K.T.S.P. Mandal's

Hutatma Rajguru Mahavidyalaya

Rajgurunagar -410505

F.Y.B.Sc

Semester-III (CBCS 2019 Pattern)

As per the New Syllabus

Subject - ST-111: Descriptive Statistics I

Prof. S.R.Thorat
(Head, Department of Statistics)

Chapter 5 – Theory of Attributes

(8) Theory of Attributes

Qualitative characteristics:

The characteristics which cannot be measured in numbers are known as qualitative characteristics.

Attribute:

The qualitative characteristics are known as attribute.

e.g (1) Blood group of a person : O, A, B, AB

(2) Examination result : Pass, Fail

(3) The unit inspected in a certain manufacturing industry : Good or bad (Non defective or defective)

Notation :

suppose

A : Male	B : Rural	C : literate
α : female	β : Urban	γ : Illiterate

AB : Rural male
$A\beta$: urban male
αB : Rural female
ABC : literate rural male
$\alpha\beta\gamma$: illiterate urban female.

Here . The presence of Attribute is denoted by capital letters A, B, C etc & its absence denoted by

Greek letters α, β, γ etc

Class : The symbols such as A, B, AB, $A\beta$ denotes the class of observations.

class frequency : The number of observations belonging to a particular class is known as frequency of that class.

It is denoted by the respective class symbol enclosed in bracket.

(A) : Indicates frequency of class A

(AB) : Indicates frequency of class AB

positive Attributes :

The attributes which are denoted by capital letters A, B, C etc are known as positive attribute.

Negative Attributes :

The attributes denoted by greek letters α, β, γ etc are known as Negative attribute.

positive classes :

The classes denoted by capital letters or combination of capital letters are known as positive classes.

for example : A, B, AB, ABC

Negative classes:

The classes which are denoted by greek letters or combination of greek lett are known as negative classes.

ex: α, $\alpha\beta$, $\alpha\beta\gamma$

order of class:

A class denoted by combination of n attributes is called nth order class.

First order class : $A, B, C, \alpha, \beta, \gamma$ etc

Second order class : $AB, AC, BC, \alpha\beta, \alpha\gamma, \beta\gamma$

Third order class : $ABC, A\beta C, \alpha\beta\gamma, \alpha\beta C$ etc.

Ultimate class frequencies:

The frequencies of the classes of highest order are known as ultimate class frequencies

	Attribute	Ulitimate Class frequencies.	No.
i)	A	(A), (α)	2
ii)	A, B	(AB), $(A\beta)$, (αB), $(\alpha\beta)$	4
iii)	A, B, C	(ABC), $(AB\gamma)$, $(A\beta C)$, (αBC), $(\alpha B\gamma)$, $(\alpha\beta C)$, $(\alpha B\gamma)$, $(A\beta\gamma)$	8
	n		2^n

Total number of class frequencies:

Let us consider two attributes A & B, using the following table we count the total No. of class frequencies.

order	class frequencies	No. of classes.
0	N	1
1	$(A),(B),(\alpha),(\beta)$	4
2	$(AB),(A\beta),(\alpha B),(\alpha\beta)$	4
	Total	9

3^2

Consider 3 attributes A, B and C

order	class frequencies	No. of classes.
0	N	1
1	$(A),(B),(C),(\alpha),(\beta),(\gamma)$	6
2	$(AB),(A\beta),(\alpha B)(\alpha\beta)$ $(AC),(A\gamma),(\alpha C),(\alpha\gamma)$ $(BC),(B\gamma),(\beta C),(\beta\gamma)$	12
3	$(ABC),(AB\gamma),(A\beta C),(A\beta\gamma)$ $(\alpha\beta\gamma),(\alpha B\gamma),(\alpha\beta C),(\alpha BC)$	8
	Total	27

3^3

In general for n attributes the total No. of class frequencies are 3^n

Relation Among the Class frequencies:

Let us consider two attributes A & B, and the total frequency is N.

$$N = (A) + (\alpha)$$

$$N = (B) + (\beta)$$

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

We can present this relations among the class frequencies in tabular form as follows.

	B	β	Total
A	(AB)	$(A\beta)$	(A)
α	(αB)	$(\alpha\beta)$	(α)
Total	(B)	(β)	N.

