UNIT 15 ASSOCIATION OF ATTRIBUTES

Structure

15.1 INTRODUCTION

You have seen in Units 13 and 14 that statistical methods deal with the quantitative measurements and the quantitative data can be obtained in two ways:

1. Where the actual magnitude of the variables can be measured for individuals or items for example, income of the people.
2. Where we can count the number of people possessing a particular attribute or the number of people not possessing this particular attribute. For example, number of persons having dark brown eyes. The quantitative character, in this case arises only by counting the number of persons with the presence or absence of certain quality or attribute.

You have also seen that the statistical methodologies for studying the nature of the relationship between two variables for the two aforesaid cases are different. For case (1), relationship between two variables can be measured by correlation coefficient, which not only gives the magnitude of the relationship but also the direction of the relationship i.e. the two variables are positively or negatively related. However, for the second case, coefficient correlation cannot be used, as the data is not numerically expressed, only we know the number possessing particular attribute or not. We have seen in Unit14, the methods used to study the relationship are different. They are covered under theory of attributes.

We have also seen that if an attribute has only two classes, it is said to be dichotomous and if it has many classes, it is called manifold classification. Hence, the need is often felt as to whether there is some association between attributes or not. For this, some measures are developed specifically known as Measures of Association of Attributes. In this unit, we will discuss the measures of association which tell us if there is any association between attributes or not and also whether association is positive or negative.

In this unit, we will focus on association of attributes i.e. the relationship between attributes and how to measure such relationship.

In Section 15.2 we shall discuss association of attributes while in Section 15.3 we shall deal with some of the measures of association commonly known as coefficient of association.

Objectives

After studying this unit, you would be able to

- define the association of attributes;
- differentiate between coefficient of correlation and coefficient of association;
- assess what kind of associations among attributes are likely to occur; and
- distinguish between different methods of measures of association.

15.2 ASSOCIATION OF ATTRIBUTES

Yule Statisti associa appear numbe be ex indeper

The meaning of association in statistical language is quiet different from the common meaning of association. Commonly, if two attributes A and B appear together number of times then they can be said to be as associated. But according to Yule and Kendall, "In Statistics A and B are associated only if they appear together in a greater number of cases than is to be expected, if they are independent."

Correlation coefficient is a measure of degree or extent of linear relationship between two variables, whereas the coefficient of association indicates association between two attributes and also whether the association is positive or negative.

Methods used to measure the association of attributes refer to those techniques, which are used to measure the relationship between two such phenomena, whose size cannot be measured and where we can only find the presence or absence of an attribute.

In the case of correlation analysis, we study the relationship between two variables, which we can measure quantitatively. Similarly, in the case of association we study the relationship between two attributes, which are not quantitatively measurable. For example, level of education and crime. In association no variables are involved. As it has been stated earlier an attribute divides the universe into two classes, one possessing the attribute and another not possessing the attribute whereas the variable can divide the universe into any number of classes. Correlation coefficient is a measure of degree or extent of linear relationship between two variables, whereas the coefficient of association indicates association between two attributes and also whether the association is positive or negative. But with the help of coefficient of association we cannot find expected change in A for a given change in B and vice-versa, as possible by regression coefficient, which is derived from correlation coefficient.

15.2.1 Types of Association

Two attributes A and B are said to be associated if they are not independent but are related in some way or the other. There are three kinds of associations, which possibly occur between attributes.

1. Positive association
2. Negative association or disassociation
3. No association or independence.

In positive association, the presence of one attribute is accompanied by the presence of other attribute. For example, health and hygiene are positively associated.

or $\quad$ if $(AB) > \frac{(A)(B)}{N}$

Then attributes A and B are positively associated.

In negative association, the presence of one attribute say A ensures the absence of another attribute say B or vice versa. For example, vaccination and occurrence of disease for which vaccine is meant are negatively associated.

or $\quad$ if $(AB) < \frac{(A)(B)}{N}$

Then attributes A and B are negatively associated.

If two attributes are such that presence or absence of one attribute has nothing to do with the absence or presence of another, they are said to independent or not associated. For example, Honesty and Boldness

or $\quad$ if $(AB) = \frac{(A)(B)}{N}$

Then attributes A and B are independent.

Note:

1. Two attributes A and B are said to be completely associated if A cannot occur without B, though B may occur without A and vice-versa. In other words, A and B are completely associated if all A's are B's i.e. (AB) = (A) or all B's are A's i.e. (AB) = (B), according as whether either A's or B's are in a minority.
2. Complete disassociation means that no A's are B's i.e. (AB) = 0 or no α's are β's i.e. (αβ) = 0.

