Frequently Asked Questions

1. Write the different methods of studying association between two attributes.

Answer:

The study of association can be done by any of the following methods.

1. Comparison of actual and observed frequencies
2. Comparison of various proportions and products.
3. Calculation of Yule's Coefficient of association
4. Calculation of coefficient of Collignation
5. Calculation of coefficient of Contingency.

2. Explain the method of comparison of actual and observed frequencies to find the association of two attributes.

Answer:

Whenever we want to study the association between two attributes A and B we try to find out whether attribute A is more commonly found with attribute B than in ordinarily expected. Thus in a study of association the first thing to be calculated is expected value of (AB). This value is calculated on the basis of simple rules of probability.

Thus if two attributes A and B are studied in a universe and if the frequency of A is represented by (A) and of B by (B), then

The probability of (A) = (A)/N and

The probability of (B) =(B)/N

The combined probability of two independent events is equal to the product of their individual probabilities. Thus the combined probability of (A) and (B) would be (A)/N X (B)/N and the expectation is obtained by multiplying the probability by N.

Hence the expectation of (A) and (B) combined would be, (A)X(B)/N.

From the above it is clear that ordinarily if attributes A and B are independent the expected frequency of (AB) would be equal to (A)X(B)/N.

3. Write the criteria for independence of two attributes in the method of comparison of actual and observed frequencies

Answer:

Two attributes A and B are said to be independent if the observed frequency of (AB) is equal to its expected frequency ie., (A)X(B)/N.

4. Write the limitation of method of comparison of actual and observed frequencies.

Answer:
The main limitation of studying association by a comparison of actual and expected values of AB is that it only determines the nature of association between A and B ie., whether the association (if any) is positive or negative. It does not tell us about the degree of association ie., whether it is high or low.

5. Explain the method of Comparison of various proportions and products.

Answer:
Under this method attributes A and B are

i. Independent if $\frac{(AB)}{(B)} = \frac{(A\beta)}{(\beta)}$

ii. Positively associated if $\frac{(AB)}{(B)} > \frac{(A\beta)}{(\beta)}$ and

iii. Negatively associated if $\frac{(AB)}{(B)} < \frac{(A\beta)}{(\beta)}$

If the relation (i) holds good, the corresponding relationship $\frac{(\alpha B)}{(B)} < \frac{(\alpha\beta)}{(\beta)}$;

$\frac{(AB)}{(A)} < \frac{(\alpha B)}{(\alpha)}$; $\frac{(A\beta)}{(A)} < \frac{(\alpha\beta)}{(\alpha)}$ will also hold good.

Further it can also be found out easily that A and B would be independent if (AB) (αβ) = (A β) (αβ).

6. Explain Yule's coefficient of association.

Answer:
For practical purpose it is enough to take a decision about whether the two attributes in question are associated, disassociated or independent. But in some cases the difference between observed and expected frequencies may be due to what are called fluctuations of sampling. Under such circumstances it becomes necessary to obtain an idea about the extent to which the difference between the observed and expected frequencies can be due to chance fluctuations. It would be convenient if the coefficient of association is such that its value is zero when the two attributes are independent, +1 when they perfectly associated and -1 when they are perfectly disassociated. Yule has given the easy formula which is as follows.

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

We know that when two attributes A and B are independent the value of (AB) (αβ) = (A β) (αB)

As such, if two attributes are independent, the value of the numerator in the above formula would be zero and the value of the coefficient of association would also be zero. Similarly if there is perfect association between the two attributes A and B the value of (Aβ) into (αB) would be zero and since it will be so both in the numerator and the denominator. It is evident that the value of the coefficient of association would be +1. Similarly if there is perfect disassociation between two attributes A and B the value of (AB) into (αβ) would be zero and since it will be so both in the numerator and denominator, the coefficient of association would be 1.

7. Show that -1 ≤Q≤1.

Answer:

Let (AB)(αβ)= a and (Aβ)(αB)=b. Then a≥0 and b≥0

$$\therefore |a-b| \le |a+b| \Rightarrow \left|\frac{a-b}{a+b}\right| \le 1$$

Hence $Q = \left|\frac{a-b}{a+b}\right| \le 1 \Rightarrow -1 \le Q \le 1$

8. Write the advantage of Yule' coefficient of association.

Answer:
An important property of Q is that it is independent of the relative proportion of A's and α's in the data. Thus if all the terms containing A in Q are multiplied by a constant k, say, its value remains unaltered. Similarly for B and β and α. This property renders it especially useful to situations where the proportions are arbitrary, for example, experiments.

