9. THEORY OF ATTRIBUTES

9.0 Introduction:

Generally statistics deal with quantitative data only. But in behavioural sciences, one often deals with the variable which are not quantitatively measurable. Literally an attribute means a quality or characteristic which are not related to quantitative measurements. Examples of attributes are health, honesty, blindness etc. They cannot be measured directly. The observer may find the presence or absence of these attributes. Statistics of attributes based on descriptive character.

9.1 Notations:

Association of attribute is studied by the presence or absence of a particular attribute. If only one attribute is studied, the population is divided into two classes according to its presence or absence and such classification is termed as division by dichotomy. If a class is divided into more than two scale-classes, such classification is called manifold classification.

Positive class which denotes the presence of attribute is generally denoted by Roman letters generally A, B, ... etc and the negative class denoting the absence of the attribute and it is denoted by the Greek letters α, β ... etc For example, A represents the attribute ‘Literacy’ and B represents ‘Criminal’. α and β represents the ‘Illiteracy’ and ‘Not Criminal’ respectively.

9.2 Classes and Class frequencies:

Different attributes, their sub-groups and combinations are called different classes and the number of observations assigned to them are called their class frequencies.

If two attributes are studied the number of classes will be 9. (i.e.,) (A), (α), (B), (β), (A β) (α β), (α B) and N.
The chart given below illustrate it clearly.

The number of observations or units belonging to class is known as its frequency are denoted within bracket. Thus (A) stands for the frequency of A and (AB) stands for the number objects possessing the attribute both A and B. The contingency table of order (2×2) for two attributes A and B can be displayed as given below

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>α</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>(AB)</td>
<td>(αB)</td>
<td>(B)</td>
</tr>
<tr>
<td>β</td>
<td>(Aβ)</td>
<td>(αβ)</td>
<td>(β)</td>
</tr>
<tr>
<td>Total</td>
<td>(A)</td>
<td>(α)</td>
<td>N</td>
</tr>
</tbody>
</table>

**Relationship between the class frequencies:**

The frequency of a lower order class can always be expressed in terms of the higher order class frequencies. 

i.e., \[ N = (A) + (\alpha) = (B) + (\beta) \]

\[ (A) = (AB) + (A\beta) \]

\[ (\alpha) = (\alpha B) + (\alpha\beta) \]

\[ (B) = (AB) + (\alpha B) \]

\[ (\beta) = (A\beta) + (\alpha \beta) \]

If the number of attributes is n, then there will be \(3^n\) classes and we have \(2^n\) cell frequencies.
9.3 Consistency of the data:

In order to find out whether the given data are consistent or not we have to apply a very simple test. The test is to find out whether any or more of the ultimate class-frequencies is negative or not. If none of the class frequencies is negative we can safely calculate that the given data are consistent (i.e the frequencies do not conflict in any way each other). On the other hand, if any of the ultimate class frequencies comes to be negative the given data are inconsistent.

Example 1:

Given \( N = 2500 \), \( (A) = 420 \), \( (AB) = 85 \) and \( (B) = 670 \). Find the missing values.

Solution:

We know \( N = (A) + (\alpha) = (B) + (\beta) \)

\[
(A) = (AB) + (A\beta) \\
(\alpha) = (\alpha B) + (\alpha\beta) \\
(B) = (AB) + (\alpha B) \\
(\beta) = (A\beta) + (\alpha \beta)
\]

From (2) \( 420 = 85 + (A\beta) \)

\[
(A\beta) = 420 - 85 \\
(A\beta) = 335
\]

From (4) \( 670 = 85 + (\alpha B) \)

\[
(\alpha B) = 670 - 85 \\
(\alpha B) = 585
\]

From (1) \( 2500 = 420 + (\alpha) \)

\[
(\alpha) = 2500 - 420 \\
(\alpha) = 2080
\]

