Sample Question Bank

Numericals given here are just for reference. Numericals based on Modules 3 and 4 can be expected. These are just sample numerical questions given here for your reference.

Chapter 3

1) The weight of the person is related to his height. Find the relationship between height and weight using linear regression and also predict the weight of the person whose height is 170cm

Height	Weight
151	63
174	81
138	56
186	91
128	47
136	57
179	76
163	72
152	62
131	48

Normal Equations

- The system of equations required to be solved for obtaining the values of constants known as Normal equations

y=a+bx

$$\sum y = Na + b\sum x$$

$$\sum xy = a\sum x + b\sum x^2$$

Example

x	y
0	2
1	1
2	3
3	2
4	4
5	3
6	5
Sum 21	20

x^2	xy
0	0
1	1
4	6
9	6
16	16
25	15
36	30
91	74

$$\sum y = Na + b\sum x$$
$$\sum xy = a\sum x + b\sum x^2$$

20=7a+21b
74=21a+91b

a=1.357 b=0.5
y=1.357+0.5x

Normal Equations

- The system of equations required to be solved for obtaining the values of constants known as Normal equations

x=a+by

$$\sum x = Na + b\sum y$$
$$\sum xy = a\sum x + b\sum y^2$$

From the following data obtain two regression equations

x 6 2 10 4 8

y 9 11 5 8 7

	x	y	x^2	y^2	xy
	6	9	36	81	54
	2	11	4	121	22
	10	5	100	25	50
	4	8	16	64	32
	8	7	64	49	56
Sum	**30**	**40**	**220**	**340**	**214**
Mean	**6**	**8**	**44**	**68**	**42.8**

$$\sum y = Na + b\sum x$$

$$\sum xy = a\sum x + b\sum x^2$$

40=5a+30b
214=30a+220b

a=11.9 b=-0.65
Y=11.9-0.65x

$$\sum x = Na + b\sum y$$

$$\sum xy = a\sum x + b\sum y^2$$

30=5a+40b
214=30a+340b

a=16.4 b=-1.3
y=16.4-1.3x

2) Following information is obtained from the records of a business organization

Sales (in '000):	91	53	45	76	89	95	80	65
Advertisement Expense (₹ in '000)	15	8	7	12	17	25	20	13

a. Compute regression coefficients
b. Obtain the two regression equations and
c. Estimate the advertisement expenditure for a sale of Rs. 1,20,000

From the following data obtain regression equations taking deviations of items from the mean of x and y

x 6 2 10 4 8

y 9 11 5 8 7

	X	$X-\bar{X}$ x	x^2
	6	0	0
	2	-4	16
	10	4	16
	4	-2	4
	8	2	4
Sum	**30**	0	40
Mean	**6**	0	8

	Y	$Y-\bar{Y}$ y	y^2	xy
	9	1	1	0
	11	3	9	-12
	5	-3	9	-12
	8	0	0	0
	7	-1	1	-2
Sum	**40**	0	20	-26
Mean	**8**	0	4	-5.2

$$X-\bar{X}=r\frac{\sigma_x}{\sigma_y}(Y-\overline{Y})$$

$$r\frac{\sigma_x}{\sigma_y}=\frac{\sum xy}{\sum y^2}=\frac{-26}{20}=-1.3$$

$X-6$ =-1.3 (Y −8)

X=-1.3Y+16.4

$$Y-\bar{Y}=r\frac{\sigma_y}{\sigma_x}(X-\bar{X})$$

$r\frac{\sigma_y}{\sigma_x}=\frac{\sum xy}{\sum y^2}$=-5.2/40=-0.65

$Y-8$ =-0.65 (X−6)

Y=-0.65.3X+11.9

From the following data obtain regression equations taking deviations of X series from 5 and Y series from 7

x 6 2 10 4 8

y 9 11 5 8 7

	X	$X-5$ dx	dx^2
	6	1	1
	2	-3	9
	10	5	25
	4	-1	1
	8	3	9
Sum	**30**	5	45
Mean	**6**	1	9

	Y	$Y-7$ dy	dy^2	dxdy
	9	2	4	2
	11	4	16	-12
	5	-2	4	-10
	8	1	1	-1
	7	0	0	0
Sum	**40**	5	25	-21
Mean	**8**	1	5	-4.25

$$X-\bar{X}=r\frac{\sigma_x}{\sigma_y}(Y-\overline{Y})$$

$$r\frac{\sigma_x}{\sigma_y}=\frac{N\sum dxdy-(\sum dx)(\sum dy)}{N\sum dy^2-(\sum dy)^2}$$

$$=\frac{5(-21)-5*5}{5*25-25}=-1.3$$

$X-6$ =-1.3 (Y −8)

X=-1.3Y+16.4

$$Y-\bar{Y}=r\frac{\sigma_y}{\sigma_x}(X-\bar{X})$$

$$r\frac{\sigma_y}{\sigma_x}=\frac{N\sum dxdy-(\sum dx)(\sum dy)}{N\sum dx^2-(\sum dx)^2}$$

$$=\frac{5(-21)-5*5}{5*45-25}=-0.65$$

$Y-8$ =-0.65 (X−6)

Y=-0.65.3X+11.9

x	y	x^2	xy	y^2
6	199	36	1194	39601
7	213	49	1491	45369
8	245	64	1960	60025
9	272	81	2448	73984
$\sum x = 45$	$\sum y = 1488$	$\sum x^2 = 285$	$\sum xy = 8924$	$\sum y^2 = 282290$

Here, $\bar{x} = \frac{\sum x}{n} = 4.5$ and $\bar{y} = \frac{\sum y}{n} = 148.8$

The regression coefficients are

$$b_{xy} = \frac{n\sum xy - \sum x \sum y}{n\sum y^2 - (\sum y)^2} = 0.0366 \quad \text{and} \quad b_{yx} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = 27.0061$$

The regression line of y on x is

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$y - 148.8 = 27.0061\,(x - 4.5)$$

$$\therefore\ y = 27.0061x + 27.2726$$

The regression line of x on y is

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

$$x - 4.5 = 0.0366\,(y - 148.8)$$

$$\therefore\ x = 0.366y - 0.9461$$

Thus, at x = 15 hours, y = 432.3641

(3) Two independent ...

Ex. 3.1.1 : Calculate the coefficient of correlation for the following data.

x	9	8	7	6	5	4	3	2	1
y	15	16	14	13	11	12	10	8	9

☑ Soln. :

Here, n = 9

x	y	x^2	y^2	xy
9	15	81	225	135
8	16	64	256	128
7	14	49	196	98
6	13	36	169	78
5	11	25	121	55
4	12	16	144	48
3	10	9	100	30
2	8	4	64	16
1	9	1	81	9
$\sum x = 45$	$\sum y = 108$	$\sum x^2 = 285$	$\sum y^2 = 1356$	$\sum xy = 597$

he coefficient of correlation is

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}} = \frac{9(597) - (45)(108)}{\sqrt{9(285) - 45^2}\sqrt{9(1356) - 108^2}}$$

$$r = 0.95$$

3) Two random variables have the regression equations $3X + 2Y - 26 = 0$

$6X + Y - 31 = 0$

Find the mean values and the coefficient of correlation between X and Y. If the variance of X=25, find the standard deviation of Y from the data given above.

After 9/11 attack, a company could partially recover following record on correlation

Variance of X=9

Eqns. Of regression 8X-10Y+66=0

40X-18Y=214

Find on the basis of above information

1) Mean values of X and Y
2) Coefficient of Correlation
3) Standard Deviation of Y

Solving two equations we get mean of X and mean of Y

$\bar{X}$ =13 $\bar{Y}$ =17

$b_{xy} = r\frac{\sigma_x}{\sigma_y}$

8X-10Y=-66
X-(10/8)Y=-66/8
b_{xy}=10/8=1.25

40X-18Y=214
-Y=214/18-(40/18)X
b_{yx}=40/18=2.22

0.45=0.6(3/σ_y)
σ_y=(0.6*3)/0.45
=4

8X-10Y=-66
10Y=8X+66
Y=(8/10)X+66/10
b_{yx}=8/10=0.8

40X=18Y+214
Y=(18/40)X+214/18
b_{xy}=18/40=0.45

R=sqrt(b_{yx}*b_{xy})=sqrt(0.8*0.45)=0.6

Chapter 4

1) Find multiple linear regression equation of Y on X_1 and X_2

.

Y	4	6	7	9	13	15
X_1	15	12	8	6	4	3
X_2	30	24	20	14	10	4

Method 1- Normal Equations

y	x_1	x_2
4	15	30
6	12	24
7	8	20
9	6	14
13	4	10
15	3	4
54	**48**	**102**

x_1y	x_2y	x_1x_2	x_1^2	X_2^2
60	120	450	225	900
72	144	288	144	576
56	140	160	64	400
54	126	84	36	196
52	130	40	16	100
45	60	12	9	16
339	**720**	**1034**	**494**	**2188**

$$54 = 6a + 48b_1 + 102b_2$$
$$339 = 48a + 494b_1 + 1034b_2$$
$$720 = 102a + 1034b_1 + 2188b_2$$

$y=16.47+0.38x_1-0.62x_2$

Ex. 4.1.1 : Fit a regression equation to estimate β_0, β_1, β_2 to the following data of a transport company on the weights of 6 shipments, the distances they were moved and the damage of the goods that was incurred. Estimate the damage when a shipment of 3700 kg. is moved to a distance of 260 km.

