1. Introduction and Attributes

Welcome to the series of E-learning modules on independence and association of attributes.

At the end of this session, you will be able to:

- Explain the meaning of Attributes
- Explain the classification of attributes
- Explain positive and negative classes
- Explain the consistence of the data
- Explain the meaning and Criteria for independence
- Explain the meaning of association
- Explain complete association and disassociation
- Explain partial association
- Explain illusory association

First let us see what we mean by attributes:
Literally, an attribute means a quality or characteristic. Theory of attributes deals with qualitative characteristics which are not amenable to quantitative measurements and hence need slightly different statistical treatment from that of the variables.

Examples of attributes are drinking, smoking, blindness, health, honest, etc.

As per the analysis of statistics relating to attributes, the first thing is the classification of data. Here data are classified on the basis of presence or absence of particular attributes. It may be pointed out that the methods of statistical analysis applicable to study of variables can also be used to a great extent in the theory of attributes and vice-versa.

For example, the presence or absence of an attributes may be regarded as changes in the values of a variable which can possess only two values namely 0 and 1.
For the sake of clarity and simplicity, the theory of attributes has been developed independently.

If only one attribute, say blindness, is being studied the population would be divided in two classes – one consisting of those people in whom this attribute is present and the other consisting of those in whom this attribute is not present. Thus, one class would be of the "blinds" and the other of "non-blinds". If more than one attribute are taken into account the number of classes would be more than two.

For example: If the attribute of deafness is also studies, there would be a number of classes in which the universe would be divided. There would be "blind", "non-blinds", "deaf", "non-deaf", "blind and not deaf", "blind and deaf", "not-blind and deaf" and "not-blind and not-deaf".

The following points should be noted about the classification of data according to attributes.

1. Classification is arbitrary and vague: It should be clearly understood that when the universe is divided in, say, two classes, "blind" and "not blind", there is no clear cut line of demarcation between them. It is very difficult to lay down such a definition of the

word blindness which may give two clear-cut classes. In practice, the attribute of blindness gradually transforms into attribute of sight and there are many cases on the border line.

The boundary in such cases is very vague and uncertain. In some cases, the line of demarcation may be arbitrary
For example: when people are classified as tall and short, an arbitrary figure is the dividing line. People above a certain height are called tall and below it as short. In all types of analysis relating to statistics of attributes this point should always be kept in mind.

2. Classification by dichotomy: If only one attribute is being studied the universe is divided into two parts – one in which the attribute is present and the other in which it is not present. These classes are mutually exclusive.

Such classification where the universe is divided into two parts is called 'classification by dichotomy'. In actual analysis usually there are more than two classes in which the universe is divided and such classification is called manifold classification

Now let us discuss about the notations and terminology.

For the sake of convenience in analysis it is necessary to use certain symbols to represent different classes and their frequencies. Usually capital letters A, B and C etc., are used to denote the presence of attributes and the Greek letters alpha, beta and gamma etc., are used to denote absence of these attributes respectively. Thus if A represents the attribute of blindness alpha would represent absence of blindness. If B represents deafness, beta would represent absence of deafness and if C represents insanity gamma would represent absence of insanity.

The number of units possessing a particular attribute represented by A would be termed as belonging to class A, and similarly those in whom this attribute is absent would be termed as belonging to class alpha.

If two attributes are being studied their combination can be represented by the combination of the letters representing the two attributes. Thus, if blindness represented by A and deafness by B then AB would represent blindness and deafness; A beta would represent blindness and absence of deafness; alpha B would represent absence of blindness and presence of deafness and alpha beta would represent absence of blindness and absence of deafness.

2. Classes

The number of units in different classes is called “class frequencies”. Thus, if number of blind and deaf people is 20, then the frequency of class AB is 20. Class frequencies are denoted by enclosing the class symbols brackets. Thus AB would represent the frequency of the class AB.

If there is one attribute represented by A the total number of classes is 3; they would be A, alpha, and N. If the number of attributes is two, represented by A and B the total number of classes (including N) would be 9. They would be N, A, B, alpha, beta, AB, A beta, alpha B and alpha beta.
If number of attributes is three, the total number of classes (including N) would be 27. The total number of classes is always equal to 3 power n where n stands for the number of attributes.