15.2.2 The Symbols $(AB)_0$ and δ

In this unit, following symbols will be used

δ is a Greek letter called Delta.

$$(AB)_0 = \frac{(A)(B)}{N}, \quad (\alpha\beta)_0 = \frac{(\alpha)(\beta)}{N}$$

$$(\alpha B)_0 = \frac{(\alpha)(B)}{N}, \quad (A\beta)_0 = \frac{(A)(\beta)}{N}$$

$$\delta = (AB) - (AB)_0 = (AB) - \frac{(A)(B)}{N}$$

If $\delta = 0$, then $(AB) = \frac{(A)(B)}{N}$

$\Rightarrow$ A and B are independent.

If $\delta > 0$ then attributes A and B are positively associated and if $\delta < 0$ then attributes A and B are negatively associated.

Remark: It is to be noted that if $\delta \neq 0$ and its value is very small then it is possible that this association (either positive or negative) is just by chance and not really significant of any real association between the attributes. This difference is significant or not should be tested by the test statistic (χ^2: Chi-square).

Example 1: Show whether A and B are independent, positively associated or negatively associated in each of the following cases:

(i) N = 1000; (A) = 450; (B) = 600; (AB) = 340

(ii) (A) = 480; (AB) = 290; (α) = 585; (αB) = 383

(iii) N = 1000; (A) = 500; (B) = 400; (AB) = 200

Solution: We have given

(i) $\frac{(A)(B)}{N} = \frac{450 \times 600}{1000} = 270 = (AB)_0$

Thus, $(AB) = 340 > \frac{(A)(B)}{N}$

Since $(AB) > (AB)_0$ hence they are positively associated.

(ii) $\because (B) = (AB) + (\alpha B) = 290 + 383 = 673$

$$N = (A) + (\alpha) = 480 + 585 = 1065$$

$$\therefore \frac{(A)(B)}{N} = \frac{480 \times 673}{1065} = 303.32 = (AB)_0$$

Thus, $(AB) = 290 < 303.32$

$$\because (AB) < (AB)_0$$

$\therefore$ A and B are negatively associated.

(iii) $\frac{(A)(B)}{N} = \frac{500 \times 400}{1000} = 200 = (AB)_0$

Thus, we find $(AB) = (AB)_0$

Hence, A and B are independent, i.e. $\delta = 0$

Example 2: The male population of certain state is 250 lakhs. The number of literate males is 26 lakhs and the total number of male criminals is 32 thousand. The number of literate male criminal is 3000. Do you find any association between literacy and criminality?

Solution: Let literate males be denoted by A so illiterate males would be denoted by α.

Let B represents male criminal so that males who are not criminal would be denoted by β.

Then in lakhs

$$(A) = 26; (B) = 0.32; (AB) = 0.03; N = 250$$

To study association between A and B let us compute

$$(AB)_0 = \frac{(A)(B)}{N} = \frac{26 \times 0.32}{250} = 0.0332$$

Since $(AB) = 0.03 < (AB)_0$. Hence literacy and criminality are negatively associated.

Example 3: 1660 candidates appeared for competitive examination 425 were successful, 252 had attended a coaching class and of these 150 came successful. Is there any association between success and utility of coaching class?

Solution: Let A denotes successful candidates and B denotes candidates attending coaching class

Given N = 1660; (A) = 425; (B) = 252; (AB) = 150

$$(AB)_0 = \frac{(A)(B)}{N} = \frac{425 \times 252}{1660} = 64.52$$

Since, $(AB) = 150 > (AB)_0$ therefore, there is a positive association between success and utility of coaching class.

Now, let us solve the following exercise:

E1) Find if A and B are independent, positively associated or negatively associated in each of the following cases:

(i) N = 100; (A) = 47; (B) = 62 and (AB) = 32

(ii) (A) = 495; (AB) = 292; (α) = 572 and ($\alpha\beta$) = 380

(iii) (AB) = 2560; (αB) = 7680;(Aβ) = 480 and ($\alpha\beta$) = 1440

E2) Out of total population of 1000 the number of vaccinated persons was 600. In all 200 had an attack of smallpox and out of these 30 were those who were vaccinated. Do you find any association between vaccination and freedom from attack?

E3) In an area with a total population of 7000 adults, 3400 are males and out of a total 600 graduates, 70 are females. Out of 120 graduate employees, 20 are females.

