Module 5: Attributes

Dr. A. N. Basugade
M.Sc. Ph.D
Head, Department of Statistics,
GopalKrishnaGokhaleCollege,Kolhapur
Email – arunb1961@gmail.com

Overview

- **Definition**
- **Order of class and class frequency**
- **Consistency of Attributes**
- **Independence &Association of Attributes**
- **Coefficients of Association**
- **Coefficient of Colligation**

- **Definition:** An attribute is a quality or a characteristic which cannot be measured but which can be marked by their presence or absence. For instance, sex, literacy, honesty, nationality etc. are attributes. Given an attribute the population can be divided into two classes; one possessing that attribute and the other not possessing it. Such a classification into two classes is called dichotomous.

- **Notations:** Suppose the population is divided into two classes according to the presence or absence of an attribute. The class possessing the attribute is called a positive class and is denoted by capital letters *A, B, C* etc. The class not possessing the attribute is called a negative class and is denoted by small Greek letters α, β, γ etc. Thus, if ‘A' denotes the class of 'males' then ' α ' will denote the class of 'females';

- **Order of Classes and Class-Frequencies:** A class representing one attribute is called a class of first order. Thus, *A, B,* C, α, β, γ are classes *of first order.* A class representing two attribute is called a class of *second order.* Thus, *AB, AC, A*β etc. are classes of *second order.* Similarly, *A*β*C, ABC,* αβC are classes *of third order. i. e.* **order denotes the number of attributes in that class.**

- **Class-Frequencies:** The number of items belonging to a class is called the frequency of that class. The class frequency is denoted by putting the letter (or letters) denoting the class in a bracket. Thus, *(A)* stands for the number of items possessing the attribute *A ; (*α*B)* stands for the number of items, not possessing *A* and possessing *B*.
- The frequency of a positive class is called positive class frequency *e.g. (AB)* and frequency of a negative class is called negative class frequency *e.g.* (αβγ).

- **Ultimate Class frequencies:** The class-frequencies of highest order are called ultimate class-frequencies. Thus, in the case of two attributes class-frequencies of order two are ultimate class-frequencies. If *A* and *B* are attributes then *(AB)*, *(A*β), (αB), (αβ) are ultimate class-frequencies. If we are considering *n* attributes, the Ultimate class frequencies will have *n* symbols.

- Thus the total number of ultimate class frequencies in case of two attributes are $2^2 = 4$ and for three attributes are $2^3 = 8$

- The total number of ultimate class frequencies in case of n attributes are 2^n

- The total number of positive class frequencies are 2^n

- The total number of class frequencies of all order are 3^n

- **Consistency of Data:** If all the class frequencies are positive then the data is said to be consistent. Or if all the ultimate class-frequencies are non-negative then the data is consistent.

- **Conditions of Consistency:** For one attribute

  (i) $(A) > 0$, (ii) $(\alpha) > 0$. But $(A)+(\alpha)= N$,

  (iii) $(A) \leq N$, (iv) $(\alpha) \leq N$

  For Two attributes

  (i) $N = (A) + (\alpha) = (B) + (\beta)$

  (ii) $(A) = (AB) + (A\beta)$, and $(B) = (AB) + (\alpha B)$

  $(\alpha) = (\alpha B) + (\alpha\beta)$, and $(\beta) = (A\beta) + (\alpha\beta)$

- **Class symbols as an operator**: If we look at the class symbols *A, B* as an operator, A stands for the ratio of items possessing the attribute A. Then *AN* means multiplying *N* by this ratio but this is the class frequency *(A)* of A. Hence, we have $AN = (A)$.

  Similarly, *A(B)* means multiplying *(B)* by the ratio *A,* but this will be the number of members having both attributes *AB i.e. (AB).* Thus, we have, $A*(B)=(AB)=AB*N$.

  Using class symbol as an operator we can obtain various relations as follows :

  i) $(AB) = N-(\alpha)-(\beta)+(\alpha\beta)$

  ii) $(\alpha\beta) = N-(A)-(B)+(AB)$

  iii) $(AB) \geq (A) + (B) - N$

  iv) $(\alpha\beta) \geq (\alpha) + (\beta) - N$

- Thus the consistency conditions for two attributes are

(i) $(AB) \geq 0$.

(ii) $(\alpha\beta) \geq 0$ $(A\beta) \geq 0$.

(iii) $(\alpha B) \geq 0$

(iv) $(\alpha\beta) \geq 0$

(v) $(A) \geq (AB)$.

(vi) $(B) \geq (AB)$

(vii) $(AB) \geq (A) + (B) - N$

- The consistency conditions for Three attributes are

(i) $(ABC) \geq 0$.