Ex (1) Given the following frequencies - obtain the remaining class frequencies.

$(AB) = 20, \quad (A\beta) = 10$

$(\alpha B) = 15, \quad (\alpha\beta) = 55$

Solution: Given $(AB) = 20, \quad (A\beta) = 10$

$(\alpha B) = 15, \quad (\alpha\beta) = 55$

	B	β	
A	$(AB) = 20$	$(A\beta) = 10$	$(A) = 30$
α	$(\alpha B) = 15$	$(\alpha\beta) = 55$	$(\alpha) = 70$
	$(B) = 35$	$(\beta) = 65$	$N = 100$

(2) A report regarding examination is given below: Total number of candidates appeared in examination is 1000. There are 550 boys among 1000 - 700 candidates were successful. Number of successful boys is 300. Find

i) No. of successful girls

ii) No. of unsuccessful girls

iii) No. of unsuccessful boys.

Solution: A : Boy $\quad \alpha$: Girl

B : Success $\quad \beta$: unsuccess

Given,

$$N = 1000$$
$$(A) = 550$$
$$(B) = 700$$
$$(AB) = 300$$

$(\alpha B) = ?$, $(\alpha \beta) = ?$, $(A\beta) = ?$

	B	β	
A	$(AB) = 300$	$(A\beta) = 250$	$(A) = 550$
α	$(\alpha B) = 400$	$(\alpha\beta) = 50$	$(\alpha) = 450$
	$(B) = 700$	$(\beta) = 300$	$N = 1000$

i) No. of successful girls $= (\alpha B) = 400$

ii) No. of unsuccessful girls $= (\alpha \beta) = 50$

iii) No. of unsuccessful boys $= (A\beta) = 250$

(3) In an examination 60 % passed in physics, 52 % passed in statistics while 32 % failed in both the subjects. Using relation among the class frequencies find the percentage of candidates

i) passed in both subjects
ii) passed only in physics
iii) passed only in statistics.

Solution:

A : passing in physics α : failing in physics
B : passing in statistics β : Failing in statistics

$(A) = 60$, $(B) = 52$, $(\alpha\beta) = 32$
$N = 100$

	B	β	
A	$(AB) = 44$	$(A\beta) = 16$	$(A) = 60$
α	$(\alpha B) = 08$	$(\alpha\beta) = 32$	$(\alpha) = 40$
	$(B) = 52$	$(\beta) = 48$	$N = 100$

i) % of candidates passed in both the subjects
$= (AB) = 44\ \%$.

ii) % of candidates passed only in physics
$= (A\beta) = 16\ \%$

iii) % of candidates passed only in statistics
$= (\alpha B) = 8\ \%$.

4) If $(A) = (B) = \frac{N}{2}$ then show that

$(AB) = (\alpha\beta)$ and

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

Solution: Given, $(A) = (B)$

$$\cancel{(AB)} + (A\beta) = \cancel{(AB)} + (\alpha B)$$

$$(A\beta) = (\alpha B) \quad \text{--- (1)}$$

further, $(A) = N/2$ --- (2)

We know that

$$(A) + (\alpha) = N$$

$$N/2 + (\alpha) = N$$

$$(\alpha) = N - N/2$$

$$(\alpha) = N/2 \quad \text{--- (3)}$$

from (2) & (3)

$$(A) = (\alpha)$$

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

from (1) $(A\beta) = (\alpha B)$

$$(AB) + \cancel{(\alpha B)} = \cancel{(\alpha B)} + (\alpha\beta)$$

$$\boxed{(AB) = (\alpha\beta)}$$

Result: If a population is classified by two Attributes A and B then

$$(A) + (B) - N \le (AB) \le \min\{(A), (B)\}$$

Proof: we know that

$$(A) = (AB) + (A\beta)$$
$$\Rightarrow (AB) \le (A) \quad \text{--- } ①$$

$$(B) = (AB) + (\alpha B)$$
$$\Rightarrow (AB) \le (B) \quad \text{--- } ②$$

	B	β	
A	(AB)	$(A\beta)$	(A)
α	(αB)	$(\alpha\beta)$	(α)
	(B)	(β)	N

from ① & ②

$$(AB) \le \min\{(A), (B)\} \quad \text{--- } ③$$

since $(\alpha\beta) \ge 0$

$$(\alpha) - (\alpha B) \ge 0$$
$$N - (A) - [(B) - (AB)] \ge 0$$
$$N - (A) - (B) + (AB) \ge 0$$
$$\Rightarrow (AB) \ge (A) + (B) - N \quad \text{--- } ④$$

by combining equation ③ & ④

$$\boxed{(A) + (B) - N \le (AB) \le \min\{(A), (B)\}}$$

Consistency of Data:

we can check whether all frequencies are non-negative. If some class frequencies are negative then the corresponding data is not consistent.

