THEORY OF ATTRIBUTES

By:

Dr. Akash Asthana

Assistant Professor,

University of Lucknow,

Lucknow

ATTRIBUTES

- The attribute means the quality or characteristic of some individual. Individual may be a person, place or thing.
- The data for which any standard measuring device is not available is known as qualitative data. Such type of data is generally used to measure quality of persons.
- Qualitative data can’t be measured in terms of quantities or numeric values.
- Qualitative data can be measured in terms of presence or absence of certain characteristic or level of the quality present.
- In statistics the term attribute is used to represent the data which is measured in terms of some quality or qualities.

TYPES OF QUALITATIVE DATA

Qualitative data can be classified into three groups:

1. **Binary Data:**

If any quality is measured in terms of its presence or absence then it is termed as Binary data. In other words if an attribute posses only two category then it is called as Binary data.

2. **Nominal Data:**

If an attribute posses more than two categories which cannot be arranged in any order then it is termed as Nominal data.

3. **Ordinal Data:**

If an attribute posses more than two categories which can be arranged in ascending/descending order then it is termed as Ordinal data.

TYPES OF ATTRIBUTES

Dichotomous:

An attribute is said to be dichotomously classified is it possessed only two categories. In other words, the binary type qualitative data is also termed as dichotomously classified.

Manifold:

An attribute is said to be manifold if it possess more than two categories. These categories may be unordered or may be ordered, i.e., The nominal and ordinal type qualitative data are termed as Manifold classified.

CONTINGENCY TABLES

- If the measurements are taken over two different attributes for same set of individuals then to represent the data contingency tables are used. In such tables rows represents the one attribute and column represents the other attribute.
- The values in the cells shows the frequencies called as cell (class) frequencies.
- If both the attributes are dichotomously classified then it is called as (2 X 2) contingency table.
- If at least one of the attribute is manifold then table will be termed as (m x n) contingency table

M X N CONTINGENCY TABLE

		Attribute B						Total
		1	2	.	j	.	m	
Attribute A	1	f_{11}	f_{12}	.	f_{1j}	.	f_{1m}	f_{1o}
	2	f_{21}	f_{22}	.	f_{2j}	.	f_{2m}	f_{2o}
	.	.	.	.	.	.	.	.
	i	f_{i1}	f_{i2}	.	f_{ij}	.	f_{im}	f_{io}
	.	.	.	.	.	.	.	.
	n	f_{n1}	f_{n2}	.	f_{nj}	.	f_{nm}	f_{no}
Total		f_{o1}	f_{o2}	.	f_{oj}	.	f_{om}	N

M X N CONTINGENCY TABLE (CONTD.)

In the above table:

f_{ij} = number of observations classified in i^{th} level of attribute A and j^{th} level of attribute B;

$i = 1, 2, 3, \ldots, n;$

$j = 1, 2, 3, \ldots, m.$

EXAMPLE

Educational Status	Sex	
	Male	Female
Illiterate	50	70
Primary	35	30
High School	75	65
Inter	65	50
Higher than Inter	35	20

This cell frequency shows no. of Illiterate Males

CLASS FREQUENCIES

- The categories of the attributes are termed as class and their frequencies are termed as class frequencies.
- For dichotomous attributes the presence of attribute is termed as the positive class and absence of the attribute is termed as the negative class.
- The positive classes are generally denoted by the capital Roman letters like, A, B, C, … and negative classes are generally denoted by the Greek letters α, β, γ, …
- The frequencies of these classes are represented by putting these letters in the braces (), like (A), (B), (C), (α), (β), (γ) …
- In same manner the higher order class frequencies can be represented, like (AB), (AC), (αβ), (αγ) etc.

CLASS FREQUENCIES

Ultimate Class frequencies:

The class-frequencies of highest order are called ultimate class-frequencies. The number of ultimate class frequencies in case of n dichotomously classified attributes would be 2^n.

For Ex: If 4 attribute would be considered then the no. of ultimate class frequency would be $2^4 = 16$.

The total number of class frequencies in case of n dichotomously classified attributes would be 3^n.

For Ex: If 4 attribute would be considered then the no. of class frequencies would be $3^4 = 81$.