Weight X_1 (1000 kg)	4.0	3.0	1.6	1.2	3.4	4.8
Distance X_2 (100 km)	1.5	2.2	1.0	2.0	0.8	1.6
Damage Y (Rs.)	160	112	69	90	123	186

Soln. :

Let weight X_1 and distance X_2 be independent variables and the damage y be the dependent variable.

Let the equation of regression be, $y = b_0 + b_1 x_1 + b_2 x_2$

Where b_0, b_1, b_2 are estimates of β_0, β_1, β_2. The three normal equations become.

$$\sum_{i=1}^{6} Y_i = nb_0 + b_1 \sum_{i=1}^{6} x_{1i} + b_2 \sum_{i=1}^{6} x_{2i}$$

$$\sum x_{1i}\, y_i = b_0 \sum x_{1i} + b_1 \sum x_{1i}^2 + b_2 \sum x_{1i} x_{2i}$$

$$\sum x_{2i}\, y_i = b_0 \sum x_{2i} + b_1 \sum x_{1i} x_{2i} + b_2 \sum x_{2i}^2$$

We prepare the table :

	x_1 (weight) (1000 kg)	x_2 distance 100 km	y damage Rs.	x_1^2	x_2^2	$x_1 x_2$	$x_1 y$	$x_2 y$
	4.0	1.5	160	16	2.25	6.0	640	240
	3.0	2.2	112	0	4.84	6.6	336	246.4
	1.6	1.0	69	2.56	1.0	1.6	110.4	69
	1.2	2.0	90	1.44	4.0	2.4	108	180
	3.4	0.8	123	11.56	0.64	2.72	418.2	98.4
	4.8	1.6	186	23.04	2.56	7.68	892.8	297.6
Total	**18**	**9.1**	**740**	**63.6**	**15.29**	**27**	**250.54**	**1131.4**

The data is : $n = 6, \sum x_{1i} = 18, \sum x_{2i} = 9.1,$

$\sum y_i = 740, \quad \sum x_{1i}^2 = 63.6, \quad \sum x_{2i}^2 = 15.29, \quad \sum x_{1i}\, x_{2i} = 27$

$\sum x_i\, y_i = 250.54, \sum x_{2i}\, y_i = 1131.4$

$\therefore$ Normal equations become

$$740 = 6b_0 + 18\, b_1 + 9.1\, b_2$$

$$250.54 = 18b_0 + 63.6\, b_1 + 27\, b_2$$

$$1131.4 = 9.1\, b_0 + 27\, b_1 + 15.29\, b_2$$

Solving, we get $b_0 = 14.56, b_1 = 30.109,$

$$b_2 = 12.16$$

Thus the required regression equation is

$$y = 14.56 + 30.109\, x_1 + 12.16\, x_2$$

Estimate

For a weight of 3700 kg. ($x_1 = 3.7$) and for a distance of 260 km. ($x_2 = 2.6$) the damage incurred in rupees is

$$y\,(x_1 = 3.7, x_2 = 2.6) = 14.56 + 30.109(3.7) + 16\,(2.6)$$

$$= 714.58 = 715 \text{ Rs.}$$

2) Given the following, determine the regression equation of X2 on X1 and X3r12 = 0.8, r13 = 0.6, r23 = 0.5, σ1 = 10, σ = 8, σ3 = 5.

4.1.12 Solved Examples on Regression Equations

Ex. 4.1.2 : Let X_1, X_2 and X_3 be the excess of heights of father, mother and son respectively in 100 samples above their respective mean values in cm. A distribution of these variables gave the following correlation coefficients r_{ij} between X_i and X_j and standard deviations σ_i for i, j = 1, 2, 3.

$$r_{12} = 0.3, \quad r_{23} = 0.4, \quad r_{31} = 0.5,$$
$$\sigma_1 = 3, \quad \sigma_2 = 2, \quad \sigma_3 = 4$$

Obtain a regression equation of X_1 on X_2 and X_3, and estimate the excess of height of father when excess of heights of mother and son are 0.7 cm and 2.1 cm respectively.

Soln. :

Step I : Given Data is :

$$r_{12} = 0.3, \quad r_{23} = 0.4, \quad r_{31} = 0.5,$$
$$\sigma_1 = 3, \quad \sigma_2 = 2, \quad \sigma_3 = 4 \qquad \text{...(i)}$$

Since X_1, X_2, X_3 denote the excess of heights of father, mother and son respectively above their respective mean values, they are measured from their means.

Hence, the equation of the plane of regression of X_1 on X_2 and X_3 is given by :

$$X_1 = b_{12.3} X_2 + b_{13.2} X_3 \qquad \text{...(ii)}$$

Step II : We have

$$b_{12.3} = \frac{\sigma_1}{\sigma_2}\left(\frac{r_{12} - r_{13}\ r_{23}}{1 - r_{23}^2}\right) \quad \text{and} \quad b_{13.2} = \frac{\sigma_1}{\sigma_3}\left(\frac{r_{13} - r_{12} \cdot r_{32}}{1 - r_{32}^2}\right)$$

Substituting the given values from Equation (i) and noting that $r_{ij} = r_{ji}$; we get

$$b_{12.3} = \frac{3}{2}\left[\frac{0.3 - 0.5 \times 0.4}{1 - (0.4)^2}\right] = \frac{3 \times 0.10}{2 \times 0.84} = \frac{10}{56} = 0.1786$$

and

$$b_{13.2} = \frac{3}{4}\left[\frac{0.5 - 0.3 \times 0.4}{1 - (0.4)^2}\right] = \frac{3 \times 0.38}{4 \times 0.84} = \frac{19}{56} = 0.3393$$

Step III : Substituting in Equation (ii), the equation of regression of X_1 on X_2 and X_3 becomes

$$X_1 = 0.1786\, X_2 + 0.3393\, X_3 \qquad \text{...(iii)}$$

The estimate of the height of father when the excess of heights of mother and son are 0.7 cm and 2.1 cm respectively, is given by

$$\hat{X}_1 = (0.1786 \times 0.7) + (0.3393 \times 2.1)$$

$$= 0.12502 + 0.71253 = 0.83755 \text{ cm}$$

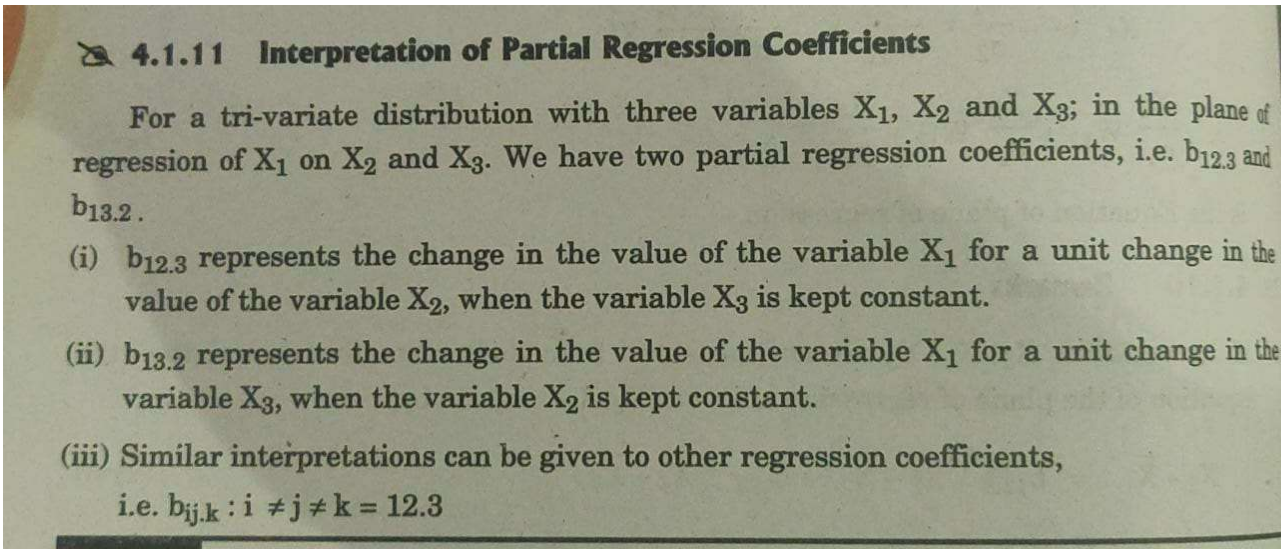
$$\therefore \hat{X}_1 = \mathbf{0.83755\ cm.}$$

3 : From heights (X_1) in inches, weights (X_2) in kg., ;

3) Define following terms and give their range.

(a) Partial regression coefficient

A partial regression coefficient (also known as a partial regression slope) is the slope coefficient of a linear regression model that measures the relationship between a predictor variable and a response variable, holding all other predictor variables constant. It indicates how much the response variable is expected to change for a unit increase in the predictor variable, when all other predictors in the model are held constant. The range of a partial regression coefficient is from negative infinity to positive infinity.



4.1.11 Interpretation of Partial Regression Coefficients

For a tri-variate distribution with three variables X_1, X_2 and X_3; in the plane of regression of X_1 on X_2 and X_3. We have two partial regression coefficients, i.e. $b_{12.3}$ and $b_{13.2}$.

(i) $b_{12.3}$ represents the change in the value of the variable X_1 for a unit change in the value of the variable X_2, when the variable X_3 is kept constant.

(ii) $b_{13.2}$ represents the change in the value of the variable X_1 for a unit change in the variable X_3, when the variable X_2 is kept constant.

(iii) Similar interpretations can be given to other regression coefficients, i.e. $b_{ij.k} : i \neq j \neq k = 12.3$

(b) Partial correlation coefficient

A partial correlation coefficient measures the strength and direction of the linear relationship between two variables, while controlling for the effects of one or more additional variables. It quantifies the degree to which two variables are associated after removing the influence of the other variables in the model. The range of a partial correlation coefficient is from -1 to +1.