(i) Is there any association between sex and education?

(ii) Is their any association between appointment and sex?

15.3 METHODS OF MEASURES OF ASSOCIATION

The methods we have discussed so far can give you an idea whether two attributes are positively associated, negatively associated or independent. Sometimes, this is enough for taking decisions for practical purposes. But most of the times, it is not sufficient as we are always interested in the extent of association, so that we can measure the degree of association mathematically. In the present section, we shall discuss the possibility of obtaining coefficient of association, which can give some idea about the extent of association between two attributes. It would be easy for taking decision if the coefficient of association is such that its value is 0 when two attributes are independent; +1 when they are perfectly associated and –1 when they are perfectly dissociated. In between –1 to +1 lie different levels of association.

Many authors have developed many such coefficients of association, but we will be discussing the one given by Yule.

15.3.1 Yule's Coefficient of Association

Yule's coefficient of association is named after its inventor G. Udny Yule. For two attributes A and B, the coefficient of association is given as

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

Value of Q lies between –1 and +1.

1. If Q = 1, A and B has perfect positive association. It can be verified that under perfect positive association

   $$(AB) = (A) \Rightarrow (A\beta) = 0$$

   $$(AB) = (B) \Rightarrow (\alpha B) = 0$$

2. If Q = –1, A and B possess perfect negative association. This leads to following relationship:

   $$(AB) = 0 \text{ or } (\alpha\beta) = 0$$

3. If Q = 0, A and B are independent. Here, we have following relation

   $$(AB)(\alpha\beta) = (A\beta)(\alpha B)$$

4. Any value between –1 to +1 tells us the degree of relationship between two attributes A and B. Conventionally, if $Q > 0.5$ the association between two attributes is considered to be of high order and the value of Q less than 0.5 shows low degree of association between two attributes.

Q =
asso
Q =
asso
Q =
Q >
high
Q <
low

Remarks: It is to be noted that Q is independent of the relative preposition of A's or α's in the data. This property of Q is useful when the prepositions are arbitrary.

15.3.2 Coefficient of Colligation

This is another important coefficient of association given by Yule. It is defined as

$$\gamma = \frac{1 - \sqrt{\dfrac{(A\beta)(\alpha B)}{(AB)(\alpha\beta)}}}{1 + \sqrt{\dfrac{(A\beta)(\alpha B)}{(AB)(\alpha\beta)}}}$$

It can be shown that

$$Q = \frac{2\gamma}{1+\gamma^2}$$

The range of γ is from –1 to +1. It can be interpreted in the same manner as Q.

Following exercises will illustrate the calculation and interpretation of above two coefficients of association.

Example 4: Calculate Yule's coefficient of association for the following data:

(i) (A) = 600; (B) = 800 ;(AB) = 480; N = 1000

(ii) (A) = 600; (B) = 800; (AB) = 600; N = 1000

(iii) (A) = 600; (B) = 800; (AB) = 400; N = 1000

(iv) (A) = 600; (B) = 800; (AB) = 500; N = 1000

Solution: We have

(i) (AB) = 480

$(A\beta) = (A) - (AB) = 600 - 480 = 120$

$(\alpha B) = (B) - (AB) = 800 - 480 = 320$

$(\alpha\beta) = (\alpha) - (\alpha B) = 400 - 320 = 80$

$$\therefore Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$= \frac{(480 \times 80) - (120 \times 320)}{(480 \times 80) + (120 \times 320)}$$

$$= \frac{0}{38400 + 38400} = 0$$

Thus, two attributes are independent

(ii) Here

$(AB) = 600$

$(A\beta) = (A) - (AB) = 600 - 600 = 0$

$(\alpha B) = (B) - (AB) = 800 - 600 = 200$

$(\alpha\beta) = (\alpha) - (\alpha B) = 400 - 200 = 200$

$$\therefore Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$= \frac{(600 \times 200) - (0 \times 200)}{(600 \times 200) + (0 \times 200)}$$

$$= \frac{120000}{120000} = +1$$

Thus, there is a perfect positive association between attributes A and B.

