UNIT 7 RANK CORRELATION

Structure

7.1 INTRODUCTION

In second unit of this block, we have discussed the correlation with its properties and also the calculation of correlation coefficient. In correlation coefficient or product moment correlation coefficient, it is assumed that both characteristics are measurable. Sometimes characteristics are not measurable but ranks may be given to individuals according to their qualities. In such situations rank correlation is used to know the association between two characteristics. In this unit, we will discuss the rank correlation and calculation of rank correlation coefficient with its merits and demerits. We will also study the method of concurrent deviation.

In Section 7.2, you will know the concept of rank correlation while Section 7.3 gives the derivation of Spearman's rank correlation coefficient formula. Merits and demerits of the rank correlation coefficient are discussed in Sub-section 7.3.1. There might be a situation when two items get same rank. This situation is called tied or repeated rank which is described in Section 7.4. You will learn the method of concurrent deviation in Section 7.5.

Objectives

After reading this unit, you would be able to

- explain the concept of rank correlation;
- derive the Spearman's rank correlation coefficient formula;
- describe the merits and demerits of rank correlation coefficient;
- calculate the rank correlation coefficient in case of tied or repeated ranks; and
- describe the method of concurrent deviation.

7.2 CONCEPT OF RANK CORRELATION

For the calculation of product moment correlation coefficient characters must be measurable. In many practical situations, characters are not measurable. They are qualitative characteristics and individuals or items can be ranked in order of their merits. This type of situation occurs when we deal with the qualitative study such as honesty, beauty, voice, etc. For example, contestants of a singing competition may be ranked by judge according to their performance. In another example, students may be ranked in different subjects according to their performance in tests.

Arrangement of individuals or items in order of merit or proficiency in the possession of a certain characteristic is called ranking and the number indicating the position of individuals or items is known as rank.

If ranks of individuals or items are available for two characteristics then correlation between ranks of these two characteristics is known as rank correlation.

With the help of rank correlation, we find the association between two qualitative characteristics. As we know that the Karl Pearson's correlation coefficient gives the intensity of linear relationship between two variables and Spearman's rank correlation coefficient gives the concentration of association between two qualitative characteristics. In fact Spearman's rank correlation coefficient measures the strength of association between two ranked variables. Derivation of the Spearman's rank correlation coefficient formula is discussed in the following section.

7.3 DERIVATION OF RANK CORRELATION COEFFICIENT FORMULA

Suppose we have a group of n individuals and let $x_1, x_2, ..., x_n$ and $y_1, y_2, ..., y_n$ be the ranks of n individuals in characteristics A and B respectively.

It is assumed that no two or more individuals have the same rank in either characteristics A or B. Suppose both characteristics X and Y are taking rank values 1, 2, 3,…, n. Then sum of ranks of characteristics A is

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + ... + x_n$$

$$= 1 + 2 + ... + n \quad \text{(since X is taking values 1, 2,…,n)}$$

$$\sum_{i=1}^{n} x_i = \frac{n(n+1)}{2} \qquad \text{…(1)}$$

From, the formula of sum of n natural numbers.

Mean of variable X is

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

$$\bar{x} = \frac{(n+1)}{2}$$

Since both variables are taking same values 1, 2, …, n then

$$\bar{x} = \bar{y} = \frac{(n+1)}{2}$$

If variance of X is denoted by $\sigma_x^{\ 2}$ then

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

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

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

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

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

$$\sigma_x^{\ 2} = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2 = \frac{1}{n}(x_1^2 + x_2^2 + ... + x_n^2) - \bar{x}^2 \qquad \ldots (2)$$

Substituting the value of $\bar{x}$ in equation (2), we have

$$\sigma_x^{\ 2} = \frac{1}{n}(1^2 + 2^2 + ... + n^2) - \left\{\frac{n+1}{2}\right\}^2 \text{ (Since X is taking values 1,2,…,n)}$$

$$\sigma_x^{\ 2} = \frac{1}{n}\left\{\frac{n(n+1)(2n+1)}{6}\right\} - \left\{\frac{n+1}{2}\right\}^2$$

