Index

Unit – 4 ⇝ Inferential Statistics – I

Unit – 4.1 ⇝ Correlation and Regression

Method – 1 ⇝ Correlation Coefficient

Example of Method-1: Correlation Coefficient

B 1 Calculate the coefficient of correlation between the given series:

x	54	57	55	57	56	52	59
y	36	35	32	34	36	38	35

Answer: $r = -0.4575$

B 2 Calculate the coefficient of correlation between the age of husband and wife for below:

Age of husband	35	34	40	43	56	20	38
Age of wife	32	30	31	32	53	20	33

Answer: $r = 0.9371$

B 3 Compute Karl Pearson's coefficient of correlation between x and y for the following data:

x	100	98	78	85	110	93	80
y	85	90	70	72	95	81	74

Answer: $r = 0.9603$

B 4 Calculate the coefficient of correlation for the following series:

x	65	66	67	67	68	69	70	72
y	67	68	65	68	72	72	69	71

Answer: $r = 0.6030$

<table>
<tr><td>B</td><td>5</td><td>Calculate the coefficient of correlation for the following series:
<table>
<tr><td>x</td><td>1100</td><td>1200</td><td>1300</td><td>1400</td><td>1500</td><td>1600</td><td>1700</td><td>1800</td><td>1900</td><td>2000</td></tr>
<tr><td>y</td><td>0.30</td><td>0.29</td><td>0.29</td><td>0.25</td><td>0.24</td><td>0.24</td><td>0.24</td><td>0.29</td><td>0.18</td><td>0.15</td></tr>
</table>
Answer: $r = -0.7906$</td></tr>
<tr><td>B</td><td>6</td><td>Find the correlation coefficient between the serum diastolic B.P. and serum cholesterol levels of 10 randomly selected data of 10 persons.
<table>
<tr><td>Person</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td></tr>
<tr><td>Cholesterol</td><td>307</td><td>259</td><td>341</td><td>317</td><td>274</td><td>416</td><td>267</td><td>320</td><td>274</td><td>336</td></tr>
<tr><td>B.P.</td><td>80</td><td>75</td><td>90</td><td>74</td><td>75</td><td>110</td><td>70</td><td>85</td><td>88</td><td>78</td></tr>
</table>
Answer: $r = 0.8088$</td></tr>
<tr><td>C</td><td>7</td><td>Find r_{xy} from given data if $n = 10, \sum(x - \bar{x})(y - \bar{y}) = 66, \sigma_x = 5.4,$ $\sigma_y = 6.2.$
Answer: $r = 0.1971$</td></tr>
<tr><td>C</td><td>8</td><td>Find r_{xy} from given data $n = 10, \sum(x - \bar{x})(y - \bar{y}) = 1650, \sigma_x^2 = 196,$ $\sigma_y^2 = 225.$
Answer: $r = 0.7857$</td></tr>
<tr><td>C</td><td>9</td><td>Determine the coefficient of correlation if $n = 8, \bar{x} = 0.5, \bar{y} = 0.5,$ $\sum x^2 = 44, \quad \sum y^2 = 44, \sum xy = -40.$
Answer: $r = -1$</td></tr>
</table>

Method – 2 ⇝ Rank Correlation Coefficient

Example of Method-2: Rank Correlation Coefficient

A 1

Two Judges in a beauty contest rank the 12 contestants as follows:

1st judge	1	2	3	4	5	6	7	8	9	10	11	12
2nd judge	12	9	6	10	3	5	4	7	8	2	11	1

What degree of agreement is there between the judges?

Answer: $\rho = -0.4545$

C 2

Ten competitors in a musical test were ranked by the three judges A, B, and C in the following order:

Rank by A	1	6	5	10	3	2	4	9	7	8
Rank by B	3	5	8	4	7	10	2	1	6	9
Rank by C	6	4	9	8	1	2	3	10	5	7

Using the rank correlation method, find which pair of judges has the nearest approach to common linking in music.