4.2 COEFFICIENT OF PARTIAL CORRELATION

In probability theory and statistics, partial correlation measures the degree of association between two random variables after eliminating the linear effect of all the remaining variables on each of these two variables.

4.2.1 Functional Relation between Partial Correlation

For example, in case of three variables (X_1, X_2, X_3), the partial correlation coefficient between X_1 and X_2 is a measure of linear relationship between X_1 and X_2 after eliminating the linear effect of X_3 on both X_1 and X_2. It is denoted by $r_{12.3}$

Let $X_i^{'s}$ $(i = 1, 2, 3)$ be measured from their respective means so that

$$E(X_i) = 0, i = 1, 2, 3 \qquad \text{...(i)}$$

and let X_1, X_2, X_3 take N observations.

The equations of the line of regression of $(X_1$ on $X_3)$ and $(X_2$ on $X_3)$ are respectively

$$X_1 = b_{13} X_3 \quad \text{...(ii)} \quad \text{and} \quad X_2 = b_{23} X_3 \quad \text{...(iii)}$$

For given value of X_3, errors of estimates of X_1 and X_2 as given by these regression lines are :

$$X_{1.3} = X_1 - b_{13} X_3 \quad \text{and} \quad X_{2.3} = X_2 - b_{23} X_3$$

Using (i), we have $E(X_{1.3}) = 0$ and $E(X_{2.3}) = 0$ ...(iv)

Thus $X_{1.3}$ and $X_{2.3}$ may be regarded as that parts of the respective variables X_1 and X_2 which are left after the linear effects of X_3 on X_1 and X_2 have been eliminated.

Hence, the partial correlation coefficient between X_1 and X_2, denoted by $r_{12.3}$ is Karl Pearson's correlation coefficient between $X_{1.3}$ and $X_{2.3}$, and is given by

$$r_{12.3} = \frac{\text{Cov}(X_{1.3}, X_{2.3})}{\sqrt{\text{Var}(X_{1.3}) \cdot \text{Var}(X_{2.3})}}) = \frac{\text{Cov}(X_{1.3}, X_{2.3})}{\sigma_{1.3} \cdot \sigma_{2.3}} \qquad \text{...(v)}$$

Now, $\text{Cov}(X_{1.3}, X_{2.3}) = E[X_{1.3} \cdot X_{2.3}] - E[X_{1.3}] \cdot E[X_{2.3}]$

$$= E[X_{1.3} \cdot X_{2.3}] \qquad [\because E(X_{1.3}) = E(X_{2.3}) = 0 \text{ from (iv)}]$$

$$= \frac{1}{N} \Sigma X_{1.3} X_{2.3}$$

$$= \frac{1}{N} \Sigma [(X_1 - b_{13} X_3)(X_2 - b_{23} X_3)]$$

$$= \frac{1}{N} \Sigma X_1 X_2 - b_{23} \frac{1}{N} \Sigma X_1 X_3 - b_{23} \frac{1}{N} \Sigma X_2 X_3 + b_{13} b_{23} \frac{1}{N} \Sigma X_3^2$$

$$= r_{12}\,\sigma_1\,\sigma_2 - r_{23}\frac{\sigma_2}{\sigma_3}\cdot r_{13}\,\sigma_1\,\sigma_3 - r_{13}\cdot\frac{\sigma_1}{\sigma_3}\,r_{23}\,\sigma_2\,\sigma_3 + r_{13}\cdot\frac{\sigma_1}{\sigma_3}\cdot r_{23}\frac{\sigma_2}{\sigma_3}\cdot\sigma_3^2$$

$$= r_{12}\,\sigma_1\,\sigma_2 - r_{13}\,r_{23}\,\sigma_1\,\sigma_2 - r_{13}\,r_{23}\,\sigma_1\,\sigma_2 + r_{13}\,r_{23}\,\sigma_1\,\sigma_2$$

$$= \sigma_1\,\sigma_2\,(r_{12} - r_{13}\,r_{23}) \qquad \text{...(vi)}$$

Remark : For linear regression of y on x, the standard error of estimate of y on x is given by

$$S_{yx} = \sigma_y\,(1 - r_{xy}^2)^{1/2} = \sigma_y \cdot \sqrt{1 - r_{xy}^2}$$

Using the above remark, we have

$$\sigma_{1.3}^2 = \sigma_1^2\,(1 - r_{13}^2) \qquad \text{...(vii)}$$

$$\text{and } \sigma_{2.3}^2 = \sigma_2^2\,(1 - r_{23}^2)$$

Substituting (vi) and (vii) in (v) ; we get

$$r_{12.3} = \frac{\sigma_1\,\sigma_2\,(r_{12} - r_{13}\,r_{23})}{\sigma_1\sqrt{1 - r_{13}^2}\cdot\sigma_2\sqrt{1 - r_{23}^2}}$$

$$\therefore\ r_{12.3} = \frac{r_{12} - r_{13}\,r_{23}}{\sqrt{1 - r_{13}^2}\cdot\sqrt{1 - r_{23}^2}} \qquad \text{...(viii)}$$

Remarks

(1) Using symmetry of the result (viii), we can write : the partial correlation coefficient between X_i and X_j, after eliminating the linear effect of X_k on each of X_i and X_j is given by :

$$r_{ij\cdot k} = \frac{r_{ij} - r_{ik}\cdot r_{jk}}{\sqrt{1 - r_{ik}^2}\cdot\sqrt{1 - r_{jk}^2}} \qquad (i \neq j \neq k = 1, 2, 3)$$

Also, $\because r_{ij} = r_{ji}$ for all i, j = 1, 2, 3;

We have

$$r_{ij\cdot k} = r_{ij\cdot k}, \quad \text{for all } i \neq j \neq k = 1, 2, 3$$

(2) Note that $r_{ij.k}$ gives the degree of the linear relationship between two variables X_i and X_j which is free from the influence of the variable X_k, $i \neq j \neq k = 1, 2, 3$

(c) Multiple correlation coefficient

The multiple correlation coefficient (also known as the coefficient of multiple determination) is a measure of the strength and direction of the linear relationship between a response variable and multiple predictor variables. It indicates the proportion of the variation in the response variable that can be explained by the predictor variables in the model. The range of a multiple correlation coefficient is from 0 to 1.

4.1.15 Coefficient of Multiple Correlation

The **coefficient of multiple correlation** is a measure of how well a given variable can be predicted using a **linear function** of a set of other variables.

It is the **correlation** between the **variable's** values and best predictions that can be found out **linearly from** the predictive variables.

- The coefficient of multiple correlation takes values between 0 and 1. Higher values indicate higher predictability of the **dependent variable** from the **independent variables**, with a value of 1 indicating that the predictions are exactly correct and a value of O indicating that no linear combination of the independent variables is a better predictor than is the fixed **mean** of the dependent variable.
- The coefficient of multiple correlation is known as the square root of the **coefficient of determination**. But it is under the particular assumption that an **intercept** is included and that the best possible linear predictors are used.
- On the contrary, coefficient of determination is defined for more general cases, including those of non-linear prediction and those in which the predicted values have not been derived from a model-fitting procedure. The coefficient of multiple correlation, denoted R, is a **scalar** that is defined as the **Pearson Correlation Coefficient** between the predicted and the actual values of the dependent variable in a linear regression model that includes an **intercept**.

4.1.16 Definition of Coefficient of Multiple Correlation

In a trivariate distribution with three variables (X_1, X_2, X_3), **Karl Pearson's Correlation Coefficient** between X_1 and the joint effect of X_2 and X_3 on X_1 is called the multiple correlation coefficient of X_1 on X_2 and X_3 and is denoted by $R_{1.23}$. Let X_1, X_2, X_3 take N observations and that the variables are measured from their means.

Now, in the usual notation

$\rho_{1.23}$ = estimated value of X_1 given by the plane of regression of X_1 on X_2 and X_3.

$$\therefore \rho_{1.23} = b_{12.3} X_2 + b_{13.2} X_3 \quad \text{...(i)}$$

Since X_i' are measured from their means,

$$E(X_1) = E(X_2) = E(X_3) = 0$$

$$\therefore E[\rho_{1.23}] = 0 \quad \text{...(ii)}$$

Now, by definition,

$$R_{1.23} = \frac{\text{Cov}(X_1, \rho_{1.23})}{\sqrt{\text{Var}(X_1) \cdot \text{Var}(\rho_{1.23})}} \quad \text{...(iii)}$$