(From the formula of sum of squares of n natural numbers)

$$\sigma_x^{\ 2} = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}$$

$$\sigma_x^{\ 2} = (n+1)\left\{\frac{2n+1}{6} - \frac{(n+1)}{4}\right\}$$

$$\sigma_x^{\ 2} = (n+1)\left\{\frac{2(2n+1) - 3(n+1)}{12}\right\}$$

$$\sigma_x^{\ 2} = (n+1)\left\{\frac{4n+2-3n-3}{12}\right\}$$

$$\sigma_x^{\ 2} = (n+1)\left\{\frac{n-1}{12}\right\}$$

$$\sigma_x^{\ 2} = \frac{n^2-1}{12} \qquad \text{(from the formula } (a-b)(a+b) = a^2 - b^2\text{)}$$

Since both variables X and Y are taking same values, they will have same variance, thus

$$\sigma_y^2 = \sigma_x^2 = \frac{n^2-1}{12}$$

Let d_i be the difference of the ranks of i^{th} individual in two characteristics, then

$$d_i = x_i - y_i$$

$$d_i = x_i - y_i - \bar{x} + \bar{y} \qquad \text{Since } \bar{x} = \bar{y}$$

$$d_i = (x_i - \bar{x}) - (y_i - \bar{y})$$

Squaring and summing d_i^2 over i = 1 to n, we have

$$\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} \{(x_i - \bar{x}) - (y_i - \bar{y})\}^2$$

$$\Rightarrow \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} \{(x_i - \bar{x})^2 + (y_i - \bar{y})^2 - 2(x_i - \bar{x})(y_i - \bar{y})\}$$

$$\Rightarrow \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + \sum_{i=1}^{n} (y_i - \bar{y})^2 - 2\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \quad \ldots (3)$$

Dividing equation (3) by n, we have

$$\frac{1}{n}\sum_{i=1}^{n} d_i^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 + \frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y})^2 - 2\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} d_i^2 = \sigma_x^2 + \sigma_y^2 - 2\text{Cov}(x, y) \quad \ldots (4)$$

We know that, $r = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y}$, which implies that $\text{Cov}(x, y) = r\sigma_x\sigma_y$.

Substituting $\text{Cov}(x, y) = r\sigma_x\sigma_y$ in equation (4), we have

$$\frac{1}{n}\sum_{i=1}^{n} d_i^2 = \sigma_x^2 + \sigma_y^2 - 2r\sigma_x\sigma_y$$

Since, $\sigma_x^2 = \sigma_y^2$, then

$$\frac{1}{n}\sum_{i=1}^{n} d_i^2 = \sigma_x^2 + \sigma_x^2 - 2r\sigma_x\sigma_x$$

$$\frac{1}{n}\sum_{i=1}^{n} d_i^2 = 2\sigma_x^2 - 2r\sigma_x^2$$

$$\frac{1}{n}\sum_{i=1}^{n} d_i^2 = 2\sigma_x^2(1 - r)$$

$$\frac{1}{2n\sigma_x^2}\sum_{i=1}^{n} d_i^2 = (1 - r)$$

$$r = 1 - \frac{\sum_{i=1}^{n} d_i^2}{2n\sigma_x^2}$$

$$r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \qquad \left(\text{Since } \sigma_x^{\ 2} = \frac{n^2-1}{12}\right)$$

We denote rank correlation coefficient by r_s, and hence

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \qquad \ldots (5)$$

This formula was given by Spearman and hence it is known as Spearman's rank correlation coefficient formula.

Let us discuss some problems on rank correlation coefficient.

Example 1: Suppose we have ranks of 8 students of B.Sc. in Statistics and Mathematics. On the basis of rank we would like to know that to what extent the knowledge of the student in Statistics and Mathematics is related.

Rank in Statistics	1	2	3	4	5	6	7	8
Rank in Mathematics	2	4	1	5	3	8	7	6

Solution: Spearman's rank correlation coefficient formula is

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

Let us denote the rank of students in Statistics by R_x and rank in Mathematics by R_y. For the calculation of rank correlation coefficient we have to find $\sum_{i=1}^{n} d_i^2$ which is obtained through the following table:

Rank in Statistics (R_x)	**Rank in Mathematics** (R_y)	**Difference of Ranks** $(d_i = R_x - R_y)$	d_i^2
1	2	−1	1
2	4	−2	4
3	1	2	4
4	5	−1	1
5	3	2	4
6	8	−2	4
7	7	0	0
8	6	2	4
			$\sum d_i^2 = 22$

Here, n = number of paired observations = 8

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6 \times 22}{8 \times 63} = 1 - \frac{132}{504} = \frac{372}{504} = 0.74$$

Thus there is a positive association between ranks of Statistics and Mathematics.