(iii) In this case

$(AB) = 400$

$(A\beta) = (A) - (AB) = 600 - 400 = 200$

$(\alpha B) = (B) - (AB) = 800 - 400 = 400$

$(\alpha\beta) = (\alpha) - (\alpha B) = 400 - 400 = 0$

$$\therefore Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$= \frac{(400 \times 0) - (200 \times 400)}{(400 \times 0) + (200 \times 400)}$$

$$= \frac{-80000}{80000} = -1$$

(iv) Here, we have

$(AB) = 500$

$(A\beta) = (A) – (AB) = 600 – 500 = 100$

$(\alpha B) = (B) – (AB) = 800 – 500 = 300$

$(\alpha\beta) = (\alpha) – (\alpha B) = 400 – 300 = 100$

$$\therefore Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$= \frac{(500 \times 100) - (100 \times 300)}{(500 \times 100) + (100 \times 300)}$$

$$= \frac{20000}{80000} = +0.25$$

Thus, there is very nominal association between A and B.

Example 5: In a sample of 1000 children, 400 came from higher income group and rest from lower income group. The number of delinquent children in these groups was 50 and 200 respectively. Calculate the coefficient of association between delinquency and income groups.

Solution: Let A denotes the higher income group, then α would denote lower income group. Let B denotes delinquent children then β would denote non-delinquent children. To get the frequencies of second order we form following nine square table (or 2×2 table):

Attributes	**A**	**α**	**Total**
B	AB 50	αB 200	B 250
β	Aβ 350	$\alpha\beta$ 400	β 750
Total	A 400	α 600	N 1000

From the table

$(\alpha) = N – (A) = 1000 – 400 = 600$

$(B) = (AB) + (\alpha B) = 50 + 200 = 250$

$(A\beta) = (A) – (AB) = 400 – 50 = 350$

$(\alpha B) = (B) – (AB) = 250 – 50 = 200$

$(\alpha\beta) = (\alpha) – (\alpha B) = 600 – 200 = 400$

$$\therefore Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$= \frac{(50 \times 400) - (350 \times 200)}{(50 \times 400) + (350 \times 200)}$$

$$= \frac{-50000}{90000} = -0.55$$

Thus, there is a negative association between income and delinquency.

Example 6: Investigate if there is any association between extravagance in father and son from the following:

Extravagant sons with extravagant fathers (AB) = 450

Miser sons with extravagant fathers (αB) = 155

Extravagant sons with miser fathers (Aβ) = 175

Miser sons with miser fathers ($\alpha\beta$) = 1150

Solution: For association between extravagance in father and son we calculate coefficient of association as

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$= \frac{(450 \times 1150) - (175 \times 155)}{(450 \times 1150) + (175 \times 155)}$$

$$= \frac{490375}{544625} = 0.90$$

Thus, there is very high degree of positive association between extravagance in father and son. An extravagant father in general has an extravagant son.

Example 7: In an examination, at which 600 candidates appeared, boys out numbered girls by 16% of all candidates. Number of passed exceeded the number of failed candidates by 310. Boys failing in the examination numbered 88. Find the Yule's coefficient of association between male sex and success in examination.

Solution: Let boys be denoted by A so girls are denoted by α.

Let the success in examination be denoted by B so that failure in examination is denoted by β.

Data given are

(i) Boys outnumber girls by 16% of the total is

$$= \frac{16 \times 600}{100} = 96$$

$\therefore$ by condition we have (A) – (α) = 96

Also we have (A) + (α) = 600

$\therefore$ 2 (A) = 696

$\Rightarrow$ (A) = 348

which is the number of boys so the number of girls (α) would be

(A) – 96 = (α)

$\therefore$ 348 – 96 = (α)

$\Rightarrow$ (α) = 252

(ii) Number of passed candidates exceeded number that failed by 310

i.e. (B) – (β) = 310

also we have (B) + (β) = 600

$$\therefore\ 2\,(B) = 910$$

$$\Rightarrow\ (B) = 455$$

Thus, number of passed candidates is 455, so the number of failures would be

$$(\beta) = (B) - 310 = 455 - 310 = 145$$

(iii) Boys failing in the examination (Aβ) = 88

Other values can be obtained from the 2 × 2 table

Attributes	**A**	**α**	**Total**
B	AB 260	α B 195	B 455
β	Aβ 88	α β 57	β 145
Total	A 348	α 252	N 600

Yule's coefficient of association is

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$= \frac{(260 \times 57) - (88 \times 195)}{(260 \times 57) + (88 \times 195)}$$

$$= \frac{-2340}{31980} = -0.07$$

Thus, there is insignificant association between male sex and success.

Example 8: Given

$$(AB) = 35 \qquad (\alpha\beta) = 7$$

$$(A\beta) = 8 \qquad (\alpha B) = 6$$

calculate the coefficient of colligation.

