UNIT 13 CLASSIFICATION OF ATTRIBUTES

Structure

13.1 INTRODUCTION

A characteristic that varies from one person or thing to another is called a variable. Income, height, weight, blindness, honesty, sickness are a few examples of variables among humans. The first three of these variables yield numerical information and an investigator can measure their actual magnitude. They are thus termed as quantitative variables. The last three yield non-numerical information and an investigator may not be able to measure their actual magnitude numerically. Hence they are called qualitative variables. They can be arranged in order or in rank. Sometimes they are also referred to as categorical variables.

Attributes are Qualitative Variables and cannot be measured directly as they do not yield any numerical information. They can be arranged in order or in rank

In the earlier blocks you must have gone through various statistical methods viz. measures of central tendency, measures of dispersion, skewness, correlation, etc. These are important statistical methods that can be used for analysis of the data, which is quantitative in nature i.e. for the first case. However, in the second case when we are dealing with the qualitative variables, which are usually referred to as attributes, the aforesaid methods cannot be used as such, as we cannot measure these qualitative characteristics numerically. All that we can do is to count the number of persons possessing a particular attribute or quality or the number of persons not possessing a particular attribute or quality. Different statistical treatment is required to numerically measure qualitative characteristics. However, they can be related in an indirect manner to a numerical data after assigning particular quantitative indicator. For example, the presence of an attribute can be represented by numeral 1 and the absence of an attribute by numeral 0. Thus, methods of statistics dealing with quantitative variables can also be used for analysing and interpreting the qualitative variables, i.e. attributes. But to have a clear insight on available data through study, analysis and interpretation of attributes there are independently developed statistical methods in the theory of attributes. This unit forms the base for understanding the theory of attributes.

In Section 13.2 the basic notations regarding the presence and absence of the attributes are explained. The dichotomy of data is defined in Sub-section 13.2.2. Concepts of classes and class frequencies are described in Section 13.4 whereas order of classes and class frequencies are elaborated in Section 13.5.

The relation between the class frequencies is described in Section 13.6 and class symbols as operators are defined in Section 13.7.

Objectives

After studying this unit, you will be able to

- describe the difference between quantitative and qualitative variables;
- explore the notations and terminology used in the classification of attributes;
- describe the classes and class frequencies;
- define the order of classes and class frequencies;
- explain the basic concepts of relation between the class frequencies; and
- describe the class symbols as operators.

13.2 NOTATIONS

Capital A, B and C denote the presence of attributes and Greek letters α, β and γ denote the absence of attributes.

For convenience in analysis it is necessary to use certain symbols for different classes and for the number of observations assigned to each class. Usually the capital letters A, B, C, etc. are used to denote the presence of attributes and Greek letters α, β, γ, etc. are used to denote the absence of these attributes respectively. Thus, if A represents the attribute of being wealthy, α would represent the attribute of being poor i.e. not wealthy. If B represents blindness, β, would represent absence of blindness. The two classes viz. A (presence of attribute) and α (absence of attribute) are called complementary classes and the attribute α is called complimentary attribute to A. Similar interpretation is for β, γ i.e. they are complementary attributes to B and C respectively. Two or more attributes present in an individual or individuals are indicated by combination of capital Latin letters AB, AC, BC, ABC, etc. Also the presence of one attribute and the absence of other will be represented by $A\beta$, αB, $AB\gamma$.

$A\beta$ represents presence of A and absence of B in an individual. Similarly, the combination $AB\gamma$ represents the absence of the attribute C and the presence of A and B. Likewise any combination of letters can be interpreted.

For example, if A denotes attribute of being wealthy and B denotes the attribute of honesty, then

AB stands for wealthy and honest

$A\beta$ stands for wealthy and dishonest

αB stands for poor and honest

$\alpha\beta$ stands for poor and dishonest.

If a third attribute is noted e.g. sickness then ABC includes those who are wealthy, honest and sick. $A\beta C$ stands for those individuals who are wealthy, dishonest and sick. Similar interpretations can be given to other combination of letters.

13.2.1 Dichotomy of Data

Dichotomous classification has only 2 sub-classes

When the individuals are divided only into two sub-classes or complementary classes and no more, with respect to each of the attributes A, B, C, etc., it is called dichotomous classification. If an attribute has many classes it is called manifold classification.