Example 2: Suppose we have ranks of 5 students in three subjects Computer, Physics and Statistics and we want to test which two subjects have the same trend.

Rank in Computer	2	4	5	1	3
Rank in Physics	5	1	2	3	4
Rank in Statistics	2	3	5	4	1

Solution: In this problem, we want to see which two subjects have same trend i.e. which two subjects have the positive rank correlation coefficient.

Here we have to calculate three rank correlation coefficients

r_{12s} = Rank correlation coefficient between the ranks of Computer and Physics

r_{23s} = Rank correlation coefficient between the ranks of Physics and Statistics

r_{13s} = Rank correlation coefficient between the ranks of Computer and Statistics

Let R_1, R_2 and R_3 be the ranks of students in Computer, Physics and Statistics respectively.

Rank in Computer (R_1)	Rank in Physics (R_2)	Rank in Statistics (R_3)	$d_{12}=R_1-R_2$	d_{12}^2	$d_{23}=R_2-R_3$	d_{23}^2	$d_{13}=R_1-R_3$	d_{13}^2
2	5	2	−3	9	3	9	0	0
4	1	3	3	9	−2	4	1	1
5	2	5	3	9	−3	9	0	0
1	3	4	−2	4	−1	1	−3	9
3	4	1	−1	1	−3	9	2	4
Total				32		32		14

Thus,

$$\sum d_{12}^2 = 32,\ \sum d_{23}^2 = 32 \text{ and } \sum d_{13}^2 = 14.$$

Now

$$r_{12s} = 1 - \frac{6\sum d_{12}^2}{n(n^2-1)} = 1 - \frac{6 \times 32}{5 \times 24} = 1 - \frac{8}{5} = -\frac{3}{5} = -0.6$$

$$r_{23s} = 1 - \frac{6\sum d_{23}^2}{n(n^2-1)} = 1 - \frac{6 \times 32}{5 \times 24} = 1 - \frac{8}{5} = -\frac{3}{5} = -0.6$$

$$r_{13s} = 1 - \frac{6\sum d_{13}^2}{n(n^2-1)} = 1 - \frac{6 \times 14}{5 \times 24} = 1 - \frac{7}{10} = \frac{3}{10} = 0.3$$

r_{12s} is negative which indicates that Computer and Physics have opposite trend. Similarly, negative rank correlation coefficient r_{23s} shows the opposite trend in Physics and Statistics. $r_{13s} = 0.3$ indicates that Computer and Statistics have same trend.

Sometimes we do not have rank but actual values of variables are available. If we are interested in rank correlation coefficient, we find ranks from the given values. Considering this case we are taking a problem and try to solve it.

Example 3: Calculate rank correlation coefficient from the following data:

x	78	89	97	69	59	79	68
y	125	137	156	112	107	136	124

Solution: We have some calculation in the following table:

x	**y**	**Rank of x (R_x)**	**Rank of y (R_y)**	**$d = R_x - R_y$**	**d^2**
78	125	4	4	0	0
89	137	2	2	0	0
97	156	1	1	0	0
69	112	5	6	-1	1
59	107	7	7	0	0
79	136	3	3	0	0
68	124	6	5	1	1
					$\sum_{i=1}^{n} d_i^2 = 2$

Spearman's Rank correlation formula is

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

$$r_s = 1 - \frac{6 \times 2}{7(49-1)} = 1 - \frac{12}{7 \times 48}$$

$$= 1 - \frac{1}{28} = \frac{27}{28} = 0.96$$

Now, let us solve a little exercise.

E1) Calculate Spearman's rank correlation coefficient from the following data:

x	20	38	30	40	50	55
y	17	45	30	35	40	25

7.3.1 Merits and Demerits of Rank Correlation Coefficient

Merits of Rank Correlation Coefficient

1. Spearman's rank correlation coefficient can be interpreted in the same way as the Karl Pearson's correlation coefficient;
2. It is easy to understand and easy to calculate;

3. If we want to see the association between qualitative characteristics, rank correlation coefficient is the only formula;
4. Rank correlation coefficient is the non-parametric version of the Karl Pearson's product moment correlation coefficient; and
5. It does not require the assumption of the normality of the population from which the sample observations are taken.