Answer: **Judges A and C has nearest approach**

$[\rho_{AB} = -0.2121, \rho_{BC} = -0.2970, \rho_{AC} = 0.6364]$

C 3

Obtain the rank correlation coefficient for the following data:

x	68	64	75	50	64	80	75	40	55	64
y	62	58	68	45	81	60	68	48	50	70

Answer: $\rho = 0.5455$

C 4

From the following data of the marks obtained by 8 students in Computer Networking (CN) and Compiler Design (CD) papers, compute rank coefficient of correlation.

CN	15	20	28	12	40	60	20	80
CD	40	30	50	30	20	10	30	60

Answer: $\rho = 0$

Method – 3 ⇝ Linear Regression

Example of Method-3: Linear Regression

B	1	Find the regression line of y on x for the following data: x: 2, 3, 4, 4, 5, 6, 6, 7, 7, 8, 10, 10 y: 1, 3, 2, 4, 4, 4, 6, 4, 6, 7, 9, 10 **Answer: $y = 0.9891x - 0.9166$**
B	2	Obtain two regression lines from the following data: x: 65, 66, 67, 67, 68, 69, 70, 72 y: 67, 68, 65, 68, 72, 72, 69, 71 **Answer: $x = 0.5455y + 30.3605$; $y = 0.6667x + 23.6644$**
B	3	The amount of chemical compound (y), which were dissolved in 100 grams of water at various temperatures (x): x: 15, 15, 30, 30, 45, 45, 60, 60 y: 12, 10, 25, 21, 31, 33, 44, 39 Find the equation of the regression line of y on x and estimate y if $x = 50°C$. **Answer: $y = 0.67x + 1.75$, 35.25**
C	4	For following data calculate the regression line of performing rating on experience and also estimate the probable performance if an operator has 11 years' experience. Operator: 1, 2, 3, 4, 5, 6 Performance rating: 78, 36, 98, 25, 75, 82 Experience: 84, 51, 91, 60, 68, 62 **Answer: $y = 0.4094\,x + 42.4494$; 46.9528 ; x = Experience**
C	5	The following values are available for the variable x & y. Obtain regression lines. $n = 10, \sum x = 30, \sum y = 40, \sum x^2 = 222, \sum y^2 = 985, \sum xy = 384.$ **Answer: $y = 2x - 2$; $x = 0.32y + 1.72$**

<table>
<tr><td>C</td><td>6</td><td>Find the lines of regression of y on x if $n = 9, \sum x = 30.3, \sum y = 91.1, \sum xy = 345.09, \sum x^2 = 115.11$. Also, find value of y(1.5) and y(5.0).

Answer: $y = 2.93x + 0.2568$; $y(1.5) = 4.6523$; $y(5.0) = 14.9083$</td></tr>
<tr><td>C</td><td>7</td><td>Find the regression lines from the following data. where, r = 0.5.
<table><tr><td></td><td>x</td><td>y</td></tr><tr><td>Mean</td><td>60</td><td>67.5</td></tr><tr><td>Standard deviation</td><td>15</td><td>13.5</td></tr></table>Answer: $y = 0.45x + 40.5$; $x = 0.5556y + 22.4970$</td></tr>
<tr><td>C</td><td>8</td><td>Find the regression equation showing the capacity utilization on production from the following data:
<table><tr><td></td><td>Average</td><td>Standard deviation</td></tr><tr><td>Production (lakh units)</td><td>35.6</td><td>10.5</td></tr><tr><td>Capacity utilization (%)</td><td>84.8</td><td>8.5</td></tr><tr><td>Correlation coefficient</td><td colspan="2">r = 0.62</td></tr></table>Estimate the production when capacity utilization is 70%.

Answer: $x = 0.5019y + 66.9324$; $y = 0.7659x - 29.3483$; 24.2627

x = Capacity utilization ; y = Production</td></tr>
</table>

Method – 4 ⇝ Curve Fitting

Example of Method-4.1: Fitting a Stright Line

B **2** By the method of least square, find the straight line that best fits the following data:

x	1	2	3	4	5
y	14	27	40	55	68

Answer: $y = 13.6x$

B **3** Fit a straight line for following data. Also, find y when x = 2.8

x	2	5	6	9	11
y	2	4	6	9	10

Answer: $y = -0.0244 + 0.9431x$; $y(2.8) = 2.6163$

B **4** Fit a straight line to the following data:

x	71	68	73	69	67	65	66	67
y	69	72	70	70	68	67	68	68

Answer: $y = 46.9394 + 0.3232x$

B **5** The weight of a calf taken at weekly intervals are given below. Fit a straight-line using method of least squares.

