution become flatter and flatter. It means that the T-distribution approaches the normal distribution as the sample size increases without limits.

(iv) The T-distribution has variance more than one but the normal distribution has variance one.

(v) The mean and variance of T-distribution is zero and n/n−2 respectively where n > 2.

F-Distribution:

Let u and v be any two independent random variables having χ^2 distribution with n_1 and n_2 are degree of freedom, then the 'F' random variable can be defined as

$$F = \frac{u/n_1}{v/n_2}$$

And its distribution or density function defined as

$$h(f) = \frac{\Gamma\frac{n_1+n_2}{2}.\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}}(f)^{\frac{n_1}{2}-1}}{\Gamma\frac{n_1}{2}.\Gamma\frac{n_2}{2}\left(1+\frac{n_1}{n_2}f\right)^{\frac{n_1+n_2}{2}}} \qquad ;0 < f < \infty$$

$$h(f) = 0 \qquad otherwise$$

$$h(f) = \frac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} (f)^{\frac{n_1}{2}-1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right) \cdot \left(1 + \frac{n_1}{n_2} f\right)^{\frac{n_1+n_2}{2}}} \qquad ; 0 < f < \infty$$

Properties of F-distribution:

(i) The random variable of F-distribution takes a non-negative value.

(ii) The range of F-distribution is 0 to ∞.

(iii) The shape of F-distribution curve is non-symmetrical and skewed to the right, but when the degree of freedom n_1 and n_2 increases then the curve of F-distribution becomes symmetrical.

(iv) The F-distribution has a unique mode at the value $\frac{n_2(n_1 - 2)}{n_1(n_2 + 2)}$; $n_2 > -2$

and it is always less than unity.

(v) The mean and variance of F-distribution is $\frac{n_2}{(n_2 - 2)}$; $n_2 > 2$

and $\frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}$; $n_2 > 4$