Solution: Coefficient of colligation

$$\gamma = \frac{1 - \sqrt{\dfrac{(A\beta)(\alpha B)}{(AB)(\alpha\beta)}}}{1 + \sqrt{\dfrac{(A\beta)(\alpha B)}{(AB)(\alpha\beta)}}} = \frac{1 - \sqrt{\dfrac{8 \times 6}{35 \times 7}}}{1 + \sqrt{\dfrac{8 \times 6}{35 \times 7}}}$$

$$= \frac{1-\sqrt{0.196}}{1+\sqrt{0.196}} = \frac{1-0.44}{1+0.44} = 0.39$$

Thus, two attributes A and B are positively associated.

Now, let us solve the following exercise:

E4) The following table is reproduced from a memoir written by Karl Pearson

	Eye colour in son	
Eye colour in father	Not light	Light
Not light	230	148
Light	151	471

Discuss whether the colour of the son's eye is associated with that of father.

E5) Can vaccination be regarded as a preventive measure for small pox from the data given below:

'Of 1482 persons in a locality exposed to small pox 368 in all were attacked'

'Of 1482 persons, 343 had been vaccinated and of these only 35 were attacked'.

E6) From the following data prepare 2×2 table and using Yule's coefficient discuss whether there is any association between literacy a unemployment.

Illiterate Unemployed	250 persons
Literate Employed	25 persons
Illiterate Employed	180 persons
Total number of persons	500 persons

E7) From the data given below calculate Yule's coefficient of association between weight of the children and then economic condition, and interpret it.

	Poor Children	Rich Children
Below Normal Weight	85	20
Above Normal Weight	6	47

This brings us end of this unit. In the next unit we will take up the study of some more aspects of theory of attributes. But before that let us briefly recall what we have studied in this unit.

15.4 SUMMARY

In this unit, we have discussed:

1. Meaning of association in Statistics is different from the general meaning of association. In Statistics, attributes A and B are associated only if they appear together in greater number of cases than is to be expected if they are independent. In common language association means if A and B occur together a number of times then A and B are associated;
2. In association, we study the relationship between two attributes, which are not quantitatively measured;
3. Correlation coefficient measures the extent of relationship between two quantitative variables, whereas coefficient of association only suggests that the association is positive or negative;
5. If there exist no relationship of any kind between two attributes then they are said to be independent otherwise are said to be associated. Attributes A and B are said to be

   Positively associated if $(AB) > \frac{(A)(B)}{N}$

   Negatively associated if $(AB) < \frac{(A)(B)}{N}$

   Independent if $(AB) = \frac{(A)(B)}{N}$
6. Some times only the knowledge of the association (whether positive or negative) or independence between attributes is not sufficient. We are interested in finding the extent or degree of association between attributes, so that we can take decision more precisely and easily. In this regard, we have discussed Yule's coefficient of association in this unit. The value of Yule's coefficient of association lies between –1 to +1. If Q = +1, A and B are perfectly associated. In between –1 to +1, are lying different degrees of association;
7. Another important coefficient of association is coefficient of colligation. Q and γ are related by the following expression

   $$Q = \frac{2\gamma}{1+\gamma^2}; \text{ and}$$
8. γ also lies between –1 to +1 and have the interpretation as that of Q.

15.5 SOLUTIONS / ANSWERS

E1) $$(AB)_0 = \frac{(A)(B)}{N} = \frac{47 \times 62}{100} = 29.14$$

$\because (AB) = 32 > (AB)_0$

$\therefore$ A and B are positively related.

(ii) We have

$N = (A) + (\alpha) = 495 + 572 = 1067$

$(B) = (AB) + (\alpha B) = 292 + 380 = 672$

$\therefore \left(AB\right)_0 = \frac{(A)(B)}{N} = \frac{495 \times 672}{1067} = 31.75$

$\because (AB) = 292 < \left(AB\right)_0$

$\therefore$ A and B are negatively related.

(iii) (A) = (AB) + (Aβ) = 2560 + 480 = 3040

(B) = (AB) + (αB) = 2560 + 7680 = 10240

N = (AB) + (Aβ) + (αB) + ($\alpha\beta$)

= 2560 + 480 + 7680 + 1440

= 12160

$\therefore \left(AB\right)_0 = \frac{(A)(B)}{N} = \frac{3040 \times 10240}{12160} = 2560$

$\because (AB) = 2560 = \left(AB\right)_0$

$\therefore$ A and B are independent.

$\therefore$ i.e. $\delta = 0$

E2) Let A represents vaccinated and B freedom from attack.