Demerits of Rank Correlation Coefficient

1. Product moment correlation coefficient can be calculated for bivariate frequency distribution but rank correlation coefficient cannot be calculated; and
2. If n >30, this formula is time consuming.

7.4 TIED OR REPEATED RANKS

In Section 7.3, it was assumed that two or more individuals or units do not have same rank. But there might be a situation when two or more individuals have same rank in one or both characteristics, then this situation is said to be tied.

If two or more individuals have same value, in this case common ranks are assigned to the repeated items. This common rank is the average of ranks they would have received if there were no repetition. For example we have a series 50, 70, 80, 80, 85, 90 then 1st rank is assigned to 90 because it is the biggest value then 2nd to 85, now there is a repetition of 80 twice. Since both values are same so the same rank will be assigned which would be average of the ranks that we would have assigned if there were no repetition. Thus, both 80 will receive the average of 3 and 4 i.e. (Average of 3 & 4 i.e. (3 + 4) / 2= 3.5) 3.5 then 5th rank is given to 70 and 6th rank to 50. Thus, the series and ranks of items are

Series	50	70	80	80	85	90
Ranks	6	5	3.5	3.5	2	1

In the above example 80 was repeated twice. It may also happen that two or more values are repeated twice or more than that.

For example, in the following series there is a repetition of 80 and 110. You observe the values, assign ranks and check with following.

Series	50	70	80	90	80	120	110	110	110	100
Ranks	10	9	7.5	6	7.5	1	3	3	3	5

When there is a repetition of ranks, a correction factor $\frac{m(m^2-1)}{12}$ is added to $\sum d^2$ in the Spearman's rank correlation coefficient formula, where m is the number of times a rank is repeated. It is very important to know that this correction factor is added for every repetition of rank in both characters.

In the first example correction factor is added once which is 2(4 - 1)/12 = 0.5, while in the second example correction factors are 2(4 - 1)/12 = 0.5 and 3 (9 - 1)/12 = 2 which are aided to $\sum d^2$.

Thus, in case of tied or repeated rank Spearman's rank correlation coefficient formula is

$$r_s = 1 - \frac{6\left\{\sum d^2 + \frac{m(m^2-1)}{12} + ...\right\}}{n(n^2-1)}$$

Example 4: Calculate rank correlation coefficient from the following data:

Expenditure on advertisement	10	15	14	25	14	14	20	22
Profit	6	25	12	18	25	40	10	7

Solution: Let us denote the expenditure on advertisement by x and profit by y

x	**Rank of x (R_x)**	**y**	**Rank of y (R_y)**	**d = R_x - R_y**	**d^2**
10	8	6	8	0	0
15	4	25	2.5	1.5	2.25
14	6	12	5	1	1
25	1	18	4	−3	9
14	6	25	2.5	3.5	12.25
14	6	40	1	5	25
20	3	10	6	−3	9
22	2	7	7	−5	25
					$\sum d^2 = 83.50$

$$r_s = 1 - \frac{6\left\{\sum d^2 + \frac{m(m^2-1)}{12} + ...\right\}}{n(n^2-1)}$$

Here rank 6 is repeated three times in rank of x and rank 2.5 is repeated twice in rank of y, so the correction factor is

$$\frac{3(3^2-1)}{12} + \frac{2(2^2-1)}{12}$$

Hence rank correlation coefficient is

$$r_s = 1 - \frac{6\left\{83.50 + \frac{3(3^2-1)}{12} + \frac{2(2^2-1)}{12}\right\}}{8(64-1)}$$

$$r_s = 1 - \frac{6\left\{83.50 + \frac{3 \times 8}{12} + \frac{2 \times 3}{12}\right\}}{8 \text{X } 63}$$

$$r_s = 1 - \frac{6(83.50 + 2.50)}{504}$$

$$r_s = 1 - \frac{516}{504}$$

$$r_s = 1 - 1.024 = -0.024$$

There is a negative association between expenditure on advertisement and profit.

Now, let us solve the following exercises.