The various classes and their frequencies are demarcated on the basis of an order. Thus,

- N is a class of zero order
- A, B, A, beta are classes of the first order
- AB, alpha beta, A beta, Alpha B are classes of second order

Similarly, the frequencies of these classes are also known as frequencies of the Zero, First or Second order. If there are only two attributes under study then the Second order classes and frequencies are called the classes or the frequencies of the ultimate order. Since these are the last set of classes and frequencies. The number of classes of the ultimate order is equal to 2 power n, where n stands for the number of attributes.

The classes which represent the presence of an attribute or attributes are called positive classes. Thus

- N, A, B and AB are positive classes

The classes which represent the absence of an attribute or attributes is called negative classes. Thus

- alpha and beta are negative classes

The classes in which one attribute is present and the other is absent are called pairs of contraries. Thus

- A beta and alpha B are pairs of contrary

Now let us discuss about the relationship between classes of various order.

In classifying statistical data according to attributes, the following simple rule should be kept in mind. Any class frequency can always be expressed in terms of class frequencies of higher order. Thus, the frequencies of first order can be expressed in terms of the frequencies of the second order which in turn can be expressed in terms of frequencies of the third order and so on.

On the basis of this rule we can set up various types of relationships between the frequencies of different orders. If there is one attribute only, represented by A, the frequency of the universe N can be divided into classes (A) and alpha.

Thus, N is equal to A plus alpha

If one more attribute B is taken into account the first order classes that is A and alpha can each be divided into two classes –

1. One in which attribute B is present
2. Other in which is not present

Thus,

A is equal to AB plus A beta

Alpha is equal to alpha B plus alpha beta

Similarly,

B is equal to AB plus alpha B

Beta is equal to A beta plus alpha beta

3. Consistence of the Data

Now let us discuss about the consistence of the data.

In statistics of attributes when frequencies of various classes are counted, no class frequency can be negative.

If any class frequency is negative the data are said to be inconsistent. Such inconsistency may be due to wrong counting or inaccurate additions or subtractions or may be the result of misprints.

In order to test whether a set of figures is consistent, various class frequencies should be found and if no class frequency is negative apparently the data are consistent.

It should be remembered that consistence of data is no proof of accurate counting or the correct calculation though the inconsistence of data proves that there is either a mistake or misprint in figures.

The easiest way to find out whether the data are consistent is to obtain ultimate class frequencies because if there is any inconsistence, one or more ultimate class frequencies would be negative. It is also possible to lay down rules for testing consistence of data. For example, the value of (A) cannot be more than N otherwise (alpha) would have negative value. Similarly, (A) cannot be less than 0; otherwise (A) itself would have a negative value.

The above rules relating to consistency of data are also used to fill the gaps if the data are incomplete. With the help of these rules it is possible to lay down the minimum and maximum limits of a particular class frequency.

Now let us discuss about the independence of attributes.

Two attributes A and B are said to be independent if there exists no relationship of any kind between them. If A and B are independent we would expect:

1. The same proportion A's amongst B's as amongst beta's
2. The proportion of B's amongst A's is same as that amongst the alpha's

For example: If insanity and deafness are independent, the proportion of the insane people among deaf and non-deaf must be same.

Following is the fundamental rule for the independence of two attributes A and B.
"If the attributes A and B are independent, the proportion of AB's in the population is equal to the product of the proportions of A's and B's in the population.

4. Association (Part 1)

Now let us discuss what is meant by association of attributes.

Association of attributes refers to such techniques by which we can measure the relationship between two such phenomena whose size cannot be measured and where we can only find out the presence or absence of an attribute.
Just as in case of correlation analysis we study the relationship between two variables, similarly in case of association we study the relationship between two attributes. As it has been pointed out earlier, an attribute divides the universe only in 2 classes –

1. One possessing the attribute
2. Other not possessing the attribute

In variables there can be any number of classes in which the universe can be divided.