From (1) \( (\beta) = 2500 - 670 \)

\[
(\beta) = 1830
\]

From (3) \( 2080 = 585 + (\alpha\beta) \)

\[
(\alpha\beta) = 1495
\]
Example 2:
Test the consistency of the following data with the symbols having their usual meaning.
\( N = 1000 \) \((A) = 600\) \((B) = 500\) \((AB) = 50\)
Solution:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>(\alpha)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>50</td>
<td>450</td>
<td>500</td>
</tr>
<tr>
<td>(\beta)</td>
<td>550</td>
<td>-50</td>
<td>500</td>
</tr>
<tr>
<td>Total</td>
<td>600</td>
<td>400</td>
<td>1000</td>
</tr>
</tbody>
</table>

Since \((\alpha \beta)) = -50\), the given data is inconsistent.

Example 3:
Examine the consistency of the given data. \( N = 60 \) \((A) = 51\) \((B) = 32\) \((AB) = 25\)
Solution:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>(\alpha)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>25</td>
<td>7</td>
<td>32</td>
</tr>
<tr>
<td>(\beta)</td>
<td>26</td>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>9</td>
<td>60</td>
</tr>
</tbody>
</table>

Since all the frequencies are positive, it can be concluded that the given data are consistent.

9.4 Independence of Attributes:
If the attributes are said to be independent the presence or absence of one attribute does not affect the presence or absence of the other. For example, the attributes skin colour and intelligence of persons are independent.
If two attributes A and B are independent then the actual frequency is equal to the expected frequency

\[(AB) = \frac{(A).(B)}{N}\]

Similarly \[(\alpha \beta) = \frac{(\alpha).(\beta)}{N}\]

### 9.4.1 Association of attributes:

Two attributes A and B are said to be associated if they are not independent but they are related with each other in some way or other.

The attributes A and B are said to be positively associated if \[(AB) > \frac{(A).(B)}{N}\]

If \[(AB) < \frac{(A).(B)}{N}\] then they are said to be negatively associated.

#### Example 4:

Show that whether A and B are independent, positively associated or negatively associated.

\[(AB) = 128, \ (\alpha B) = 384, \ (A\beta) = 24 \text{ and } (\alpha\beta) = 72\]

**Solution:**

\[(A) = (AB) + (A\beta) = 128 + 24 = 152\]

\[(B) = (AB) + (\alpha B) = 128 + 384 = 512\]

\[(\alpha) = (\alpha B) + (\alpha\beta) = 384 + 72 = 456\]

\[
\therefore (\alpha) = 456
\]

\[(N) = (A) + (\alpha) = 152 + 456 = 608\]

239
\[
\frac{(A) \times (B)}{N} = \frac{152 \times 512}{608} = 128
\]

\[
(AB) = 128
\]

\[
\therefore \ (AB) = \frac{(A) \times (B)}{N}
\]

Hence A and B are independent

**Example 5:**

From the following data, find out the types of association of A and B.

1) \(N = 200\) \((A) = 30\) \((B) = 100\) \((AB) = 15\)
2) \(N = 400\) \((A) = 50\) \((B) = 160\) \((AB) = 20\)
3) \(N = 800\) \((A) = 160\) \((B) = 300\) \((AB) = 50\)

**Solution:**

1. Expected frequency of \((AB) = \frac{(A).(B)}{N} = \frac{(30)(100)}{200} = 15\)

Since the actual frequency is equal to the expected frequency, i.e., \(15 = 15\), therefore A and B are independent.

2. Expected frequency of \((AB) = \frac{(A).(B)}{N} = \frac{(50)(160)}{400} = 20\)

Since the actual frequency is greater than expected frequency. i.e., \(25 > 20\), therefore A and B are positively associated.