ORDER OF CLASS FREQUENCIES

The number of attributes considered simultaneously define the order of the class frequencies. The number of class frequencies of order r in case of n dichotomously classified attributes is given by ${}^{n}C_{r}2^{r}$.

Let us consider the case of three dichotomously classified attributes:

Order	Frequencies	Number of frequencies
0	N (Total no. of Individuals)	$1 = {}^{3}C_{0}2^{0}$
1	(A), (α), (B), (β), (C), (γ)	$6 = {}^{3}C_{1}2^{1}$
2	(AB), (Aβ), (αB), (αβ), (AC), (Aγ), (αC), (αγ), (BC), (Bγ), (βC), (βγ)	$12 = {}^{3}C_{2}2^{2}$
3	(ABC), (ABγ), (AβC), (Aβγ) (αBC), (αBγ), (αβC), (αβγ)	$8 = {}^{3}C_{3}2^{3}$

RELATIONS BETWEEN CLASS FREQUENCIES

The Lower order class frequencies can be expressed as the sum of the higher order cell frequencies. Let us consider the case of three dichotomously classified attributes A, B and C.

N = (A) + (B) + (C) + (α) + (β) + (γ)	
(A) = (AB) + (Aβ),	(α) = (αB) + (αβ)
(B) = (AB) + (αB)	(β) = (Aβ) + (αβ)
(AB) = (ABC) + (ABγ)	(Aβ) = (AβC) + (Aβγ)
(αB) = (αBC) + (αBγ)	(αβ) = (αβC) + (αβγ)
(AC) = (ABC) + (AβC)	(A γ) = (ABγ) + (Aβγ)
(αC) = (αBC) + (αβC)	(α γ) = (αBγ) + (αβγ)
(BC) =(ABC) + (αBC)	(Bγ) = (ABγ) + (αBγ)
(βC) = (AβC) + (αβC)	(β γ) = (Aβγ) + (αβγ)

Further substituting the values of lower order frequencies in terms of higher order frequencies in expression of (A), (B) etc. more expressions can be obtained.

INDEPENDENCE/ASSOCIATION OF ATTRIBUTES

Two attributes A and B are said to be independent if the proportion of A's amongst B's is same as the proportion of A's among β's or the proportion of B's amongst A's is same as the proportion of B's among α's, i.e.,

$$\frac{(AB)}{(B)} = \frac{(A\beta)}{(\beta)} \quad \text{or} \quad \frac{(AB)}{(A)} = \frac{(\alpha B)}{(\alpha)} \quad \text{or} \quad (AB) = \frac{(A)(B)}{N}$$

Otherwise attributes are said to be associated.

Two attributes A and B are said to be positively associated if

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

Two attributes A and B are said to be negatively associated if

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

MEASURES OF ASSOCIATION

To measure the strength of relationship among two attributes the measures of associations are used. The main measures of the association are:

For Dichotomous classified attributes:

1. Yule's Coefficient
2. Coefficient of Colligation

For Manifold attributes:

1. Chi-Square coefficient
2. Karl Pearson's Coefficient
3. Tschuprow's Coefficient

MEASURES OF ASSOCIATION

For computing the coefficient of association for two dicotomously classified attributes A and B the data is represented in the form of following (2X2) contingency table:

Attribute A	**Attribute B**		**Total**
	B	β	
A	(AB)	(Aβ)	(A)
α	(αB)	(αβ)	(α)
Total	(B)	(β)	N

MEASURES OF ASSOCIATION

Yule's coefficient:

This measure was suggested by Mr. G. U. Yule. The measure is defined as:

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

The value of Q lies between –1 and +1.

A negative value of Q indicates the presence of negative association. Value of Q near to –1 shows the strong negative association, which indicates that the proportion of off diagonal frequencies are high enough than the proportion of diagonal frequencies.

Q = –1 indicates the perfect negative association. In this case the diagonal frequencies would be zero.

A positive value of Q indicates the presence of positive association. Value of Q near to +1 shows the strong positive association, which indicates that the proportion of diagonal frequencies are high enough than the proportion of off diagonal frequencies.