13.3 CLASSES AND CLASS FREQUENCIES

Different attributes are themselves called classes. For example, if only one attribute, say, tallness is being studied the population would be divided into two classes-one consisting of those possessing this attribute and the other consisting of persons not possessing this attribute i.e. one class would be of tall persons and the other class would be of dwarfs. If more than one attribute were taken into account then the number of classes would be more than two. If, for example, the attribute of sex were also studied along with tallness, then there would be more than two classes in which the total population would be divided. There would be "tall", "dwarf", "male", "female", "tall female", "tall male", "dwarf male" and "dwarf female".

The number of observations assigned to any class is termed as class frequency. Putting a letter or letters within brackets generally denotes the class frequencies. Thus, (A) stands for the frequency of the attribute A. $(A\beta)$ denotes the number possessing the attribute A but not B. Class frequencies (A) (AB) (ABC), etc. are known as positive frequencies and class frequencies of type (α) $(\alpha\beta)$ $(\alpha\beta\gamma)$ are called negative frequencies whereas the frequencies of the type (αB) $(A\beta)$ $(A\beta C)$, etc. are called contrary frequencies.

13.4 ORDER OF CLASSES AND CLASS FREQUENCIES

The total of all frequencies denoted by N is called the class of zero order. The classes A, α, B, β, etc. are called the class of first order whereas the combination of any two letters showing the presence or absence of attributes are called class of second order e.g. AB, $A\beta$, αB, $\alpha\beta$, etc. Similarly, combinations like ABC, $AB\gamma$, $A\beta\gamma$, $\alpha\beta\gamma$, etc. are known as class of third order and so on. The frequencies of these classes are known as frequencies of zero order, first order, second order, third order respectively.

13.4.1 Total Number of Class Frequencies

Only one class with N number of members is known as a class of order 0. In a class of order 1 there are 2n number of classes, because there are n attributes, each of them contributing two symbols i.e. one of type A and other of type α.

By binomial theorem

$(1+2)^n = 1 + {}^nC_1\, 2 + {}^nC_2\, 2^2 + \ldots + {}^nC_n 2^n$

Similarly, a class of order 2 has ${}^nC_2 \times 2^2$ classes. Since each class contains two symbols, two attributes can be chosen from n in ${}^nC_2$ ways, and each pair give rise to 2^2 different frequencies of the types (AB), $(A\beta)$, (αB) and $(\alpha\beta)$.

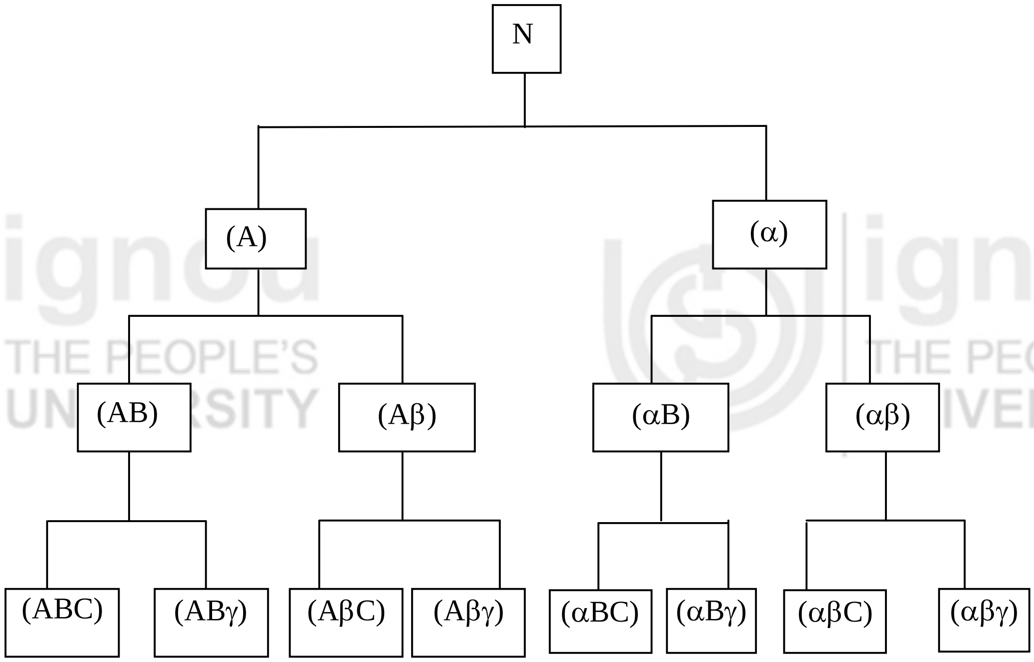
In the same way, it can be shown that total number of classes of order r, there are ${}^nC_r \times 2^r$ classes. Thus, total number of class frequencies of all orders, for n attributes