Now, $\text{Cov}(X_1, \rho_{1.23}) = E[X_1 \rho_{1.23}] - E(X_1) \cdot E(\rho_{1.23})$

$$= E[X_1 \cdot \rho_{1.23}] \qquad [\because E(X_1) = 0]$$

$$= \frac{1}{N} \Sigma X_1 \rho_{1.23} \qquad [\because E(\rho_{1.23}) = 0]$$

$$= \frac{1}{N} \Sigma X_1 (b_{12.3} X_2 + b_{13.2} X_3)$$

$$= b_{12.3}\frac{\Sigma X_1 X_2}{N} + b_{13.2}\frac{1}{N}\Sigma X_1 X_3$$

$$= \frac{\sigma_1}{\sigma_2}\left(\frac{r_{12} - r_{13}\, r_{23}}{1 - r_{23}^2}\right)\cdot r_{12}\,\sigma_1\,\sigma_2 + \frac{\sigma_1}{\sigma_3}\left(\frac{r_{13} - r_{12}\, r_{23}}{1 - r_{32}^2}\right)\cdot r_{13}\,\sigma_1\,\sigma_3$$

$$= \sigma_1^2\left[\frac{r_{12}^2 + r_{13}^2 - 2\cdot r_{12}\cdot r_{13}\cdot r_{23}}{1 - r_{23}^2}\right] \qquad \text{...(iv)}$$

$$\text{Var}(\rho_{1.23}) = E\left[\rho_{1.23}^2\right] - \left[E/\rho_{1.23}\right]^2 \qquad \text{...(by definition)}$$

$$= E\left[\rho_{1.23}^2\right] \qquad \left[\because E(\rho_{1.23}) = 0 \text{ From (ii)}\right]$$

$$= \frac{1}{N}\Sigma\rho_{1.23}^2$$

$$= \frac{1}{N}\Sigma\,(b_{12.3}X_2 + b_{13.2}X_3)^2$$

$$= b_{12.3}^2\frac{1}{N}\Sigma X_2^2 + b_{13.2}^2\frac{1}{N}\Sigma X_3^2 + 2\,b_{12.3}\;b_{13.3}\frac{1}{N}\Sigma X_2\cdot X_3$$

$$= \left[\frac{\sigma_1}{\sigma_2}\left(\frac{r_{12} - r_{13}\cdot r_{23}}{1 - r_{23}^2}\right)\right]^2\cdot\sigma_2^2 + \left[\frac{\sigma_1}{\sigma_3}\left(\frac{r_{13} - r_{12}\cdot r_{23}}{1 - r_{23}^2}\right)\right]^2\cdot\sigma_3^2$$

$$+ 2\cdot\frac{\sigma_1}{\sigma_2}\left(\frac{r_{12} - r_{13}\cdot r_{23}}{1 - r_{23}^2}\right)\cdot\frac{\sigma_1}{\sigma_3}\left(\frac{r_{13} - r_{12}\cdot r_{23}}{1 - r_{23}^2}\right)\cdot r_{23}\,\sigma_2\cdot\sigma_3$$

$$= \frac{\sigma_1^2}{\left(1 - r_{23}^2\right)^2}\left[r_{12}^2 + r_{13}^2\, r_{23}^2 - 2\, r_{12}\, r_{13}\, r_{23} + \left(r_{13}^2 + r_{12}^2\, r_{23}^2 - 2\, r_{12}\, r_{13}\, r_{23}\right)\right.$$

$$\left. + 2\left(r_{12}\, r_{13} - r_{12}^2\, r_{23} - r_{13}^2\, r_{23} + r_{12}\, r_{13}\, r_{23}^2\right) r_{23}\right]$$

On simplification,

$$= \frac{\sigma_1^2}{\left(1 - r_{23}^2\right)^2}\left[r_{12}^2 + r_{13}^2 - 2\, r_{12}\, r_{13}\, r_{23}\left(-r_{12}^2\, r_{23}^2 + r_{13}^2\cdot r_{23}^2 - 2\, r_{12} + r_{13}\right) r_{23}^2\right]$$

$$= \frac{\sigma_1^2}{\left(1 - r_{23}^2\right)^2}\left[r_{12}^2 + r_{13}^2 - 2\, r_{12}\, r_{13}\, r_{23} - r_{23}^2\left(r_{12}^2 + r_{13}^2 - 2\, r_{12}\, r_{13}\, r_{23}\right)\right]$$

$$= \frac{\sigma_1^2}{\left(1-r_{23}^2\right)^2}\left[\left(r_{12}^2+r_{13}^2-2\,r_{12}\,r_{13}\,r_{23}\right)\left(1-r_{23}^2\right)\right]$$

$$= \frac{\sigma_1^2}{\left(1-r_{23}^2\right)^2}\left[r_{12}^2+r_{13}^2-2\,r_{12}\,r_{13}\,r_{23}\right] \qquad \text{...(v)}$$

Substituting from Equations (iv) and (v) in (iii), we get

$$R_{1.23} = \frac{\left(r_{12}^2+r_{13}^2-2\,r_{12}\cdot r_{13}\cdot r_{23}\right)\cdot \sigma_1^2/(1-r_{23})^2}{\sqrt{\sigma_1^2\left(r_{12}^2+r_{13}^2-2r_{12}\cdot r_{13}\cdot r_{23}\right)\cdot \sigma_1^2/(1-r_{23})^2}}$$

Squaring, we get

$$R_{1.23}^2 = \frac{r_{12}^2+r_{13}^2-2\,r_{12}\cdot r_{13}\cdot r_{23}}{1-r_{23}^2}$$

$$\therefore 1-R_{1.23}^2 = \frac{1-r_{12}^2-r_{13}^2-r_{23}^2+2\,r_{12}\,r_{13}\,r_{23}}{1-r_{23}^2} = \frac{W}{W_{11}}$$

Where w and w_{11} have usual meaning,

$$\therefore\ R_{1.23}^2 = 1-\frac{W}{W_{11}} \qquad \text{...(vi)}$$

$$\therefore\ R_{1.23}^2 = \sqrt{1-\frac{W}{W_{11}}} \qquad \text{...(vii)}$$

☞ Remarks

(1). In multiple correlation coefficient, the primary subscript stands for dependent variable and the secondary subscripts refer to the independent variables.

Ordering of secondary subscripts is immaterial

For example

$R_{1.23}^2 = R_{1.32}$: $\qquad R_{2.13} = R_{2.31}$, etc

(2) We have, Cov $(X_1, \rho_{1.23})$ = var $(\rho_{1.23}) \geq 0$,

Since var (X) is always non-negative,

$\therefore$ Cov $(X_1, \rho_{1.23}) \geq 0$; $\quad \therefore\ R_{1.23} \geq 0$

Also, Karl Pearson's correlation coefficient between X_1 and $\rho_{1.23}$ cannot exceed 1.

$\therefore 0 \leq R_{1.23} \leq 1$

Thus, multiple correlation coefficient is always non-negative and lies between 0 and 1.

$R_{1.23}$ = Multiple Correlation Coefficient coefficient of X_1 on X_2 and X3

- $$R^2{}_{1.23} = \frac{r^2{}_{12} + r^2{}_{13} - 2r_{12}r_{13}r_{23}}{1 - r^2{}_{23}}$$

$$R^2{}_{2.13} = \frac{r^2{}_{21} + r^2{}_{23} - 2r_{12}r_{13}r_{23}}{1 - r^2{}_{13}}$$

$$R^2{}_{3.12} = \frac{r^2{}_{21} + r^2{}_{31} - 2r_{12}r_{13}r_{23}}{1 - r^2{}_{21}}$$

- In a trivariate distribution, if $r_{12} = 0.7, r_{13} = 0.61\ and\ r_{23} = 0.4$

  Find all multiple correlation coefficients.

$$R^2{}_{1.23} = \frac{r^2{}_{12} + r^2{}_{13} - 2r_{12}r_{13}r_{23}}{1 - r^2{}_{23}}$$

$R_{1.23} = 0.6196$

$R_{2.13} = 0.4912$

$R_{1.23} = 0.6111$

Chapter 5

1) A random sample of n=6 has the element 6,10,13,14,18,20. Compute the following the following point estimation

1) Population mean
2) Population standard deviation
3) The standard error of mean

1. Population mean:
   The point estimate of the population mean is the sample mean, which is given by:

$\bar{x} = (6+10+13+14+18+20) / 6 = 81 / 6 = 13.5$

Therefore, the point estimate of the population mean is 13.5.

2. Population standard deviation:
   The point estimate of the population standard deviation is the sample standard deviation, which is given by:

$s = \sqrt{[\Sigma(xi - \bar{x})^2 / (n-1)]}$

where xi is the i-th element of the sample.

First, we need to calculate the sample mean:

$\bar{x} = (6+10+13+14+18+20) / 6 = 81 / 6 = 13.5$

Next, we can calculate the sample variance:

$s^2 = \Sigma(xi - \bar{x})^2 / (n-1) = [(6-13.5)^2 + (10-13.5)^2 + (13-13.5)^2 + (14-13.5)^2 + (18-13.5)^2 + (20-13.5)^2] / (6-1) = 48.5$

Finally, we can calculate the sample standard deviation:

$s = \sqrt{48.5} \approx 6.963$

Therefore, the point estimate of the population standard deviation is approximately 6.963.

3. Standard error of mean:
   The standard error of the mean is given by:

$SEM = s / \sqrt{n}$

where s is the sample standard deviation and n is the sample size.

In this case, we have:

$s = \sqrt{48.5} \approx 6.963$
$n = 6$

Therefore, the standard error of the mean is:

$SEM = 6.963 / \sqrt{6} \approx 2.839$

Therefore, the point estimate of the standard error of the mean is approximately 2.839.

2) Explain Characteristics of Estimation.

Estimation

- In most statistical research studies, population parameters are usually unknown and have to be estimated from a sample.
- Estimators = random variables used to estimate population parameters (mean, variance)
- Estimates = specific values of the population parameters

Types of estimates

- Point estimate = estimate that specifies a single value of the population
- Interval estimate = estimate that specifies a range of values

Properties of a good estimator

Let θ = a population parameter

Let = $\hat{\theta}$ a sample estimate of that parameter. Desirable properties of $\hat{\theta}$ are:

1. Unbiased:

   Expected value = the true value of the parameter

   $E(\hat{\theta}) = \theta$.

   For example, $E(\bar{X}) = \mu$, $E(s^2) = \sigma^2$.

2. Efficiency:

   The most efficient estimator among a group of unbiased estimators is the one with the smallest variance.