E2) Calculate rank correlation coefficient from the following data:

x	10	20	30	30	40	45	50
y	15	20	25	30	40	40	40

E3) Calculate rank correlation coefficient from the following data:

x	70	70	80	80	80	90	100
y	90	90	90	80	70	60	50

7.5 CONCURRENT DEVIATION

Sometimes we are not interested in the actual amount of correlation coefficient but we want to know the direction of change i.e. whether correlation is positive or negative, coefficient of concurrent deviation serves our purpose. In this method correlation is calculated between the direction of deviations and not their magnitudes. Coefficient of concurrent deviation is denoted by r_c and given by

$$r_c = \pm\sqrt{\pm\frac{(2c - k)}{k}} \qquad \ldots (6)$$

where, c is the number of concurrent deviation or the number of + sign in the product of two deviations, $k = n - 1$ i.e. total number of paired observation minus one. This is also called coefficient of correlation by concurrent deviation method. Steps for the calculation of concurrent deviation (see the Example 5 simultaneously) are:

1. The first value of series x is taken as a base and it is compared with next value i.e. second value of series x. If second value is greater than first value, '+' sign is assigned in the Column titled D_x. If second value is less than the first value then '-' sign is assigned in the column D_x.
2. If first and second values are equal then '=' sign is assigned.
3. Now second value is taken as base and it is compared with the third value of the series. If third value is less than second '-' is assigned against the

third value. If the third value is greater than the second value '+' is assigned. If second and third values or equal than '=' sign is assigned.

4. This procedure is repeated upto the last value of the series.
5. Similarly, we obtain column D_y for series y.
6. We multiply the column D_x and D_y and obtain column $D_x D_y$. Multiplication of same sign results '+'sign and that of different sign is '-' sign.
7. Finally number of '+' sign are counted in the column $D_x D_y$, it is called c and we get coefficient concurrent deviation by the formula (6).
8. In the formula, inside and outside the square root, sign '+' and '-' depends on the value of $(2c - k)$. If this value is positive than '+' sign is taken at both places if $(2c - k)$ is negative '-'sign is considered at both the places.

Let us discuss the following problem.

Example 5: We have data of income and expenditure of 11 workers of an organization in the following table:

Income	65	40	35	75	63	79	35	20	80	60	50
Expenditure	60	55	50	66	30	71	40	35	80	75	80

Find whether correlation is positive or negative by coefficient of concurrent deviation.

Solution: Coefficient of concurrent deviation is given by

$$r_c = \pm\sqrt{\pm\frac{(2c - k)}{k}}$$

Let us denote the income by x and expenditure by y and we calculate c by the following table:

x	Change of direction sign for x (D_x)	y	Change of direction sign for y (D_y)	D_xD_y
65		60		
40	−	55	−	+
35	−	50	−	+
75	+	66	+	+
63	−	30	−	+
79	+	71	+	+
35	−	40	−	+
20	−	35	−	+
80	+	81	+	+
60	−	75	−	+
50	−	80	+	−
				c = 9

Here, c = 9 and k = n −1 = 10 then we have

$$r = \pm\sqrt{\pm\frac{2\times 9-10}{10}} \qquad = +\sqrt{+\frac{8}{10}}$$

(Both signs are + because 2c − k is positive)

$$= \sqrt{0.8} = 0.89$$

Thus correlation is positive.

Now, let us solve the following exercises.

E 4) Find the coefficient of correlation between supply and price by concurrent deviation method for the following data:

Year	1970	1971	1972	1973	1974	1975
Supply	114	127	128	121	120	124
Price	108	104	105	106	100	99

E5) Calculate coefficient of concurrent deviation for the following data:

x	368	384	385	360	347	384
y	122	121	124	125	122	126

7.6 SUMMARY

In this unit, we have discussed:

1. The rank correlation which is used to see the association between two qualitative characteristics;
2. Derivation of the Spearman's rank correlation coefficient formula;
3. Calculation of rank correlation coefficient in different situations- (i) when values of variables are given, (ii) when ranks of individuals in different characteristics are given and (iii) when repeated ranks are given;
4. Properties of rank correlation coefficient; and
5. Concurrent deviation which provides the direction of correlation.