$$= 1 + {}^nC_1\, 2 + {}^nC_2\, 2^2 + \cdots + {}^nC_r\, 2^r + \cdots + {}^nC_n\, 2^n$$

$$= (1+2)^n = 3^n$$

13.5 RELATION BETWEEN CLASS FREQUENCIES

The frequencies of lower order class can always be expressed in terms of higher order class frequencies. For three factors A, B and C, all possible twenty seven combinations of attributes belonging to different classes in the form of a pedigree can be displayed in the following manner.



Similar relations can be given by taking N, B, γ and N, C, γ.

From the above relations we find that

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

$$(AB) = (ABC) + (AB\gamma); \ (A\beta) = (A\beta C) + (A\beta\gamma)$$

$$(\alpha B) = (\alpha BC) + (\alpha B\gamma); \ (\alpha\beta) = (\alpha\beta C) + (\alpha\beta\gamma)$$

$$N = (A) + (\alpha) = (B) + (\beta) = (C) + (\gamma)$$

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

$$N = (ABC) + (AB\gamma) + (A\beta C) + (A\beta\gamma) + (\alpha BC) + (\alpha B\gamma) + (\alpha\beta C) + (\alpha\beta\gamma)$$

Similarly, other relations can be given. The classes of higher order are known as ultimate classes and their frequencies as the ultimate class frequencies. In case of n attributes the ultimate class frequencies will be of n order. For example, if there are three attributes A, B and C the ultimate frequencies will be (ABC), (ABγ), (AβC), (Aβγ), (αBC), (αBγ), (αβC) and (αβγ).

13.6 CLASS SYMBOLS AS OPERATORS

Symbol AN is taken for the operation of dichotomizing N according to the attribute A and is written as AN = (A)
Similarly, we can write $\alpha N = (\alpha)$

Adding these two expressions we get

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

$\Rightarrow$ N (A+α) = N

$\therefore$ A + α = 1

Thus, A can be replaced by (1–α) and α can be replaced by (1–A) respectively. Similarly, B can be replaced by (1–β) and β by (1–B) and so on.

Similarly, we may write ABN = (AB) or αβN = (αβ)

Thus, (αβ) = (1–A) (1–B) N

= (1–A– B + AB) N

= N – (A) – (B) + (AB)

Again

(αβγ) = (1–A) (1–B) (1–C) N

= (1– A – B – C + AB + BC + AC – ABC) N

= N – (A) – (B) – (C) + (AB) + (BC) + (AC) – (ABC)

Let us consider some problems.

Example 1: Given that (AB) = 150, (Aβ) = 230, (αB) = 260, (αβ) = 2340. Find other frequencies and the value of N.

Solution: We have

(A) = (AB) + (Aβ) = 150 + 230 =380

(α) = (αB) + (αβ) =260 + 2340 = 2600

(B) = (AB) + (αB) =150 +260 = 410

(β) = (Aβ) + (αβ) = 230 + 2340 = 2570

N = (A) + (α) = 380 + 2600 = 2980

N = (B) + (β) = 410 + 2570 = 2980

(A) = (AB) + (Aβ)

(α) = (αB) + (αβ)

(B) = (AB) + (αB)

(β) = (Aβ) + (αβ)

(AB) = (ABC) + (ABγ);

(Aβ) = (AβC) + (Aβγ)

(αB) = (αBC) + (αBγ);

(αβ) = (αβC) + (αβγ)

N = (A) + (α) = (B) + (β) = (C) + (γ)

Example 2: A number of school children were examined for the presence or absence of certain defects of which three chief descriptions were noted. Let A development defects; B nerve sign; C low nutrition. Given the following ultimate frequencies, find the frequencies of the class defined by the presence of the defects.