Conditions for Consistency of Data

case (i) For single Attribute A

i) $(A) \geq 0$

ii) $(\alpha) \geq 0 \Rightarrow N-(A) \geq 0$

$\Rightarrow (A) \leq N$

case (ii) For two Attributes A and B.

	B	β	Total
A	(AB)	(Aβ)	(A)
α	(αB)	($\alpha\beta$)	(α)
Total	(B)	(β)	N

i) $(AB) \geq 0$

ii) $(A\beta) \geq 0 \Rightarrow (A)-(AB) \geq 0$

$\Rightarrow (AB) \leq (A)$.

iii) $(\alpha B) \geq 0 \Rightarrow (B)-(AB) \geq 0$

$\Rightarrow (AB) \leq (B)$

iv) $(\alpha\beta) \geq 0 \Rightarrow (\alpha)-(\alpha B) \geq 0$

$N-(A)-[(B)-(AB)] \geq 0$

$N-(A)-(B)+(AB) \geq 0$

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

Examine whether the following data are consistent

$N = 100$, $(A) = 30$, $(B) = 80$, $(AB) = 40$

Solution:

	B	β	
A	$(AB) = 40$	$(A\beta) = -10$	$(A) = 30$
α	$(\alpha B) = 40$	$(\alpha\beta) = 30$	$(\alpha) = 70$
	$(B) = 80$	$(\beta) = 20$	$N = 100$

Here, $(A\beta) = -10 \ngeq 0$

Hence, the given data is not consistent.

Alternately since

$(AB) \nleq (A)$

Hence, the given data is not consistent.

(2) Examine whether the following data are consistent :

$N = 200$, $(A) = 150$, $(B) = 80$, $(AB) = 25$

Solution :

	B	β	
A	$(AB) = 25$	$(A\beta) = \underline{125}$	$(A) = 150$
α	$(\alpha B) = \underline{55}$	$(\alpha\beta) = \underline{-5}$	$(\alpha) = \underline{50}$
	$(B) = 80$	$(\beta) = \underline{120}$	$N = 200$

Here, $(\alpha\beta) = -5 \ngtr 0$

Hence the given data is inconsistent.

Independence of Attribute:

Attributes A and B are said be independent if

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

or

$$\frac{(AB)}{(A)} = \frac{(\alpha B)}{(\alpha)}$$

Result: If the attributes A and B are independent then

i) A & β are independent
ii) α & B are independent
iii) α & β are independent

Ex. Test whether the attributes A and B are independent.

Given that, $(AB) = 10$, $(A\beta) = 30$, $(\alpha B) = 40$, $(\alpha\beta) = 120$

Solution:

	B	β	
A	$(AB) = 10$	$(A\beta) = 30$	$(A) = 40$
α	$(\alpha B) = 40$	$(\alpha\beta) = 120$	$(\alpha) = 160$
	$(B) = 50$	$(\beta) = 150$	$N = 200$

Consider,

$$\frac{(A)(B)}{N} = \frac{40 \times 50}{200}$$

$$= 10$$

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

i.e. $$(AB) = \frac{(A)(B)}{N}$$

Since the condition for independence is satisfied the attributes A and B are independent.

(2) Given that, $(AB) = 20$, $(A\beta) = 10$, $(\alpha B) = 15$, $(\alpha\beta) = 55$ verify that the attributes A & B are independent.

Solution :

	B	β	
A	$(AB) = 20$	$(A\beta) = 10$	$(A) = 30$
α	$(\alpha B) = 15$	$(\alpha\beta) = 55$	$(\alpha) = 70$
	$(B) = 35$	$(\beta) = 65$	$N = 100$

Here, $(AB) = 20$

consider, $\dfrac{(A)(B)}{N} = \dfrac{30 \times 35}{100} = \dfrac{1050}{100}$

$= 10.5$

$\neq (AB)$

i.e $(AB) \neq \dfrac{(A)(B)}{N}$

Hence the two attributes A and B are not independent.