3. Expected frequency of \((AB) = \frac{(A).(B)}{N} = \frac{(160)(300)}{800} = 60\)

Since actual frequency is less than expected frequency i.e., \(50 < 60\) therefore A and B are negatively associated.
9.5 Yules’ co-efficient of association:

The above example gives a rough idea about association but not the degree of association. For this Prof. G. Undy Yule has suggested a formula to measure the degree of association. It is a relative measure of association between two attributes A and B.

If \((AB), (\alpha B), (A\beta)\) and \((\alpha \beta)\) are the four distinct combination of \(A, B, \alpha\) and \(\beta\) then Yules’ co-efficient of association is

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

Note:

I. If \(Q = +1\) there is perfect positive association
   If \(Q = -1\) there is perfect negative association
   If \(Q = 0\) there is no association (ie) A and B are independent

1. For rememberance of the above formula, we use the table below

<table>
<thead>
<tr>
<th>(\alpha)</th>
<th>(B)</th>
<th>(B)</th>
<th>(\alpha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha)</td>
<td>(A)</td>
<td>(AB)</td>
<td>(\alpha B)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(A\beta)</td>
<td>(\alpha \beta)</td>
<td></td>
</tr>
</tbody>
</table>

Example 6:

Investigate the association between darkness of eye colour in father and son from the following data.

- Fathers’ with dark eyes and sons’ with dark eyes = 50
- Fathers’ with dark eyes and sons’ with no dark eyes = 79
- Fathers’ with no dark eyes and sons with dark eyes = 89
- Neither son nor father having dark eyes = 782

Solution:

Let A denote the dark eye colour of father and B denote dark eye colour of son.

<table>
<thead>
<tr>
<th>(A)</th>
<th>(\alpha)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(B)</td>
<td>50</td>
<td>89</td>
</tr>
<tr>
<td>(\beta)</td>
<td>79</td>
<td>782</td>
</tr>
<tr>
<td>Total</td>
<td>129</td>
<td>871</td>
</tr>
</tbody>
</table>
Yules’ co-efficient of association is
\[ Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)} \]
\[ = \frac{50 \times 782 - 79 \times 89}{50 \times 782 + 79 \times 89} \]
\[ = \frac{32069}{46131} = 0.69 \]
∴ there is a positive association between the eye colour of fathers’ and sons’.

Example 7:
Can vaccination be regarded as a preventive measure of small pox from the data given below.
Of 1482 persons in a locality, exposed to small pox, 368 in all were attacked, among the 1482 persons 343 had been vaccinated among these only 35 were attacked.

Solution:
Let A denote the attribute of vaccination and B denote that of attacked.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>α</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>35</td>
<td>333</td>
<td>368</td>
</tr>
<tr>
<td>β</td>
<td>308</td>
<td>806</td>
<td>1114</td>
</tr>
<tr>
<td>Total</td>
<td>343</td>
<td>1139</td>
<td>1482</td>
</tr>
</tbody>
</table>

Yules’ co-efficient of association is
\[ Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)} \]
\[ = \frac{35 \times 806 - 308 \times 333}{35 \times 806 + 308 \times 333} \]
\[ = \frac{-74354}{130774} = -0.57 \]
i.e., there is a negative association between attacked and vaccinated. In other words there is a positive association between not attacked and vaccinated. Hence vaccination can be regarded as a preventive measure for small pox.
Example 8:
In a co-educational institution, out of 200 students, 150 were boys. They took an examination and it was found that 120 passed, 10 girls failed. Is there any association between sex and success in the examination.

Solution:
Let A denote boys and α denote girls. Let B denote those who passed the examination and β denote those who failed.
We have given \( N = 200 \) \( (A) = 150 \) \( (AB) = 120 \) \( (\alpha\beta) = 10 \)
Other frequencies can be obtained from the following table

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>α</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>120</td>
<td>40</td>
<td>160</td>
</tr>
<tr>
<td>β</td>
<td>30</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>Total</td>
<td>150</td>
<td>50</td>
<td>200</td>
</tr>
</tbody>
</table>

Yule’s co-efficient of association is
\[
Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}
\]
\[
= \frac{120 \times 10 - 30 \times 40}{120 \times 10 + 30 \times 40} = 0
\]
Therefore, there is no association between sex and success in the examination.