(ABC) = 60, (αBC) = 75, (ABγ) = 250, (αBγ) = 650,

(AβC)= 80, (αβC) = 55, (Aβγ) = 350, (αβγ) = 8200

Solution: We have

(A)= (AB) + (Aβ)

= (ABC) + (ABγ) + (AβC) + (Aβγ)

= 60 + 250 + 80 + 350

= 740

Similarly,

(A)= (AB) + (αB)

= (ABC) + (ABγ) + (αBC) + (αBγ)

$$= 60+250+75+650$$
$$= 1035$$
$$(B) = (AC) + (\alpha C)$$
$$= (ABC) + (A\beta C) + (\alpha BC) + (\alpha\beta C)$$
$$= 60 + 80 + 75 + 55$$
$$= 270$$

Again

$$(AB) = (ABC) + (AB\gamma) = 60+250 = 310$$
$$(AC) = (ABC) + (A\beta C) = 60 + 80 = 140$$
$$(BC) = (ABC) + (\alpha BC) = 60 + 75 = 135$$
$$N = (ABC) + (AB\gamma) + (A\beta C) + (A\beta\gamma) + (\alpha BC) + (\alpha B\gamma) + (\alpha\beta C) + (\alpha\beta\gamma)$$
$$= 60 + 250 + 80 + 350 + 75 + 650 + 55 + 8200$$
$$= 9720$$

Example 3: Measurements are made on a thousand husbands and a thousand wives. If the measurement of husbands exceeds the measurement of wives in 600 cases for one measurement, in 500 cases for another and in 480 cases for both measurements, then in how many cases would both measurements on wives exceed the measurement on husbands?

Solution: Let (A) denotes the husbands exceeding wives in first measurement and (B) denotes husbands exceeding wives in second measurement. Then given N = 1000, (A) = 600, (B) = 500, (AB) = 480

We have to find $(\alpha\beta)$

$$(\alpha\beta) = N - (A) - (B) + (AB)$$
$$= 1000 - 600 - 500 + 480$$
$$= 380$$

Now, let us solve following exercise:

E1) Given, the following frequencies of the positive classes. Find the frequencies of the rest of the classes:

(A) = 975, (AB) = 455, (ABC) = 125, (B) = 1,187, (AC) = 290,

N= 12,000, (C) = 585 and (BC) = 250

E2) In an observation of 100 cases it was found that the number of unmarried students was 40, number failing the examination was 55 and the number of married who failed was 30. From the information given above find out:

1. The number of married students,
2. The number of students passing the examination,
3. The number of married students who passed,
4. The number of unmarried students who passed,
5. The number of unmarried students who failed.

E3) In a Girls' High school there were 200 students. Their results in the quarterly, half yearly and annual examinations were as follows

85 passed the quarterly examination.

80 passed the half yearly examination.

94 passed the annual examination.

28 passed all the three and 40 failed all the three.

25 passed the first two and failed in the annual examination.

43 failed the first two but passed the annual examination.

Find how many students passed the annual examination.

13.7 SUMMARY

In this unit, we have discussed:

1. Variables can be classified as quantitative and qualitative. The magnitude of quantitative variables can be measured whereas qualitative variables are non, numerical in nature and their magnitude cannot be measured numerically. They are called attributes. The qualitative data can be quantified by assigning number 1 to a person possessing the particular attribute and number 0 to a person not possessing that attribute. Thus, total number of ones would denote total number of persons possessing that attribute. Similarly, total number of zeros would denote total number of persons not possessing that attribute;

2. The statistical methods used to study the qualitative variables are different from the methods to study quantitative variables;

3. The data for the analysis of qualitative variables can be dichotomous where the individuals are classified only into two sub-classes or complementary classes with respect to each attribute A, B, C, etc. On the other hand if the individuals are classified into many classes on the basis of an attribute, the classification is manifold;

4. The number of observations belonging to each class is called class frequency of that class. Putting the letter or letters within brackets generally denotes the class frequencies;

5. The total number of observation denoted by N is called the class of zero order. The class denoted by single letter e.g. A, B, C, α, β, γ, etc. are called class of first order, whereas class represented by combination of two letters e.g. AB, Aβ, αB, etc. are classes of second order. Similarly combinations of three letters indicating presence or absence of attributes are class of third order and so on. The frequencies of these classes are known as frequencies of zero, first, second and third order respectively; and

6. For order r, there are ${}^{n}C_{r} \times 2^{r}$ classes. If there were n attributes, total number of class frequencies would be 3^{n}. The frequencies of lower order class can be expressed in terms of higher order class frequencies.