Age (x)	1	2	3	4	5	6	7	8	9	10
Weight (y)	52.5	58.7	65	70.2	75.4	81.1	87.2	95.5	102.5	108.4

Answer: $y = 45.6867 + 6.1752x$

B **6** The following show the gain in reading speed of 3 students in a speed-reading program, and the number of weeks they have been in the program:

No. of weeks	3	5	2	8	6	9	3	4
Speed gain	86	118	49	193	164	232	73	109

Find a straight line by the method of least squares.

Answer: $y = 3.3409 + 24.9318x$

C	9	If P is the pull required to lift a load W by means of a pulley block, find a linear approximation of the form $P = mW + c$ connecting P and W, using the following data:

P	13	18	23	27
W	51	75	102	119

Answer: $P = 0.2028W + 2.6580$

Example of Method-4.2: Fitting a Parabola

B **1** Fit a second-degree polynomial of y on x to the following data:

x	50	70	100	120
y	12	15	21	25

Answer: $\mathbf{y = 5.5259 + 0.1029x + 0.0005x^2}$

B **2** Fit a parabola to the following observations:

x	1	2	3	4	5	6
y	3.13	3.76	6.94	12.62	20.86	31.53

Answer: $\mathbf{y = 4.982 - 3.1199x + 1.2579x^2}$

B **3** Fit a parabola $y = a + bx + cx^2$ to the following data:

x	1	2	3	5	6
y	1.1	5.8	17.5	55.9	86.7

Answer: $\mathbf{y = 2.7227 - 4.5528x + 3.0771x^2}$

B **4** Fit a second-degree parabola $y = a + bx + cx^2$ to the following data:

x	1.0	1.5	2.0	2.5	3.0	3.5	4.0
y	1.1	1.3	1.6	2.0	2.7	3.4	4.1

Answer: $\mathbf{y = 1.0357 - 0.1929\,x + 0.2429\,x^2}$

C **5** For 10 randomly selected observations, the following data were recorded.

Observation Number	1	2	3	4	5	6	7	8	9	10
Overtime Hours (x)	1	1	2	2	3	3	4	5	6	7
Additional units (y)	2	7	7	10	8	12	10	14	11	14

Determine the coefficient of regression using the non-linear $y = a + b_1x + b_2x^2$.

Answer: $\mathbf{y = 1.8022 + 3.4823x - 0.2690x^2}$

<table>
<tr>
<td>C</td>
<td>6</td>
<td>The following are the data on the drying time of a certain varnish and the amount of an additive that is intended to reduce the drying time?
<table>
<tr><td>Amount of varnish additive(grams)"x"</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td></tr>
<tr><td>Drying time(hr.) "y"</td><td>12</td><td>10.5</td><td>10</td><td>8</td><td>7</td><td>8</td><td>7.5</td><td>8.5</td><td>9</td></tr>
</table>
Fit a second-degree polynomial by the method of least square.
Use the result to predict the drying time of the varnish when 6.5 gm of the additive is being used.

Answer: $y = 12.1848 - 1.8465x + 0.1829x^2$; $y(6.5) = 7.9101$</td>
</tr>
<tr>
<td>B</td>
<td>7</td>
<td>Fit a second – degree parabola $y = ax^2 + bx + c$ to the following data:
<table>
<tr><td>x</td><td>−1</td><td>0</td><td>1</td><td>2</td><td>3</td></tr>
<tr><td>y</td><td>5</td><td>6</td><td>21</td><td>50</td><td>93</td></tr>
</table>
Answer: $y = 7x^2 + 8x + 6$</td>
</tr>
</table>