   For example, both the sample mean and the sample median are unbiased estimators of the mean of a normally distributed variable. However, $\bar{X}$ *has the smallest variance.*

3. Sufficiency:

An estimator is said to be sufficient if it uses all the information about the population parameter that the sample can provide.

The sample median is not sufficient, because it only uses information about the ranking of observations. The sample mean is sufficient.

4. Consistency

An estimator is said to be consistent if it yields estimates that converge in probability to the population parameter being estimated as N becomes larger.

That is, as N tends to infinity,

$E(\hat{\theta}) = \theta$, $V(\hat{\theta}) = 0$.

For example, as N tends to infinity,

$V(\overline{X}) = \sigma^2/N = 0.$

3) Explain types of estimation with examples

Estimation

- In most statistical research studies, population parameters are usually unknown and have to be estimated from a sample.
- Estimators = random variables used to estimate population parameters (mean, variance)
- Estimates = specific values of the population parameters

Types of estimates

- Point estimate = estimate that specifies a single value of the population
- Interval estimate = estimate that specifies a range of values

Example

A poll may seek to estimate the proportion of adult residents of a city that support a proposition to build a new sports stadium.

Out of a random sample of 200 people, 106 say they support the proposition.

Thus in the sample, 0.53 of the people supported the proposition.

This value of 0.53 is called a point estimate of the population proportion.

It is called a point estimate because the estimate consists of a single value or point.

Interval Estimate

- Point estimates are usually supplemented by interval estimates called confidence intervals.
- Confidence intervals are intervals constructed using a method that contains the population parameter a specified proportion of the time.
- For example, if the pollster used a method that contains the parameter 95% of the time it is used, he or she would arrive at the following 95% confidence interval: $0.46 < \pi < 0.60$.
- The pollster would then conclude that somewhere between 0.46 and 0.60 of the population supports the proposal.
- The media usually reports this type of result by saying that 53% favor the proposition with a margin of error of 7%.

In statistics, estimation refers to the process of using sample data to make inferences or predictions about a population parameter. There are different types of estimation methods that are used depending on the type of data and the research question. Some of the most common types of estimation are:

1. Point estimation:

Point estimation involves estimating a single value as the best approximation of a population parameter. This method is used when the goal is to estimate a specific value of a parameter. For example, estimating the mean height of a population based on a sample of individuals. A common example of point estimation is the sample mean, which is used to estimate the population mean.

2. Interval estimation:

Interval estimation involves estimating a range of values within which the population parameter is likely to fall. This method is used when the goal is to estimate the range of values that the parameter is likely to take. For example, estimating the population mean height along with a confidence interval. A common example of interval estimation is a confidence interval, which provides a range of values that the population parameter is likely to take.

4) Prove that sample mean is unbiased estimator of population mean

To prove that the sample mean is an unbiased estimator of the population mean, we need to show that the expected value of the sample mean is equal to the population mean.

Let $X1, X2, \ldots, Xn$ be a random sample of size n from a population with mean μ and variance $σ^2$. Then, the sample mean is defined as:

$\bar{x} = (X1 + X2 + \ldots + Xn) / n$

The expected value of the sample mean is:

$E(\bar{x}) = E[(X1 + X2 + \ldots + Xn) / n]$

Using linearity of expectation, we can write:

$E(\bar{x}) = (1/n) * [E(X1) + E(X2) + \ldots + E(Xn)]$

Since the sample is drawn from a population with mean μ, we know that:

$E(Xi) = \mu$, for $i = 1, 2, \ldots, n$

Therefore, we can simplify the above equation as:

$E(\bar{x}) = (1/n) * [\mu + \mu + \ldots + \mu]$

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

Hence, we have shown that the expected value of the sample mean is equal to the population mean. Therefore, the sample mean is an unbiased estimator of the population mean.

Prove that sample mean is unbiased estimator of population mean

We have:

$$\begin{aligned}
\mathsf{E}\left(\bar{X}\right) &= \mathsf{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}\left(X_i\right) && \text{Expectation is Linear} \\
&= \frac{1}{n}\sum_{i=1}^{n} \mu && \text{as } \mathsf{E}\left(X_i\right) = \mu \\
&= \frac{n}{n}\mu && \text{as } \sum_{i=1}^{n} 1 = n \\
&= \mu
\end{aligned}$$

So $\bar{X}$ is an unbiased estimator of μ.

5) Discuss in brief about Method of Moments estimators

The method of moments (MoM) is a statistical technique used to estimate population parameters. This technique involves equating the sample moments (such as the sample mean or sample variance) to the corresponding population moments and solving for the unknown parameters.

The basic idea behind the MoM estimator is that the sample moments are calculated from the sample data and are assumed to be close to the corresponding population moments. The population moments are functions of the population parameters, which are unknown. The MoM estimator uses the sample moments to estimate the unknown population parameters.

The MoM estimator can be used to estimate various population parameters, such as the mean, variance, skewness, and kurtosis. For example, if we want to estimate the population mean, we can use the sample mean as the MoM estimator. Similarly, if we want to estimate the population variance, we can use the sample variance as the MoM estimator.

The MoM estimator has some advantages and disadvantages. One advantage is that it is easy to compute and does not require any assumptions about the probability distribution of the population. Another advantage is that it is consistent, meaning that as the sample size increases, the MoM estimator converges to the true population parameter.

However, the MoM estimator has some limitations. One limitation is that it may not be the most efficient estimator, meaning that it may have higher variance than other estimators. Another limitation is that it may not be robust to outliers or other deviations from the assumed population distribution.

In summary, the method of moments is a simple and widely used technique for estimating population parameters. It can be used to estimate various parameters, but it has some limitations and may not always be the best estimator.

The method of moments is a technique for estimating the parameters of a probability distribution based on matching the moments of the estimated distribution to the corresponding moments of the sample. The method of moments estimators (MMEs) are the estimates obtained using this technique.

The method of moments is based on the idea that the moments of the probability distribution are related to the parameters of the distribution. Specifically, if μk represents the kth moment of the distribution, then we can equate the sample moments to their corresponding population moments to obtain estimates of the parameters.

For example, suppose we have a random sample X1, X2, ..., Xn from a normal distribution with unknown mean μ and variance σ^2. The first two sample moments are:

$\hat{\mu}_1$ = (X1 + X2 + ... + Xn) / n

$\hat{\mu}_2$ = ($X_1^2 + X_2^2 + ... + Xn^2$) / n

The first two population moments of a normal distribution are:

$\mu_1 = \mu$

μ_2 = μ^2 + σ^2

By equating the sample moments to their corresponding population moments, we obtain the following estimators for the parameters:

$\hat{\mu} = \hat{\mu}_1$

$\hat{\sigma}$^2 = $\hat{\mu}_2 - \hat{\mu}_1^2$

These are the method of moments estimators for the mean and variance of a normal distribution.

The advantages of the method of moments are that it is easy to understand and apply, and it often provides reasonable estimates for the parameters of a distribution. However, the method of moments has limitations in that it may not always produce consistent or efficient estimators, especially for complex distributions with many parameters.

Overall, the method of moments is a useful technique for obtaining initial estimates of the parameters of a distribution, which can be refined using other methods such as maximum likelihood estimation or Bayesian inference.

5.3 METHOD OF MOMENTS

This method was discovered in detail by karl pearson. Let $f(x ; \theta_1, \theta_2, \ldots, \theta_k)$ be the density function of the parent population with k parameters $\theta_1, \theta_2, \ldots, \theta_k$.

If μ_r' denotes the r^{th} Moment about origin,

then $\mu_r' = \int_{-\infty}^{\infty} x^r f(x ; \theta_1, \theta_2, \ldots, \theta_k)\, dx, \ (r = 1, 2, \ldots, k)$

In general, $\mu_1', \mu_2', \ldots, \mu_k'$ will be function of the parameters $\theta_1, \theta_2, \ldots, \theta_k$.

5.3.1 Working Method (of Moments)

(1) Let x_i, $i = 1, 2, \ldots$, be a random sample of size n from the given population.

(2) Solve k-equations for $\theta_1, \theta_2, \ldots, \theta_k$ in terms of $\mu_1', \mu_2', \ldots, \mu_k'$

(3) Replace these moments u_r', $r = 1, 2, \ldots k$ by the sample moments,

e.g. $\hat{\theta}_i = \theta_i\left(\hat{\mu}_1', \hat{\mu}_2', \ldots, \hat{\mu}_k'\right) = \theta_i\left(m_1', m_2', \ldots, m_k'\right)$;

$i = 1, 2, \ldots, k.$

Where m_i' is the i^{th} Moment about origin in the sample.

(4) By the method of moments $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ are the required estimates of $\theta_1, \theta_2, \ldots, \theta$ respectively.

Remarks

(1) Sample moments are consistent estimators of the corresponding population moments

(2) The estimates obtained by the method of moments are asymptotically normal but no in general, efficient.

(3) Generally the method of moments yields less efficient estimators than those obtain from the principle of maximum likelihood.

6) Explain Point Estimation in detail with its properties -same as above

7) Explain method of maximum likelihood with its advantageous and disadvantages.

The method of maximum likelihood (ML) is a statistical method used to estimate the parameters of a statistical model based on observed data. The basic idea of the ML method is to find the values of the model parameters that maximize the likelihood function, which is the probability of observing the given data, assuming that the model is correct and the parameters take certain values.