(ii) $(AB) \geq (ABC)$

(iii) $(AC) \geq (ABC)$

(iv) $(BC) \geq (ABC)$

(v) $(ABC) \geq (AB) +(AC) - (B)$

(vi) $(ABC) \geq (AB) + +(AC) - (A)$

(vii) $(ABC) \geq (AC) + (BC) - (C)$

(viii) $(ABC) \leq (AB) + (BC) + (AC) - (A) - (B) - (C) + N$

(ix) $(AB) + (AC) + (BC) \geq (A) + (B) + (C) - N$

(x) $(AC) + (BC) - (AB) \leq (C)$

(xi) $(AB) + (BC) - (AC) \leq (B)$

(xii) $(AB) + (AC) - (BC) \leq (A)$

Example 1: From the following data check whether the data are consistent or not. $(A) = 120$, $(B) = 165$, $(AB) = 160$, $N = 400$.

Solution : $(\alpha B) = (B) - (AB) = 165-160 = 5$,

$$(\alpha) = N - (A) = 400-120 = 280$$

$$(\alpha\beta) = (\alpha) - (\alpha B) = 280 - 5 = 275$$

$$(A\beta) = (A) - (AB) = 120 - 160 = -40$$

since one of the ultimate class frequency is negative the given data is not consistent.

- ## **Independence of Attributes:**

The two attributes are said to be independent if one is not affected by the presence or absence of other.

If two attributes *A* and *B* are independent, we expect the proportion of A's amongst B's is same the proportion of A's amongst β's.

i. e. $(AB)/(B) = (A\beta)/(\beta)$

Similarly, $(A\beta)/(A) = (\alpha B)/(\alpha)$

- **Criterion of independence** : If A and B are independent then by the above definition of independence we get,

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

and $(A\beta)/(A) = (\alpha B)/(\alpha)$ which gives

(i) $(A\beta)/(A) = (\alpha\beta)/(\alpha)$

(ii) $(AB) = (A) * (B)/(N)$

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

These are the three criterions of independence of attributes.

- **Symbols $(AB)_0$ and δ :** If A & B are independent then we have $(AB) = (A) * (B)/(N)$, we denote $(AB)_0 = (A)(B)/(N)$
 and the difference between (AB) & $(AB)_0$ is denoted by δ,
 i.e. $\delta = (AB) - (AB)_0 = (AB) - (A)(B)/(N)$

 If A & B are independent, then $\delta = 0$,
 δ can also be represented as
 $\delta = (1/N)\,[(AB)\,(\alpha\beta) - (A\beta) - (\alpha B)]$

- **Association of Attributes:**

To study relationship if the characteristics cannot be measured *i.e.* to study relationship between two attributes we use the technique called association **of attributes.**

If the attributes are not independent and they are related with each other in some way then they are said to be associated to one another.

i) **Positive Association**: If 'A' occurs large number of times with 'B' than β then A & B are said to be Positively Associated.

i.e. $(AB) > (A)(B)/(N)$ then A & B are Positively Associated.

ii) **Negative Association**: If 'A' occurs small number of times with 'B' or α occurs large number of times with 'B' then they are said to be negatively Associated.

i.e. $(AB) < (A)(B)/(N)$ then A & B are Negatively Associated.

i.e. if $\delta > 0$ then A & B are Positively Associated and

if $\delta < 0$ then A & B are Negatively Associated.

iii) If A cannot occur without B or all A's are B's then (AB) = (A) then A & B are completely associated.

iv) If all A's are β's then (αB) = (α) then A & B are completely disassociated.

- **Measures of Association**

  There are two methods of measuring association:

  a) Yule's coefficient of Association, b) Coefficient of Colligation.

a) Yule's Coefficient of association : This is the most commonly used method of studying association and it is denoted by Q and is defined by

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

i) It A and *B* are independent then (AB)(αβ) - (Aβ)(αB) = 0, hence *Q* = *0*.

ii) If A and B are completely associated, then (AB) = (A) and (AB) = (B) i.e. (Aβ) = 0 and (αB) = 0. Thus Q =1

iii) If *A* and *B* are completely dissociated, then *(AB)* = 0 & (αβ) = 0 Thus Q = -1

i.e. Q lies between -1 to 1.

b) Coefficient of Colligation : Another measure of association coefficient suggested by Yule is the coefficient of colligation denoted by Y and is given by

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

i) If A & B are independent then (AB)(αβ) = (Aβ)(αB)

Thus Y = 0

ii) If A and B are completely associated, then (AB) = (A) and (AB) = (B) i.e. (Aβ) = 0 and (αB) = 0. Thus Y =1

iii) If *A* and *B* are completely dissociated, then *(AB)* = 0 & (αβ) = 0 Thus Y = -1

i.e. Y lies between -1 to 1.

Relation between Q and Y is

$$Q = 2Y/(1+y^2)$$

and

$$Y = \frac{\sqrt{1+Q} - \sqrt{1-Q}}{\sqrt{1+Q} + \sqrt{1-Q}}$$

Summary: (Learning Outcomes)

At the end of this module student must be able to

- **Define Attributes.**
- **Explain Order of class and class frequency.**
- **Describe Consistency of Attributes.**
- **Explain Independence &Association of Attributes.**
- **Define Coefficients of Association and Coefficient of Colligation.**