When we say that there is association between two attributes, it may be, complete association or disassociation or partial association or illusory association. Hence we study different types of associations one by one.

Consider complete association and disassociation.

There is complete association (or in other words perfect positive association) between two attributes A and B if all A's are B's and all B's are A's in which case (A) is equal to (B).

If (A) and (B) are not equal in number then there would be complete association between them if either all A's are B's or all B's are A's depending on whether the number of A's is less than the number of B's or vice versa.

Thus, if the number of A's is 40 and of B's is 60 then A and B would be completely associated if (AB) is equal to 40.

In the same way there would be complete disassociation or in other words perfect negative association between A and B if none of the A is B and none of the alpha is beta.

Thus, if the number of A's is 40 and the number of B's is 60, A and B would be completely disassociated if (AB) is equal to zero. In such case (alpha beta) would also be zero.

5. Association (Part 2)

Now let us consider the partial association.

The association between A and B may not be direct association; it may be the result of their association with a third attribute C. Thus, if A is positively associated with C and if B is also associated with C, A may be found to be positively associated with B.

This association between A and B is not direct. It is the effect of their association with another attribute C. To find out whether the association between A and B is real, and not merely due to their association with a third attribute C, it would be necessary to study the association of A and B in the sub-population of C and gamma.

If A and B are associated in both the sub-population of C and Gamma it would indicate that A and B are really associated with each other. The associations of A and B in the sub-populations are called partial association to distinguish them from total association in the universe as a whole.

An illustration would make the point clear.
Suppose that an association is observed between “B.C.G vaccination” and “exemption from Tuberculosis”, which means that a larger number of those people who are vaccinated are exempt from an attack of tuberculosis than those who are not vaccinated by BCG. It may be argued that this association is not real, it is illusory.

It may be due to their common association with some third attribute, say, economic condition of the people. Most of the vaccinated people may be those who are rich and well-to-do people and most of the unvaccinated people may be poor. Well-to-do people generally live in healthy surroundings and take rich diet and as such may be less exposed to the attack of tuberculosis.

On the other hand poor people generally live in unhygienic surroundings and take poor diet and as such are more liable to attack of tuberculosis.

Thus, the association between vaccination and exemption from tuberculosis may not be direct one.

It may be due to the fact that the attribute of richness, hygienic condition and good diet and exemption from tuberculosis is also associated with these conditions.

In this way vaccination is associated with exemption of tuberculosis due to their common association with richness, hygienic conditions and good diet.

In actual practise the ambiguity referred to above is of a more complex nature because the population under discussion may contain not only those units which possess a third attribute alone but a mixture of units with and without it.

For example: In the above case richness and hygienic conditions may be present side by side with poverty and unhygienic conditions at the place where the observations are being made.

However, the study is made only amongst those units where the third attribute is either present or not present; this type of ambiguity would not be there.

Thus, if the data in the previous illustration refer to only such a group of persons all of whom are rich and live in hygienic conditions and take rich diet or to such a group of persons all of whom are poor and live in unhygienic conditions and take poor diet, there would not be any complication in the study of association between A and B.

In order to be sure about the associations of A and B, generally their association in the sub-population C and γ are separately studied and then only conclusions are drawn about the association of A and B in the universe at large. As it has been said earlier, the association of A and B in the sub-populations of C and γ are called partial association and the association of A and B in the universe as a whole, total association.

Finally let us consider the illusory association.

It is clear from the previous discussion about partial associations that sometimes misleading or illusory associations may be observed between tow attributes which are not directly associated but which are both individually associated with third attribute. Thus if A and B are two independent attributes but both of them are positively associated with a third attribute C, it would appear as if A and B are directly associated.

If A is positively associated with C, and B is negatively associated with C, the association between A and B would appear to be negative.

Misleading conclusions may be arrived at if the partial associations are not studied.

Here's a summary of our learning in this session:

- Meaning of Attributes
- Classification of attributes
- Positive and negative classes
- Consistence of the data
- Meaning and Criteria for independence
- Meaning of association
- Complete association and disassociation
- Partial association
- Illusory association