Advantages of the Method of Maximum Likelihood:

1. Efficiency: The ML method is known to be the most efficient method of estimation in terms of asymptotic variance, meaning that as the sample size increases, the variance of the estimator approaches zero faster than any other unbiased estimator.
2. Consistency: The ML estimator is consistent, meaning that as the sample size increases, the estimator converges to the true value of the parameter being estimated.
3. Asymptotic Normality: The ML estimator is asymptotically normal, meaning that as the sample size increases, the distribution of the estimator approaches a normal distribution. This makes it possible to use statistical tests and confidence intervals based on the normal distribution to make inferences about the population parameters.
4. Flexibility: The ML method can be applied to a wide range of statistical models, including linear regression models, generalized linear models, and nonlinear models.

Disadvantages of the Method of Maximum Likelihood:

1. Complex Computation: The ML method often requires complex computations, such as finding the maximum of a complex likelihood function or solving a system of nonlinear equations.
2. Sensitivity to Model Assumptions: The ML method assumes that the data are generated by a specific probability distribution, and if this assumption is incorrect, the ML estimator may be biased or inefficient.
3. Sample Size: The ML method may require large sample sizes to produce accurate estimates of the parameters.
4. Uniqueness of Maximum: The likelihood function may have multiple maxima or saddle points, which can make it difficult to find the global maximum and lead to nonunique solutions.

In summary, the method of maximum likelihood is a powerful method of estimation that has many advantages, including efficiency, consistency, asymptotic normality, and flexibility. However, it also has some disadvantages, including complex computation, sensitivity to model assumptions, sample size requirements, and the possibility of nonunique solutions.

Advantages and disadvantages of maximum likelihood methods

List the advantages and disadvantages of maximum likelihood methods

Advantages.

- Simple to apply
- Lower variance than other methods (i.e. estimation method least affected by sampling error) and unbiased as the sample size increases.
- The method is statistically well understood
- Able to analyze statistical models with different characters on the same basis. Maximum likelihood provides a consistent approach to parameter estimation problems. This means that maximum likelihood estimates can be developed for a large variety of estimation situations.
- Once a maximum-likelihood estimator is derived, the general theory of maximum-likelihood estimation provides standard errors, statistical tests, and other results useful for statistical inference.

Disadvantages.

- Computationally intensive and so extremely slow (though this is becoming much less of an issue)
- Frequently requires strong assumptions about the structure of the data
- The estimates that are obtained using this method are often biased. That is, they contain a systematic error of estimation. This is true for small samples. The optimality properties may not apply for small samples.
- MLE is inapplicable for the analysis of non-regular populations (Non-regular distributions are models where a parameter value is constrained by a single observed value)

5.4 METHOD OF MAXIMUM LIKELIHOOD ESTIMATION

This is the most general method of estimation and was formulated by C.F. Gauss.

5.4.1 Likelihood Function

Definition : Let $x_1, x_2, \ldots, x_n$ be a ranodm sample of size n from a population with density function f (x, θ). Then the likelihood funciton of the sample values $x_1, x_2, \ldots, x_n$, usually denoted by L = L (θ) is their joint density function given by :

$$L = f(x_1, \theta) \cdot f(x_2, \theta) \ldots f(x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta) \quad \ldots(i)$$

L gives the relative likelihood that the random variables assume a particular set of values $x_1, x_2, \ldots, x_n$.

For a given sample $x_1, x_2, \ldots, x_n$, L becomes a function of the variable θ, the parameter.

The principle of maximum likelihood consists in finding an estimator for the unknown parameter $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$, which maximizes the likelihood function L (θ) ; i.e. we want to find $\hat{\theta} = \left(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k\right)$ so that

$$L\left(\hat{\theta}\right) > L(\theta), \forall \theta \in \Theta \text{ ; i.e., } L\left(\hat{\theta}\right) = \text{Sup. } L(\theta), \quad \forall \theta \in \Theta.$$

Thus $\hat{\theta}$ is the solution of

$$\frac{\partial L}{\partial \theta} = 0 \text{ and } \frac{\partial^2 L}{\partial \theta^2} < 0 \quad \ldots(ii)$$

8) Explain in details moments and its type

https://www.analyticsvidhya.com/blog/2022/01/moments-a-must-known-statistical-concept-for-data-science/#:~:text=%E2%80%93%20Standardized%20Moments-,What%20is%20the%20Moment%20in%20Statistics%3F,X%E2%81%B4)%2C%E2%80%A6%2C%20etc.

Moments are a set of numerical values that describe the shape and location of a probability distribution or a set of data. Moments can be used to estimate various parameters of a distribution, such as the mean, variance, and skewness. The moment of a distribution is calculated by taking a mathematical expectation of the distribution raised to some power.

There are several types of moments, including:

1. Raw Moments: The raw moment of a probability distribution is calculated as the expected value of the distribution raised to some power, usually k. The kth raw moment of a continuous random variable X is defined as:

Mk = E(X^k)

The first raw moment is the mean of the distribution, and the second raw moment is the variance of the distribution.

2. Central Moments: The central moment of a probability distribution is calculated as the expected value of the deviation of the distribution from its mean, raised to some power, usually k. The kth central moment of a continuous random variable X is defined as:

Mk = E[(X - μ)^k]

where μ is the mean of the distribution. The second central moment is the variance of the distribution, and the third central moment is the skewness of the distribution.

3. Standardized Moments: The standardized moment of a probability distribution is calculated as the ratio of a central moment to a power of the standard deviation. The kth standardized moment of a continuous random variable X is defined as:

$$\gamma k = E[(X - \mu)^k] / \sigma^k$$

where μ is the mean of the distribution and σ is the standard deviation of the distribution. The standardized moments are useful for comparing distributions with different scales and locations.

4. Cumulant Moments: The cumulant moment of a probability distribution is defined as the logarithm of the moment-generating function of the distribution. The kth cumulant moment of a continuous random variable X is denoted by ck and is defined as:

$$ck = (d^k/dt^k) \ln[M(t)]$$

where M(t) is the moment-generating function of the distribution. The cumulant moments are useful for characterizing the shape and location of a distribution.

In summary, moments are a useful tool for describing the properties of a probability distribution or a set of data. The different types of moments, including raw moments, central moments, standardized moments, and cumulant moments, provide different measures of the shape and location of the distribution, and they can be used to estimate various parameters of the distribution.

9) Show that the sample variance (S^2) is an unbiased estimator of σ^2.

To show that the sample variance (S^2) is an unbiased estimator of σ^2, we need to show that the expected value of S^2 equals σ^2.

The sample variance is defined as:

$S^2 = 1/(n-1) * \Sigma(xi - \bar{x})^2$

where n is the sample size, xi is the ith observation, $\bar{x}$ is the sample mean, and Σ denotes the sum over all observations.

The expected value of S^2 is:

$E(S^2) = E[1/(n-1) * \Sigma(xi - \bar{x})^2]$

Using the properties of the expected value operator, we can simplify this expression as follows:

$E(S^2) = 1/(n-1) * E[\Sigma(xi - \bar{x})^2]$

Now, we can expand the expression inside the expected value operator as follows:

$\Sigma(xi - \bar{x})^2 = \Sigma(xi^2 - 2xi\ \bar{x} + \bar{x}^2)$

$= \Sigma(xi^2) - 2\bar{x}\ \Sigma(xi) + n\bar{x}^2$

Substituting this expression back into the expected value expression, we get:

$E(S^2) = 1/(n-1) * E[\Sigma(xi^2) - 2\bar{x}\ \Sigma(xi) + n\bar{x}^2]$

Using the linearity property of the expected value operator, we can simplify this expression as follows:

$E(S^2) = 1/(n-1) * [E(\Sigma(xi^2)) - 2\bar{x}\ E(\Sigma(xi)) + n\ E(\bar{x}^2)]$

Since the observations are assumed to be independent and identically distributed with a common variance σ^2, we have:

$E(xi) = \mu$ (the population mean)

$E(xi^2) = Var(xi) + E(xi)^2 = \sigma^2 + \mu^2$

$E(\bar{x}) = \mu$ (the population mean)

$E(\bar{x}^2) = Var(\bar{x}) + E(\bar{x})^2 = \sigma^2/n + \mu^2$

Substituting these expressions back into the previous expression, we get:

$E(S^2) = 1/(n-1) * [E(\Sigma(xi^2)) - 2n\ \bar{x}^2 + n\ \sigma^2/n + n\ \mu^2]$

Simplifying this expression, we get:

$E(S^2) = 1/(n-1) * [\Sigma[E(xi^2)] - n\ \bar{x}^2 - \sigma^2]$

Substituting $E(xi^2) = \sigma^2 + \mu^2$ and $E(\bar{x}) = \mu$, we get:

$E(S^2) = 1/(n-1) * [(n-1)\sigma^2 - n\sigma^2/(n-1)]$

Simplifying this expression, we get:

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

Therefore, the expected value of S^2 is equal to σ^2, which means that S^2 is an unbiased estimator of σ^2.

Prove that sample variance is unbiased estimator of population variance

Imagine that we are drawing a sample of size (n) from a population which has parameters μ and σ.

The variance of a particular sample (of size n) is given by the following equation.

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

Obviously, the value of individual s^2 is likely to be different from sample to sample.

While the sample variance itself is a random variable, the question is “what is the theoretical average or expectation of s^2 ?”.

That is, if we are to draw a large number of times (say 1 million) -each time a sample of size n, then what can be expected as the average value of s^2?