Recall
(A) (B) denote positive attributes
(α) (β) denote negative attributes
2 × 2 contingency table.

<table>
<thead>
<tr>
<th>X</th>
<th>A</th>
<th>α</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>(AB)</td>
<td>(αB)</td>
<td>(B)</td>
</tr>
<tr>
<td>β</td>
<td>(Aβ)</td>
<td>(αβ)</td>
<td>(β)</td>
</tr>
<tr>
<td>Total</td>
<td>(A)</td>
<td>(α)</td>
<td>N</td>
</tr>
</tbody>
</table>
Vertical Total
(AB) + (Aβ) = (A)
(αB) + (αβ) = (α)
(A) + (α) = N

Horizontal Total
(AB) + (αB) = B
(Aβ) + (αβ) = β
(B) + (β) = N

Types of Association

Positive Association if (AB) > \( \frac{(A)(B)}{N} \)

Negative Association if (AB) < \( \frac{(A)(B)}{N} \)

Independent if (AB) = \( \frac{(A)(B)}{N} \)

Yule’s co-efficient of Association
\[
Q = \frac{(AB)(\alpha \beta) - (A\beta)(\alpha B)}{(AB)(\alpha \beta) + (A\beta)(\alpha B)}
\]

Exercise – 9

I. Choose the best answer:
1. Measures of association in usually deal with
   (a) Attributes (b) Quantitative factors
   (c) Variables (d) Numbers
2. The frequency of class can always be expressed as a sum of frequencies of
   (a) Lower order classes (b) Higher order classes
   (c) Zero order classes (d) None of the above
3. With the two attributes the total number of class frequencies is
   (a) Two (b) Four (c) Eight (d) Nine
4. If for two the attributes are A and B, (AB) > \( \frac{(A)(B)}{N} \), the attributes are
   (a) Independent (b) Positively associated
   (c) Negatively associated (d) No conclusion
5. In case of two attributes A and B the class frequency 
   \((AB) = 0\) the value of \(Q\) is
   \( (a) \ 1 \quad (b) -1 \quad (c) \ 0 \quad (d) -1 \leq Q \leq 1\)

II. Fill in the blanks:
6. If an attribute has two classes it is said to be ____________
7. In case of consistent data, no class frequency can be 
   ____________
8. If A and B are independent Yule’s co-efficient is equal to 
   ____________
9. If A and B are negatively associated then ____________
10. If \(N = 500, (A) = 300, (B) = 250\) and \((AB) = 40\) the data are 
    ____________

III. Answer the following:
11. Give a brief idea of notations used in classification of 
    attributes
12. How can the frequencies for various attributes be displayed 
    in contingency table
13. What do you understand by consistency of data.
14. Write briefly about association of attributes.
15. Give Yule’s co-efficient of association

IV. Problems
16. For two attributes A and B, we have \((AB) = 35, (A) = 55; \) 
   \(N=100\) and \((B) = 65\). Calculate the missing values.
17. From the following ultimate class frequencies, find the 
   frequencies of positive and negative classes and the total 
   number of observations. \((AB) = 9, (A\beta) = 14, (\alphaB) = 4 \) 
   and \((\alpha\beta) = 37\)
18. Verify whether the given data \(N = 100, (A) = 75, (B) = 60\) 
   and \((AB) = 15\) are consistent.
19. Find whether A and B are independent in the following data 
   \((AB) = 256 \quad (\alphaB) = 768 \quad (A\beta) = 48 \quad (\alpha\beta) = 144\)
20. In a report on consumer’s preference it was given that out of 
    500 persons surveyed 410 preferred variety A 380 preferred
variety B and 270 persons linked both. Are the data consistent?

