Frequently Asked Questions

1. What is class frequency?

   Answer:
   The number of units in different classes is called "class frequencies". Thus if number of blind and deaf people is 20 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.

2. What do you mean by consistence of the data?

   Answer:
   It is obvious that 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 e inconsistent.

   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

3. Explain different types of classes.

   Answer:

   The classes which represent the presence of an attribute or attributes are called positive classes. The classes which represent the absence of an attribute or attributes are called negative classes. The classes in which one attribute is present and the other is absent are called pairs of contraries.

   Thus: N, A, B and AB are positive classes

   α and β are negative classes

   Aβ and αB are pairs of contrary.

4. When you can say two attributes are independent?

   Answer:
   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

   a) The same proportion A's amongst B's as amongst β's and

   b) The proportion of B's amongst A's is same as that amongst the α's

5. Give an expression for finding the Coefficient of Association?

   Answer:

   The expression for finding the coefficient of association is given by Yule and is given by,

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

6. An investigation of 23,713 households was made in an urban and rural mixed locality. Of these 1,618 were farmers, 2015 well-to-do and 770 families were having at least one graduate. Of these graduate families, 335 were those of farmers and 428 were well-to-do, also 587 well-to-do families were those of farmers and out of them 156 were having at least one of their family member as graduate. Obtain all ultimate class frequencies.

   Answer:

   Let the attribute 'farming' be denoted by A, the attribute 'well-to-do' by B and 'having at least one graduate' by C.

   Given N=23713, (A) =1618, (B) =2015, (C) =770

   (AB)=587, (BC) =428, (AC) =335, (ABC) =156

   For three attributes A, B, C the number of ultimate class frequencies is 2^3=8, one of them being (ABC) =156.

   (ABγ) = (AB)-(ABC) =587-156=431

   (AβC)= (AC)-(ABC) =335-156=179

   (Aβγ)= (A)-(AB)-(AC) + (ABC) =1618-587-335+156=852

   (ABC)= (BC)-(ABC) =428-156=272

   (αBγ)= (B)-(AB)-(BC) + (ABC) =2015-587-428+156=1156

   (αβC)=(C)-(AC)-(BC) + (ABC) =770-335-428+456=163

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

   =23713-1618-2015-770+587+335+428-156

   =20504

7. Given the following ultimate class frequencies, find the frequencies of the positive and negative classes and total number of observations. (AB)=250, (Aβ)=120, (αB)=200, (αβ)=70

   Answer:

   N=(A)+(α)=(AB)+(Aβ)+(αB)+(αβ)=250+120+200+70=640

   (A)= (AB) + (Aβ) =250+120=370

   (B)= (AB) + (αB) =250+200=450

   (α)= (αB) + (αβ) =200+70=270

   (β)= (Aβ) + (αβ) =120+70=190

8. Given the following, calculate the frequencies of the remaining classes (A) =50, (AB) =30, (B) =40, (N) =100.

   Answer:

   Here we need find the frequencies of the classes, (α), (β), (Aβ), (αB) and (αβ)

   Instead of using the formula and finding as we have given only two attributes, we can write the given values in the table and find the missing values.

	A	α	
B	(AB) (30)	(αB) (10)	(B) (40)
B	(Aβ) (20)	(αβ) (40)	(β) (60)
	(A) (50)	(α) (50)	N (100)

   In the table the numbers written in black are the given frequencies. The other frequencies are written by filling the table. The last row and last columns give the totals of corresponding rows and columns. We are also given the grand total N in the table. Hence the other values are found by subtraction.

9. Husband with light eyes – 400 do you find any inconsistency in the above data. To investigate the association between eye-colour of husband and eye colour of wife the following data are available

   Husband with light eyes and wives with not-light eyes – 414

   Husbands with not light eyes and wives with light eyes – 260

Husbands with not light eyes and wives with not- light eyes – 238

Husband with light eyes – 400 do you find any inconsistency in the above data.

Answer:
Denoting Husband with light eyes by A and Husband with not-light eyes by α
Wives with light eyes by B and wives with not-light eyes by β, the given data are,
$(A\beta)=414$, $(\alpha B)=260$, $(\alpha\beta)=238$, $(A)=400$
From above, $(AB) = (A)-(A\beta) =400-414=-14$
Thus AB has got a negative value. Hence the given data are inconsistent as it is obvious that no class-frequency occurring by counting real attribute can be negative.