Taking the expectation both sides,

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

We need some basic results:

$$Var(x_i) = E(x_i^2) - [E(x_i)]^2$$

$$\sigma^2 = E(x_i^2) - \mu^2$$

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

$$Var(\bar{X}) = E(\bar{X}^2) - [E(\bar{X})]^2$$

$$\frac{\sigma^2}{n} = E(\bar{X}^2) - \mu^2$$

$$\therefore E(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2$$

Using the results

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

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

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

$$E[s^2] = \frac{1}{n-1}[\sum_{i=1}^{n} E(x_i^2) - 2nE(\bar{X}^2) + nE(\bar{X}^2)]$$

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

$$E[s^2] = \frac{1}{n-1}[nE(x_i^2) - nE(\bar{X}^2)$$

Using earlier results

$$E[s^2] = \frac{1}{n-1}[n(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2))$$

$$E[s^2] = \frac{1}{n-1}[n(\sigma^2 + \mu^2) - \sigma^2 - n\mu^2)$$

$$\therefore E[s^2] = \sigma^2$$

Hence Proved

Chapter 6

1. Define and explain Hypothesis and tests of Hypothesis

Introduction to Hypothesis Testing

- Hypothesis testing is one of the most important concepts in Statistics
- Heavily used by **Statisticians**, **Machine Learning Engineers**, and **Data Scientists**
- Statistical tests are used to check whether the **null hypothesis** is rejected or not rejected.
- Statistical tests assume a null hypothesis **of no relationship or no difference between groups.**

- Definition:

A hypothesis is defined as a formal statement, which gives the explanation about the relationship between the two or more variables of the specified population

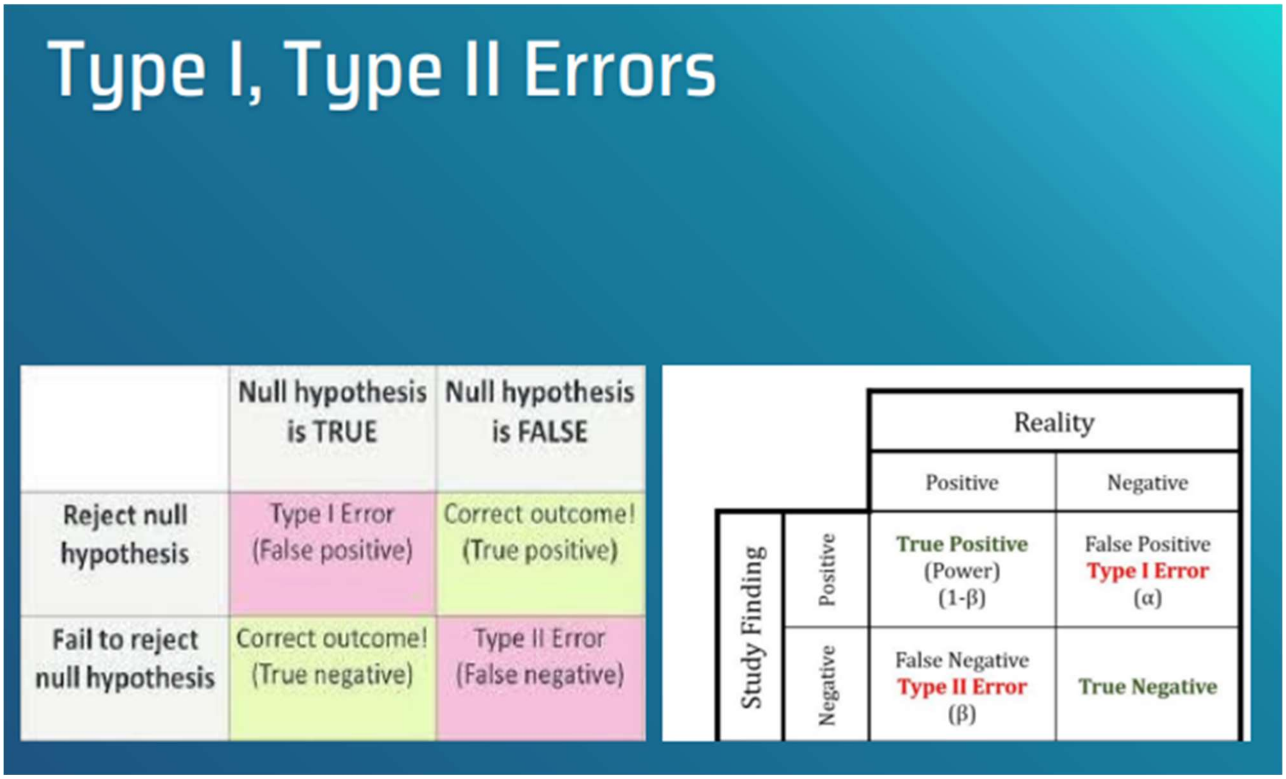
Example based on a sample data we may wish to decide whether a serum is really effective in curing Corona

What is test of Hypothesis?

- If on the supposition that a particular hypothesis is true we find that results observed in a random sample differ markedly from those expected, we say that observed differences are significant and we reject the hypothesis
- Procedures that enable us to decide to accept or reject hypothesis are called **test of hypothesis, test of significance, decision rules**

Type I and Type II Errors

- **Type I error**:- Rejecting a hypothesis when it happens to be true
- **Type II error**:- Accepting a hypothesis when it is to be rejected
- These errors have to be minimized but the decrease in one causes the increase in the other
- The best solution is to increase the sample size



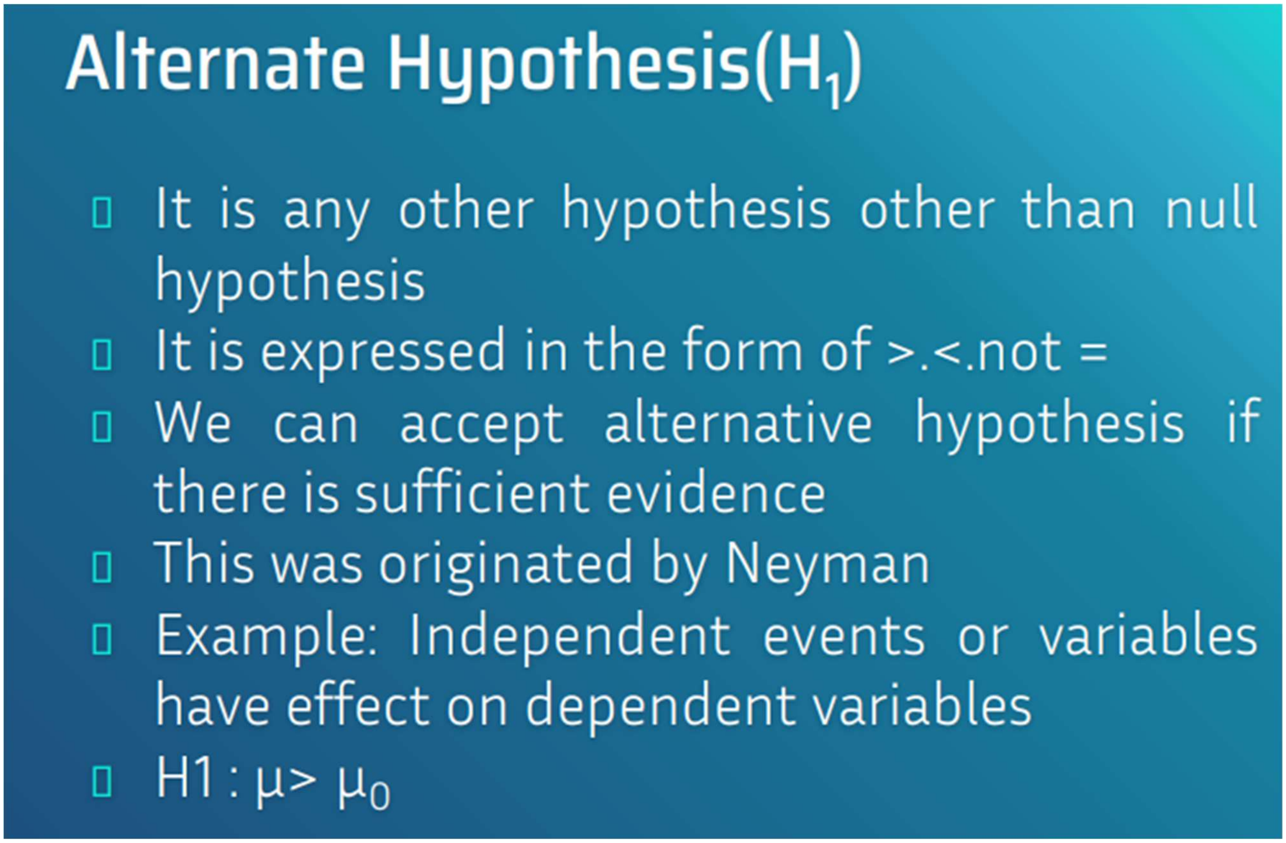
2. <u>What do you mean by Null and alternative Hypothesis Explain</u>

Null Hypothesis(H_0)

- It is a statistical hypothesis which is to be actually tested for acceptance or rejection
- It is the hypothesis which is tested for possible rejection under the assumption that it is true
- It is expressed in the form of equality
- Example:- Independent variables have no effect on the dependent variables.

Examples of Null Hypothesis

- Null hypothesis is always a simple hypothesis stated as an equality specifying an exact value of the parameter
- Examples
  - Population mean equals to a specified constant μ_0
  - The difference between the sample means equals to a constant



3. <u>Note on Types of Errors -same as above</u>

4. Note on MP and UMP tests

MP and UMP are two types of statistical hypothesis tests that are commonly used in hypothesis testing.

MP Test:
MP stands for the most powerful test, which is a test that has the highest probability of rejecting the null hypothesis when it is false, given a certain level of significance. MP tests are typically used when the alternative hypothesis is one-sided. The MP test is based on the likelihood ratio, which is the ratio of the likelihoods of the data under the null and alternative hypotheses. The test statistic is compared to a critical value, which is determined based on the level of significance and the distribution of the test statistic. MP tests are used to test a single parameter of interest.