9. Write the formula for finding Coefficient of Collignation.

Answer:
This coefficient is denoted by γ. The formula of calculation is as follows.

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

10. Express, Yule's Coefficient in terms of gamma.

Answer:

Let $\frac{(A\beta)(\alpha B)}{(AB)(\alpha\beta)} = k$ so that $\gamma = \frac{1-\sqrt{k}}{1+\sqrt{k}} \Rightarrow \gamma^2 = \frac{1+k-2\sqrt{k}}{1+k+2\sqrt{k}}$

$$\therefore 1+\gamma^2 = \frac{2(1+k)}{1+k+2\sqrt{k}} = \frac{2(1+k)}{(1+\sqrt{k})^2}$$

$$\therefore \frac{2\gamma}{1+\gamma^2} = \frac{2(1-\sqrt{k})(1+\sqrt{k})}{2(1+k)} = \frac{1-k}{1+k} = \frac{1-\frac{(A\beta)(\alpha B)}{(AB)(\alpha\beta)}}{1+\frac{(A\beta)(\alpha B)}{(AB)(\alpha\beta)}} = \frac{(AB)(\alpha\beta)-(A\beta)(\alpha B)}{(AB)(\alpha\beta)+(A\beta)(\alpha B)} = Q$$

Hence $Q = \frac{2\gamma}{1+\gamma^2}$

11. What do you mean by coefficient of contingency?

Answer:

Thus attribute A can be divided into a number of groups A_1, A_2, A_3 …,A_5 and similarly the attribute B can be sub-divided as B_1, B_2, B_3 …,B_5. It will be observed that each one of the classes A_1, A_2, A_3 etc., of the first attribute would be divided into a number of heads like B_1, B_2, B_3 etc., when a second attribute B is taken into account. Such classification is called Manifold Classification.

12. Explain the method of coefficient of contingency to find the association between two attributes.

Answer:

If A and B are completely independent of each other in the universe at large then the actual values $A_1 B_1$, $A_2 B_2$, etc., must be equal to their expected values which are in turn equal to $\frac{(A_1)(B_1)}{N}$ and $\frac{(A_2)(B_2)}{N}$ respectively. In other words if the observed frequency in each of the cells of a contingency table is equal to the expected frequency of that cell, A and B would be completely independent of each other.

If these values are not equal in all the cells it is an indication of association between the attributes A and B. In order to test the intensity of association, the difference between the actual and expected frequencies of various cells is calculated. With these differences the value of Chi-square is obtained. The value of chi-square is represented by

χ^2 = Σ[(differences of actual and expected frequencies)^2/expected frequencies]

Thus, Square contingency = χ^2

Mean Square Contingency $\phi^2 = \frac{\chi^2}{N}$

Karl Pearson has given the following formula for the calculation of "coefficient of mean square contingency". According to it the coefficient of mean square contingency,

$$C = \sqrt{\frac{\chi^2}{N+\chi^2}} = \sqrt{\frac{\phi^2}{1+\chi^2}}$$

13. What is the disadvantage of coefficient of contingency?

Answer:
The coefficient of contingency has a drawback and it is that it never reaches the limit of 1. The limit of 1 is reached by it only if the number classes are infinite. Ordinarily its maximum value depends on the values of s and t. That is the number of sub-divisions of the two attributes A and B.

14. Give the maximum value of C in tXt contingency table.

Answer
In a t X t contingency table, the maximum value of C is given by

$$\sqrt{\frac{t-1}{t}}$$

15. Give Tschuprow's Coefficient.

Answer:
Since the Pearsonian coefficient of mean square contingency does not reach the maximum limit of 1 and since this a drawback, Tschurprow has suggested the coefficient T. It is calculated as follows.

$$T = \sqrt{\frac{C^2}{(1-C^2)\sqrt{(s-1)(t-1)}}}$$ where C stands for coefficient of attribute A and t stands for sub-divisions of attribute B.