7.7 SOLUTIONS /ANSWERS

E1) We have some calculations in the following table:

x	**y**	**Rank of x (R_x)**	**Rank of y (R_y)**	**d = $R_x - R_y$**	**d^2**
20	17	6	6	0	0
38	45	4	1	3	9
30	30	5	4	1	1
40	35	3	3	0	0
50	40	2	2	0	0
55	25	1	5	−4	16
					$\sum_{i=1}^{n} d_i^2 = 26$

$$r_s = 1 - \frac{6 \times 26}{6(36-1)}$$

$$= 1 - \frac{26}{35} = \frac{9}{35} = 0.26$$

E2) We have some calculations in the following table:

x	Rank of x (R_x)	y	Rank of y (R_y)	$d = R_x - R_y$	d^2
10	7	15	7	0	0
20	6	20	6	0	0
30	4.5	25	5	−0.5	0.25
30	4.5	30	4	0.5	0.25
40	3	40	2	1	1
45	2	40	2	0	0
50	1	40	2	−1	1
					$\sum d^2 = 2.5$

$$r_s = 1 - \frac{6\left\{\sum d^2 + \frac{m(m^2-1)}{12} + ...\right\}}{n(n^2-1)}$$

Here, rank 4.5 is repeated twice in rank of x and rank 2 is repeated thrice in rank of y so the correction factor is

$$\frac{2(2^2-1)}{12} + \frac{3(3^2-1)}{12}$$

and therefore, rank correlation coefficient is

$$r_s = 1 - \frac{6\left\{2.5 + \frac{2(2^2-1)}{12} + \frac{3(3^2-1)}{12}\right\}}{7(49-1)}$$

$$r_s = 1 - \frac{6\left\{2.5 + \frac{2 \times 3}{12} + \frac{3 \times 8}{12}\right\}}{7 \times 48}$$

$$r_s = 1 - \frac{6(2.5 + 2.5)}{336}$$

$$r_s = 1 - \frac{30}{336} = \frac{306}{336}$$

$$r_s = 0.91$$

E3) We have some calculations in the following table:

x	Rank of x (R_x)	y	Rank of y (R_y)	$d = R_x - R_y$	d^2
70	6.5	90	2	4.5	20.25
70	6.5	90	2	4.5	20.25
80	4	90	2	2	4
80	4	80	4	0	0
80	4	70	5	−1	1
90	2	60	6	−4	16
100	1	50	7	−6	36
					$\sum d^2 = 97.5$

Rank correlation coefficient is

$$r_s = 1 - \frac{6\left\{\sum d^2 + \frac{m(m^2-1)}{12} + \ldots\right\}}{n(n^2-1)}$$

Here, rank 4 and 6.5 is repeated thrice and twice respectively in rank of x and rank 2 is repeated thrice in rank of y, so the correction factor is

$$\frac{3(3^2-1)}{12} + \frac{2(2^2-1)}{12} + \frac{3(3^2-1)}{12}$$

and therefore, rank correlation coefficient is

$$r_s = 1 - \frac{6\left\{97.5 + \frac{3(3^2-1)}{12} + \frac{2(2^2-1)}{12} + \frac{3(3^2-1)}{12}\right\}}{7(49-1)}$$

$$r_s = 1 - \frac{6\{97.5 + 4.5\}}{7 \times 48}$$

$$r_s = 1 - \frac{6(102)}{336}$$

$$r_s = 1 - \frac{102}{56} = -0.82$$

E4) Coefficient of concurrent deviation is given

$$r_c = \pm\sqrt{\pm\frac{(2c-k)}{k}}$$

Let us denote the supply by x and price by y and we calculate c by the following table:

x	Change of Direction sign for x (D_x)	y	Change of Direction sign for y (D_y)	D_xD_y
114		108		
127	+	104	−	−
128	+	105	+	+
121	−	106	+	−
120	−	100	−	+
124	+	99	−	−
				c = 2

Now c = 2 and k = n - 1 = 5

$$r = \pm\sqrt{\pm\frac{2\times2-5}{5}} \quad = -\sqrt{\frac{1}{5}}$$

(Both signs are '-' because 2c − k is negative)

= − 0.45

Thus, correlation is negative.

E5) Coefficient of concurrent deviation is given

$$r_c = \pm\sqrt{\pm\frac{(2c-k)}{k}}$$

Let us denote the supply by x and price by y and we calculate c by the following table:

x	Change of Direction sign for x (D_x)	y	Change of Direction sign for y (D_y)	D_xD_y
368		122		
384	+	121	−	−
385	+	124	+	+
360	−	125	+	−
347	−	122	−	+
384	+	126	+	+
				c = 3

Now c = 3 and k = n −1= 5

$$r = \pm\sqrt{\pm\frac{2\times3-5}{5}} \quad = \sqrt{\frac{1}{5}}$$

(Both signs are '+' because 2c−k is positive)

r = 0.45

Thus, correlation is positive.