UMP Test:
UMP stands for uniformly most powerful test, which is a test that is the most powerful among all tests of a certain level of significance. UMP tests are typically used when the alternative hypothesis is two-sided. The UMP test is based on the Neyman-Pearson lemma, which states that the likelihood ratio test is the most powerful test among all tests of a certain level of significance. The test statistic is compared to a critical value, which is determined based on the level of significance and the distribution of the test statistic. UMP tests are used to test a single parameter of interest.

The main difference between MP and UMP tests is that MP tests are used when the alternative hypothesis is one-sided, while UMP tests are used when the alternative hypothesis is two-sided. MP tests are also used when the sample size is small, while UMP tests are used when the sample size is large. In addition, MP tests are more powerful than UMP tests, but they are also more complex and difficult to calculate.

In summary, MP and UMP tests are two different types of hypothesis tests that are used in statistical inference. MP tests are used when the alternative hypothesis is one-sided, while UMP tests are used when the alternative hypothesis is two-sided. Both tests are used to test a single parameter of interest, and the choice of test depends on the nature of the alternative hypothesis and the sample size.

...maximum, the likelihood ratio is bounded between zero and one. ... as the constrained maximum cannot exceed

6.9 MP AND UMP-TEST

Uniformly most powerful test

In statistical hypothesis testing, a uniformly most powerful test is a hypothesis test which has the greatest power among all possible tests of a given size α. For example, according to Neyman-Pearson lemma, the likelihood-ratio test is UMP for testing simple hypothesis.

6.9.1 Setting of UMP

Let X denote a random vector, taken from a parameters family of probability density function $f_\theta(x)$: And it depends on the unknown deterministic parameter $\theta \in \Theta$.

The parameter space Θ is partitioned into two disjoint sets Θ_0 and Θ_1.

Let H_0 denote the hypothesis that $\theta \in \Theta_0$ and let H_1 denote the hypothesis that $\theta \in \Theta_1$.

The binary test of hypothesis is performed using a test function $\phi(x)$ with a reject region R (it is a subset of measurement space).

and $$\phi(x) = \begin{cases} 1, & \text{if } x \in R \\ 0, & \text{if } x \in R^c \end{cases}$$

Note that R and R^C disjoint covering of the measurement space.

6.9.2 Definition of Test Function

A test function $\phi(x)$ is UMP of size α if for any other test function $\phi'(x)$ satisfying :

$$\sup_{\theta \in \Theta_0} E[\phi'(x) \mid \theta] = \alpha' < \alpha = \sup_{\theta \in \Theta_0} E[\phi(x) \mid \theta]$$

6.9.3 Property of UMP

In general, UMP tests do not exist for vector parameters or for two-sided tests (a t... in which one hypothesis lies on both sides of the alternative)

In these situations, the most powerful test of a given size for one possible value of parameter (e.g. for θ_1 where $\theta_1 > \theta_0$) is different from the most powerful test of the sa... size for a different value of the parameter (e.g. for θ_2 where $\theta_2 < \theta_1$)

Thus no test is **uniformly most powerful** in such situations. **Uniformly** means **regardless** of the values of the unobservable parameters.

One test may be the most powerful one for a particular value of an unobservable parameter while a different test is the most powerful one for a different value of the parameter. A uniformly more powerful test remains the most powerful one regardless of the value of the parameters.

6.9.4 Most Powerful Test (MP)

It is a test, say δ if $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ if size α is MP, if it has the greatest power $\pi(\theta_1 \mid \delta)$ among all tests of size α or less.

Here π is the power function. And δ has the greatest capacity of detecting H_1 among tests of size at most α and these specific hypothesis.

6.9.5 Uniformly Most Powerful Test (UMP)

A test δ of $H_0 : \theta \in \Theta_0$ versus

$H_1 : \theta \in \Theta - \Theta_0$, size α is UMP if it has the greatest power.

$\pi_{\theta \in \Theta_1}(\theta \mid \delta)$ among all tests of size α or less.

"Uniformly" refers to all values of θ. We note the difference in the two statements with respect to the hypothesis and power.

A non-UMP test can be most powerful just for a specific value of θ

A UMP test is the **'most powerful'** test for each value of θ in H_1.

Level MP-test obeys the likelihood ratio inequalities. Note that the **most powerful test may not always be unique** as can be deduced from the lemma.

In fact, it may not exist at all.

A Uniformly most powerful Test is always unbiased , if it exists

Unbiasedness is the property that the probability of rejection is greater under any alternative distribution than it is under null distribution.

5. <u>Note on Neyman-Pearson Lemma</u>

6.8 NEYMAN-PEARSON LEMMA

- The Neyman-Pearson lemma is part of the Neyman-Pearson theory of statistical testing. It introduced concepts like errors of second kind, power function and inductive behaviour.
- The theory of significance testing allowed only one hypothesis. It involved only one type of error. The Neyman-Pearsonian theory of statistical testing allows investigating the **two types of errors.**
- Neyman and pearson proceeded to restrict their attention to the class of all α level tests to minimise type II errors. It is denoted by β. The Neyman-Pearson Lemma is a way to find out if the hypothesis test you are using is the one with the greatest statistical power.
- The lemma tells us that **good hypothesis tests are likelihood ratio tests.** The power of a hypothesis test is the probability that test correctly **rejects the null hypothesis** when the **alternate hypothesis** is true. The goal is to maximise this power, so that null hypothesis is rejected as much as possible when the alternate is true.

6.8.1 Proposition of Neyman-Pearson Lemma

Consider a test with hypothesis

$H_0 : \ \theta \equiv \theta_0$ and

$H_1 : \ \theta \equiv \theta_1$ where the probability density function (or probability mass function

is $f(x / \theta_i)$ for $i = 0, 1$

We denote the rejection region by R_a.

The Neyman-Pearson lemma states that a most powerful (MP) test satisfies the following : for some $\eta \geq 0$,

(i) $x \in R$ if $f(x|\theta_1) > \eta f(x|\theta_0)$, (ii) $x \in R^c$ if $f(x|\theta_1) < \eta f(x|\theta_0)$

and(iii) $P_{\theta_0}(x \in R) = \alpha$ for a prefixed significance level α.

Also, if there is at least one MP test that satisfies the two conditions, the lemma states that every existing α-level MP test should obey the likelihood ratio inequalities.

Note that the most powerful test may not always be unique. Also it is possible that it may not exist at all.

In practice, the likelihood ratio is often used directly to construct tests.

6.8.2 Likelihood-ratio Test

- In statistics, the likelihood-ratio test assesses the goodness of fit of two competing **statistical models** based on the ratio of their likelihoods.
- One is found by maximization over the entire parameter space and another is found after imposing some constraint.
- If the constraint (i.e., the null hypothesis) is supported by the **observed data,** the two likelihoods should not differ by more than **sampling error.**
- Thus the likelihood-ratio test tests whether this ratio is **significantly different** from one, or equivalently whether its natural logarithm is significantly different from zero.

6.8.3 Interpretation

Suppose that we have a **statistical model** with parameter space Θ.

A **null hypothesis** is stated by saying that the parameter Θ is in a specified subset Θ_0 of Θ.

The alternate hypothesis : θ is in complement of Θ_0, i.e., in $\Theta \sim \Theta_0$ and denoted by Θ_0^c.

The likelihood ratio test statistic for null hypothesis $H_0 : \theta \in \Theta_0$ is given by

$$\lambda_{LR} = -2 \log_e \left[\frac{\sup\limits_{\theta \in \Theta_o} L(\theta)}{\sup\limits_{\theta \in \Theta_o} L(\theta)} \right]$$

Where the quantity inside the bracket is likelihood ratio. Here 'sup' refers to supremum. As all likelihoods are positive, and as the constrained maximum cannot exceed the unconstrained maximum, the likelihood ratio is bounded between zero and one.

6.9 MP AND UMP-TEST

The Neyman-Pearson Lemma is a fundamental result in mathematical statistics that provides a framework for constructing tests of statistical hypotheses. The lemma was introduced by Jerzy Neyman and Egon Pearson in 1933, and it is named after them.

The Neyman-Pearson Lemma states that, among all tests of a given level of significance, the likelihood ratio test is the most powerful. This means that the likelihood ratio test has the highest probability of rejecting the null hypothesis when it is false, given a certain level of significance. The likelihood ratio test compares the likelihood of the observed data under the null hypothesis to the likelihood of the observed data under the alternative hypothesis. If the ratio of the two likelihoods is sufficiently large, the null hypothesis is rejected in favor of the alternative hypothesis.

The Neyman-Pearson Lemma is important because it provides a systematic and objective way to construct tests of statistical hypotheses. The likelihood ratio test is based on the maximum likelihood principle, which is a general principle that underlies many statistical methods. The lemma also provides a way to compare different tests and to choose the most powerful test for a given level of significance.

However, the Neyman-Pearson Lemma has some limitations. One limitation is that it only applies to tests of simple hypotheses, which are hypotheses that specify a single value for the parameter of interest. In addition, the likelihood ratio test can be computationally intensive and may not always be feasible to implement in practice. Finally, the Neyman-Pearson Lemma does not provide any guidance on how to choose the level of significance or how to interpret the results of a test.

In summary, the Neyman-Pearson Lemma is a fundamental result in statistical inference that provides a framework for constructing tests of statistical hypotheses. The lemma states that the likelihood ratio test is the most powerful among all tests of a given level of significance. While the lemma has some limitations, it provides a powerful and objective tool for hypothesis testing.