13.8 SOLUTIONS /ANSWERS

E1) We have

$(\alpha) = N - (A) = 12{,}000 - 975 = 11025$

$(\gamma) = N - (C) = 12000 - 585 = 11415$

$(\beta) = N - (B) = 12000 - 1187 = 10815$

$(AB\gamma) = (AB) - (ABC) = 455 - 125 = 330$

$(A\beta C) = (AC) - (ABC) = 290 - 125 = 165$

$(A\beta\gamma) = (A) - (AB) - (AC) + (ABC)$

$= 975 - 455 - 290 + 125 = 355$

$(\alpha BC) = (BC) - (ABC)$

$= 250 - 125 = 125$

$(\alpha B\gamma) = (B) - (AB) - (BC) + (ABC)$

$= 1187 - 455 - 250 + 125 = 607$

$(\alpha\beta C) = (C) - (AC) - (BC) + (ABC)$

$= 585 - 290 - 250 + 125 = 170$

$(\alpha\beta\gamma) = N - (A) - (B) - (C) + (AB) + (AC) + (BC) - (ABC)$

$= 12000 - 975 - 1187 - 585 + 455 + 290 + 250 - 125 = 10123$

$(A\beta) = (A\beta C) + (A\beta\gamma)$

$= 165 + 355 = 520$

$(\alpha B) = (\alpha BC) + (\alpha B\gamma)$

$= 125 + 607 = 732$

$(\alpha\beta) = (\alpha\beta C) + (\alpha\beta\gamma)$

$= 170 + 10123 = 10293$

$(A\gamma) = (A) - (AC)$

$= 975 - 290 = 685$

$(\alpha C) = (C) - (AC)$

$= 585 - 290 = 295$

$(\alpha\gamma) = (\alpha) - (\alpha C)$

$= 11025 - 295 = 10730$

$(B\gamma) = (B) - (BC)$

$= 1187 - 250 = 937$

$(\beta C) = (C) - (BC)$

$= 585 - 250 = 335$

$(\beta\gamma) = (\beta) - (\beta C)$

$= 10815 - 335 = 10480$

E2) Let A represents married

Then, α represents unmarried

Let B represents passed

Then, β represents failed

The given data are

$N = 100, (\alpha) = 40, (\beta) = 55, (A\beta) = 30$

(i) Number of married students

$(A) = N - (\alpha) = 100 - 40 = 60$

(ii) Number of students passing the examination

$(B) = N - (\beta) = 100 - 55 = 45$

(iii) Number of married students who passed

$(AB) = (A) - (A\beta) = 60 - 30 = 30$

(iv) Number of unmarried students who passed

$(\alpha B) = (B) - (AB) = 45 - 30 = 15$

(v) Number of unmarried students who failed

$(\alpha\beta) = (\alpha) - (\alpha B) = 40 - 15 = 25$

E3) Let us denote

Success in quarterly examination by A and failure by α

Success in half yearly examination by B and failure by β

Success in annual examination by C and failure by γ

Thus, we have data in the question as

$N = 200, (A) = 85, (B) = 80, (C) = 94, (ABC) = 28,$

$(\alpha\beta\gamma) = 40, (AB\gamma) = 25, (\alpha\beta C) = 43$

Now we have to find the value of

$(\alpha BC) + (A\beta C) + (AB\gamma) + (ABC)$

We have the relation

$(\alpha BC) + (A\beta C) + (ABC) + (\alpha\beta C) = (C)$

$\therefore\ (\alpha BC) + (A\beta C) = (C) - (ABC) - (\alpha\beta C)$

$= 94 - 28 - 43 = 23$

$\therefore\ (\alpha BC) + (A\beta C) + (AB\gamma) + (ABC)$

$= 23 + 25 + 28 = 76$

Thus, the number of students who passed at least two examinations is 76.

GLOSSARY

Attribute	:	A qualitative measurement assigned to objects or individuals.
Dichotomy	:	A sharp division into two opposed groups or classes.
Frequency	:	The number of observations in some given category.
Manifold classification	:	Numerous or various classes.
Variable	:	A single one-dimensional property, which may vary along a scale.