The given data are

N = 1000; (A) = 600; (β) = 200; (Aβ) = 30

We have (AB) + (Aβ) = N

$\therefore$ (AB) = N – (Aβ) = 1000 – 30 = 970

Again, we have

(B) + (β) = N

$\therefore$ (B) = N – (β) = 1000 –200 = 800

Thus $\left(AB\right)_0 = \frac{(A)(B)}{N} = \frac{600 \times 800}{1000} = 480$

Since $(AB) = 970 > \left(AB\right)_0$

Hence, A and B are positively associated.

E3) (i) Let A represents males; α will be females

Let B represents graduates; β will be non graduates

The given data are

N = 7000, (A) = 3400, (B) = 600, (αB) = 70

(AB) = (αB) + (B) = 600 + 70 = 670

$\left(AB\right)_0 = \frac{(A)(B)}{N} = \frac{3400 \times 600}{7000} = 291.43$

Since $(AB) = 670 > \left(AB\right)_0$

Hence, A and B are positively associated.

(ii) Let A represents male graduates then α will be female graduates.

Let B represents employed then β will be unemployed.

Given (A) = 530; (β) = 70; (B) = 120; (αB) = 20

$$N = (\beta) + (B) = 70 + 120 = 190$$

$$(AB) + (\alpha B) = (B)$$

$$\Rightarrow \quad (AB) = 120 - 20 = 100$$

Now, $$(AB)_0 = \frac{(A)(B)}{N} = \frac{530 \times 120}{100} = 636$$

$$\because \quad (AB) = 100 < 636$$

$\therefore$ A and B are negatively related.

E4) Let A represents the light eye colour of father an B represents the light eyecolour of son. Then α represents not light eye colour of father and β represents not light eye colour of son. Then the given data is

$$(\alpha\beta) = 230, (\alpha B) = 148$$

$$(A\beta) = 151, (AB) = 471$$

Coefficient of Association is

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$= \frac{(471 \times 230) - (151 \times 148)}{(471 \times 230) + (151 \times 148)}$$

$$= \frac{85982}{130678} = +0.657$$

This shows that there is fairly high degree of positive association between eye colour of father and son.

E5) Let A denotes attribute of vaccination, and B that of attack

Then the given data are N = 1482; (B) = 368; (A) = 343; (AB) = 35

Now,

$$(\alpha\beta) = N - (A) - (B) + (AB) = 1482 - 343 - 368 + 35 = 806$$

$$(A\beta) = (A) - (AB) = 343 - 35 = 308$$

$$(\alpha B) = (B) - (AB) = 368 - 35 = 333$$

Yule's coefficient of association

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$= \frac{(35 \times 806) - (308 \times 333)}{(35 \times 806) + (308 \times 333)}$$

$$= \frac{-74354}{130774} = -0.57$$

Thus, we find that vaccination and small pox are in a high degree of negative association. Hence, vaccination can be regarded as a preventive measure of small pox.

E6) Let A denotes literacy and B denotes unemployment so that α denotes illiteracy and β denotes employment.

Now we have

$(\alpha B) = 250; (A\beta) = 25; (\alpha\beta) = 180; N = 500$

We put these figures in 2×2 table and get the frequencies of the remaining class

Attributes	**A**	**α**	**Total**
B	AB 45	αB 250	B 295
β	Aβ 25	α β 180	β 205
Total	A 70	α 430	N 500

Yule's Coefficient of Association is

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$= \frac{(45\times 180) - (25\times 250)}{(45\times 180) + (25\times 250)}$$

$$= \frac{1850}{14350} = +0.13$$

This shows that there is only marginal positive association between literacy and unemployment.

E7) Let A denotes poor children and B denotes children below normal weight. Then α would denote rich children and β would denote children above normal weight.

The data given are

$(AB) = 85; (\alpha B) = 20; (A\beta) = 6; (\alpha\beta) = 47;$

Yule's coefficient of association is

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$= \frac{(85\times 47) - (6\times 20)}{(85\times 47) + (6\times 20)}$$

$$= \frac{3875}{4115} = +0.94$$

There is high degree of association between poor children and children below normal weight.

This means chances of poor children being below normal weight are very high. Rich children will generally be above normal weight.

GLOSSARY

Correlation	:	Study of relationship between two variables say x and y.
Correlation Coefficient	:	A measure that gives the degree of in lines relationship between x and y.
Regression Coefficient	:	Measure of change in variable y corresponding to unit change in variable x and vice versa.