Association and Dissociation:

The two attributes A and B are independent if

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

otherwise

$$(AB) \neq \frac{(A)(B)}{N}$$

If $(AB) > \frac{(A)(B)}{N}$

then we say that the two attributes are positively associated or associated.

If $(AB) < \frac{(A)(B)}{N}$

then we say that the two attributes are Negatively associated or dissociated.

Complete Association:

Attributes A and B are said to be completely associated if

$$(AB) = (A) \quad \text{or} \quad (AB) = (B)$$

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

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

Complete Dissociation:

Two attributes A and B are said to be completely dissociated if

$(AB) = 0$ or $(\alpha\beta) = 0$

Yule's Coefficient of Association:

Yules coefficient of association between two attributes A and B is denoted by Q_{AB} and is defined as

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

by using this formula we get the idea about the association among two attributes A & B also the amount of association

Interpretation:

i) If $Q > 0$ $\Rightarrow$ positive Association between Attributes A & B

ii) If $Q < 0$ $\Rightarrow$ Negative Association between attributes A & B

iii) If $Q = 0$ $\Rightarrow$ There is no association between attributes A & B

or

The attributes A & B are independent.

iv) If $Q = 1$ $\Rightarrow$ The attributes A & B are completely associated

v) If $Q = -1$ $\Rightarrow$ The attributes A & B are completely dissociated.

Result: The coefficient of association Q_{AB} lies between -1 and 1

i.e $-1 \leq Q_{AB} \leq 1$

proof:

$$Q_{AB} = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)} \quad \text{--- (1)}$$

let $(AB)(\alpha\beta) = x$ &

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

hence,

$$Q_{AB} = \frac{x-y}{x+y} \quad \text{--- } (*)$$

since $x \geq 0$, $y \geq 0$

$-y \leq y$

Adding x on both sides.

$$x-y \le x+y$$

$$\frac{x-y}{x+y} \le 1 \quad \text{——} \quad (2)$$

Also, $-x \le x$

subtract y from both sides

$$-x-y \le x-y$$

$$-(x+y) \le x-y$$

$$-1 \le \frac{x-y}{x+y} \quad \text{——} \quad (3)$$

Combining equation (2) & (3)

we get

$$-1 \le \frac{x-y}{x+y} \le 1$$

i.e. $\boxed{-1 \le Q_{AB} \le 1}$

Ex (1) If $N = 100$, $(A) = 47$, $(B) = 62$, $(AB) = 32$ find the coefficient of association between A and B and interpret

Solution:

Given: $N = 100$, $(A) = 47$, $(B) = 62$, $(AB) = 32$

	B	β	
A	$(AB) = 32$	$(A\beta) = 15$	$(A) = 47$
α	$(\alpha B) = 30$	$(\alpha\beta) = 23$	$(\alpha) = 53$
	$(B) = 62$	$(\beta) = 38$	$N = 100$

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

$$= \frac{32 \times 23 - 15 \times 30}{32 \times 23 + 15 \times 30}$$

$$= \frac{736 - 450}{736 + 450}$$

$$= \frac{286}{1186}$$

$$Q_{AB} = 0.2411$$

Since, $Q_{AB} > 0$

The attributes A and B are positively associated or associated.

(2) The following is the information on employment and education

Employed graduates = 286
Unemployed graduates = 48
Employed undergraduate = 450
Unemployed undergraduates = 216

Compute the coefficient of association between employment and education and comment on it.

Solution :

A : Employment α : Unemployment
B : Graduate β : Undergraduate

$(AB) = 286$, $(\alpha B) = 48$
$(A\beta) = 450$, $(\alpha\beta) = 216$

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

$$= \frac{286 \times 216 - 450 \times 48}{286 \times 216 + 450 \times 48}$$

$$= \frac{61776 - 21600}{61776 + 21600} = \frac{40176}{83376}$$

$Q_{AB} = 0.4818 > 0$

Hence, employment & education are positively associated.

Result :- $\quad a_{AB} = a_{\alpha\beta} = -a_{A\beta} = -a_{\alpha B}$