Unit – 4.2 ⇝ Hypothesis Testing – I

Hypothesis Testing for Large Sample – I

Method – 5 ⇝ Test for Single Proportion

Example of Method-5: Test for Single Proportion

A	1	A coin was tossed 400 times and the head turned up 216 times. Test the hypothesis that the coin is unbiased. ($\lvert Z_{0.05} \rvert = 1.96$) **Answer: The coin is unbiased.**
B	2	In a hospital 480 female and 520 male babies were born in a week. Do these figures confirm the hypothesis that males and females were born in equal numbers? ($\lvert Z_{0.05} \rvert = 1.96$) **Answer: Males and females were born in equal proportions.**
A	3	In a study designed to investigate whether certain detonators used with explosives in a coal mining meet the requirement that at least 90% will ignite the explosives when charged. It is found that 174 of 200 detonators function properly. Test the null hypothesis $P = 0.9$ against the alternative hypothesis $P < 0.9$ at the 0.05 level of significance. ($Z_{0.05} = -1.645$) **Answer: Null hypothesis is accepted.**
B	4	A salesman in a departmental store claim that at most 60% of the shoppers entering the store leave without making a purchase. A random sample of 50 shoppers showed that 35 of them left without making a purchase. Are these sample results consistent with the claim of the salesman? Use a level of significance of 0.05. ($Z_{0.05} = 1.645$) **Answer: Yes, these sample results are consistent with the claim of salesman.**

A	5	In a random sample of 125 cold drinkers, 68 said they prefer 'Thumsup to Pepsi'. Test the null hypothesis $P = 0.5$ against the alternative hypothesis $P > 0.5$. ($Z_{0.05} = 1.645$) **Answer: Null hypothesis is accepted.**
C	6	A social worker believes that fewer than 25% of the couples in a certain area have ever used any form of birth control. A random sample of 120 couples was conducted. 20 of them said they have used. Test the belief of the social worker at 0.05 level. ($Z_{0.05} = 1.645$) **Answer: Social worker's belief is true.**

Method – 6 ⇝ Test for Difference of Proportions

Example of Method-6: Test for Difference of Proportions

A	1	100 articles from a factory are examined and 10 are found to be defective. Out of 500 similar articles from a second factory 15 are found to be defective. Test the significance between the difference of two proportions at 5% level. ($\lvert Z_{0.05} \rvert = 1.96$) **Answer: There is a significant difference between the two proportions**
A	2	In a random sample of 1000 persons from town A, 400 are found to be consumers of wheat. In a sample of 800 from town B, 400 are found to be consumers of wheat. Do these data reveal a significant difference between town A and town B, so far as the proportion of wheat consumers is concerned? ($\lvert Z_{0.05} \rvert = 1.96$) **Answer: There is significant difference between town A and town B as the proportion of wheat consumers is concerned.**
A	3	A manufacturer of electronic equipment subjects' samples of two completing brands of transistors to an accelerated performance test. If 45 of 180 transistors of the first kind and 34 of 120 transistors of second kind fails the test, what can be conclude at the level of significance $\alpha = 0.05$ about the difference between the corresponding sample proportion? ($\lvert Z_{0.05} \rvert = 1.96$) **Answer: The difference between the proportion is not significant.**
C	4	500 Articles from a factory are examined and found to be 2% defective. 800 similar articles from a second factory are found to have only 1.5% defective. Can it reasonably have concluded that the product of first factory is inferior than those of second? ($Z_{0.05} = 1.645$) **Answer: Products do not differ in quality.**

C	5	Before an increase in excise duty on tea, 800 people out of a sample of 1000 persons were found to be tea drinkers. After an increase in the duty, 800 persons were known to be tea drinkers in a sample of 1200 persons. Do you think that there is a significant decrease in the consumption of tea after the increase in the excise duty? ($Z_{0.01} = 2.33$) **Answer: There is significant decrease in consumption of tea.**
B	6	A cigarette manufacturing firm claims that its brand A of the cigarettes outsells its brand B by 8%. If it is found that 42 out of a sample of 200 smokers prefers brand A and 18 out of another random sample of 100 smokers prefer brand B, test whether the 8% difference is a valid claim. ($Z_{0.05} = 1.645$) **Answer: The claim of manufacturer is valid.**

*** * * * * End of the Unit * * * * ***