21. For two attributes A and B, we have \((AB) = 35\), \((A) = 55\), 
\(N=100\), \((\alpha\beta) = 20\). Calculate the Yule’s co-efficient of association.

22. Given \(N = 1500\), \((A) = 383\), \((B) = 360\) and \((AB) = 35\). 
Prepare \(2 \times 2\) contingency table and compute Yule’s co-efficient of association and interpret the result.

23. In an experiment on immunization of cattle from tuberculosis the following results were obtained.

<table>
<thead>
<tr>
<th></th>
<th>Affected</th>
<th>Unaffected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inoculated</td>
<td>12</td>
<td>26</td>
</tr>
<tr>
<td>Not inoculated</td>
<td>16</td>
<td>6</td>
</tr>
</tbody>
</table>

By calculating Yule’s co-efficient of association, examine the effect of vaccine is in controlling the disease.

24. Calculate the co-efficient of association between the intelligence of fathers and sons from the following data

Intelligent fathers with intelligent sons = 300
Intelligent fathers with dull sons = 100
Dull fathers with intelligent sons = 50
Dull fathers with dull sons = 500

25. Out of 3000 unskilled workers of a factory, 2000 come from rural area and out of 1200 skilled workers 300 come from rural area. Determine the association between skill and residence

26. In an anti-malarial campaign in a certain area, quinine was administrated to 812 persons out of a total population of 3428. The number of fever cases is shown below:

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Fever</th>
<th>No Fever</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quinine</td>
<td>20</td>
<td>792</td>
</tr>
<tr>
<td>No quinine</td>
<td>220</td>
<td>2216</td>
</tr>
</tbody>
</table>

Examine the effect of quinine on controlling malaria.

27. 1500 candidates appeared for competitive examinations 425 were successful. 250 had attended a coaching class and of
these 150 came out successful. Estimate the utility of the coaching class.

28. In an examination at which 600 candidates appeared of them 348 were boys. Number of passed candidates exceeded the number of failed candidates by 310. Boys failing in the examination numbered 88. Find the coefficient of association between male sex and success in examination.

29. Following data relate to literacy and unemployment in a group of 500 persons. Calculate Yule’s co-efficient of association between literacy and unemployment and interpret it

<table>
<thead>
<tr>
<th>Literate unemployed</th>
<th>= 220</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literate employed</td>
<td>= 20</td>
</tr>
<tr>
<td>Illiterate Employed</td>
<td>= 180</td>
</tr>
</tbody>
</table>

30. In a group of 400 students, the number of married is 160. Out of 120 students who failed 48 belonged to the married group. Find out whether the attributes of marriage and failure are independent.

Answers

I.
1. (a) 2. (b) 3. (d) 4. (b) 5. (b)

II.
6. Dichotomy 7. Negative 8. 0
9. \( AB < \frac{(A).(B)}{N} \) 10. Inconsistent

IV.
16

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>( \alpha )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>35</td>
<td>30</td>
<td>65</td>
</tr>
<tr>
<td>( \beta )</td>
<td>20</td>
<td>15</td>
<td>35</td>
</tr>
<tr>
<td>Total</td>
<td>55</td>
<td>45</td>
<td>100</td>
</tr>
</tbody>
</table>
17.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>α</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>9</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>β</td>
<td>14</td>
<td>37</td>
<td>51</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>41</td>
<td>64</td>
</tr>
</tbody>
</table>

Total No of observations = 64

18. Inconsistent
19. A and B are independent
20. Inconsistent
21. 0.167
22. – 0.606, Negative association
23. – 0.705, Vaccine is effective
24. + 0.935
25. Negative association between skill and residence.
26. – 0.59. Negative association ∴ quinine is effective.
27. + 0.68. Coaching class are useful
28. – 0.07
29. 0.92 Positive association between literacy and unemployment
30. Q = 0, Marriage and failure are independent.