Q = +1 indicates the perfect positive association. In this case the off diagonal frequencies would be zero.

Q = 0 indicates that the attributes are independent.

MEASURES OF ASSOCIATION

Coefficient of Colligation:

This measure was also suggested by Mr. G. U. Yule. It is also termed as Yule's Y. The measure is defined as:

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

The value of Y lies between –1 and +1.

A negative value of Y indicates the presence of negative association. Value of Y near to –1 shows the strong negative association, which indicates that the proportion of off diagonal frequencies are high enough than the proportion of diagonal frequencies.

MEASURES OF ASSOCIATION

Coefficient of Colligation (Contd.):

Y = –1 indicates the perfect negative association. In this case the diagonal frequencies would be zero.

A positive value of Y indicates the presence of positive association. Value of Y near to +1 shows the strong positive association, which indicates that the proportion of diagonal frequencies are high enough than the proportion of off diagonal frequencies.

Y = +1 indicates the perfect positive association. In this case the off diagonal frequencies would be zero.

Y = 0 indicates that the attributes are independent.

Relation between Yule's Coefficient and Coefficient of colligation:

$$Q = \frac{2Y}{1+Y^2} \quad \text{or} \quad Y = \frac{1-\sqrt{1-Q^2}}{Q}$$

MEASURES OF ASSOCIATION

For computing the measure of association for manifold classified attributes the data is represented in the form of following (m x n) contingency table:

		Attribute B						Total
		1	2	.	j	.	m	
Attribute A	1	O_{11}	O_{12}	.	O_{1j}	.	O_{1m}	O_{1o}
	2	O_{21}	O_{22}	.	O_{2j}	.	O_{2m}	O_{2o}
	.	.	.	.	.	.	.	.
	i	O_{i1}	O_{i2}	.	O_{ij}	.	O_{im}	O_{io}
	.	.	.	.	.	.	.	.
	n	O_{n1}	O_{n2}	.	O_{nj}	.	O_{nm}	O_{no}
Total		O_{o1}	O_{o2}	.	O_{oj}	.	O_{om}	N

MEASURES OF ASSOCIATION

Chi –square measure of association:

The chi-square measure of association is defined by:

$$\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Where,

O_{ij} is called as observed frequency

E_{ij} is called as expected frequency defined by:

$E_{ij} = (O_{io} \text{ X } O_{oj})/N$

χ^2 ranges between 0 and ∞. Higher value of χ^2 shows the stronger association. The value of 0 indicates that the attributes are independent.

This measure has a drawback that it has no upper-bound.

MEASURES OF ASSOCIATION

Karl Pearson's measure of association:

To overcome from the drawback of χ^2 association Karl Pearson defines another measure which is given by the formula:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

The value of this measure lies between 0 and 1. Higher value of the measure indicate the stronger association. The value 0 indicates that both the attributes are independent.

This measure has a drawback that it never attains its upper bound, i.e., even if both the attributes are perfectly associated its values comes out to be smaller than 1.

MEASURES OF ASSOCIATION

Tschuprow's measure of association:

To overcome from the drawback of Karl Pearson's measure of association Tschuprow defines another measure which is given by the formula:

$$T = \left\{ \frac{\chi^2}{N\sqrt{(m-1)(n-1)}} \right\}^{1/2}$$

The value of this measure lies between 0 and 1. Higher value of the measure indicate the stronger association. The value 0 indicates that both the attributes are independent. The value of 1 indicates that attributes are perfectly associated.

REFERENCES

1. Kothari CR, Research Methodology: Methods and Techniques, 2nd Edition, New Age International (P) Ltd, New Delhi.
2. Malhotra NK, Birks DF, Marketing Research an Applied Approach, 4th Edition, Prentice Hall, New Delhi.
3. A. M. Goon, M. K. Gupta and B. Dasgupta, Fundamentals of Statistics Vol.1, 2008, World Press Organization (P) Ltd, India.
4. A.K. Sharma, Text Book of Elementary Statistics, 2005, Discovery Publishing House, New Delhi.
5. https://conjointly.com/kb/probability-sampling/

Note: The example consider in the present notes are taken from the reference books mentioned at serial no. 3 and 4.