10. If in a village actually involved by anthrax, 70 percent of the goats are attacked and 85 per cent have been inoculated with vaccine; what is the lowest percentage of the inoculated that must have been attacked?

Answer:

Denoting the attribute of attack by A and the attribute of inoculation by B,

The given data are $(A) =70$, $(B) =85$ and $N=100$

We have to find the lowest percentage of (AB)

Now according to conditions of consistence,

$(AB)\geq 0$ and $(AB) \geq (A) + (B)-N= 70+85 -100 = 55$

Hence the lowest percentage of the inoculated that must have been attacked is,

$(AB)*100/B=55*100/85=65$ per cent.

11. In a population, of 200 students the number of married is 80. Out of 60 students who failed, 24 belonged to the married group. It is required to find out whether the attributes of marriage and failure independent.

Answer:

Suppose the attributes of marriage are represented by A and failure by B, the actual value of $(AB) =24$.

Expected value of $(AB) = (A) (B)/ (AB) = (80) (60)/200=24$

The actual value of (AB) and its expected values are equal. It means that attributes A and B are independent.

12. Given (A) =40, (B) =30, (AB) =20, N=100. Study the association between A and B; α and β; A and β; α and B.

Answer:

We can represent the given data in the shape of a table and obtain the frequencies of the missing classes.

	A	A	
B	(AB) (20)	(αB) (10)	(B) (30)
B	(Aβ) (20)	(αβ) (50)	(β) (70)
	(A) (40)	(α) (60)	N (100)

The above values are those which have been observed. We can now calculate expected frequencies. The expected frequency of

$$(AB) = \frac{(A)\times(B)}{N} = \frac{40\times 30}{100} = 12$$

$$(\alpha\beta) = \frac{(\alpha)\times(\beta)}{N} = \frac{60\times 70}{100} = 42$$

$$(A\beta) = \frac{(A)\times(\beta)}{N} = \frac{40\times 70}{100} = 28$$

$$(\alpha B) = \frac{(\alpha)\times(B)}{N} = \frac{60\times 30}{100} = 18$$

Now we can easily study the association between various attributes if we lay down the actual and expected values together. Thus

Class	Actual Values	Expected values
(AB)	20	12
(αB)	50	42
(Aβ)	20	28
(αβ)	10	18

From the above table it is clear that the actual values of (AB) and (αβ) are more than the expected values and as such A and B is positively associated. Similarly, α and β are also positively associated. In case of (Aβ) and (αB), the actual values are less than the expected values. As such A and β are negatively associated and similarly there is negative association between α and B.

13. In a sample of 500 children, 200 came from higher income group, and the rest from lower income group. The number of delinquent children in these groups respectively was 25 and 100. Calculate the coefficient of association between delinquency and income groups.

Answer:

Let A denote the higher income group then α would denote the lower income group.

Let B denote delinquent children then β would denote non-delinquent children.

To get the frequencies of the second order, we form the following table

	A	A	
B	(AB) (25)	(αB) (100)	(B) (125)
B	(Aβ) (175)	(αβ) (200)	(β) (375)
	(A) (200)	(α) (300)	N (500)

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)} = \frac{(25 \times 200) - (175 \times 100)}{(25 \times 200) + (175 \times 100)} = \frac{-12500}{22500} = -0.55$$

Thus there is negative association between income and delinquency.

14. From the data given below calculate Yule's coefficient of association between weight of children and their economic condition, and interpret it.

	POOR CHILDREN	RICH CHILDREN
BELOW NORMAL WEIGHT	75	23
ABOVE NORMAL WEIGHT	5	42

Answer

Let A denote poor children and B denote children below normal weight. Then α would denote rich children and β children above normal weight.

The given data are (AB) =75, (Aβ) =5; (αB) =23; (αβ) =42

Substituting the values in Yule's coefficient of association, we get,

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)} = \frac{(75 \times 42) - (5 \times 23)}{(75 \times 42) + (5 \times 23)} = \frac{3035}{3265} = 0.93$$

There is high degree of positive association between poor children and children below normal weight. This means that the chances of poor children being below normal weight are very high. Rich children will general have above normal weight.

15. Among the adult population of a certain town 50 per cent are males, 60 percent are wage earners and 50 percent are 45 years of age or over, 10 percent of the males are not wage earners and 40 percent of the males are under 45. Make a best possible inference about the limits within which the percentage of persons (male or female) of 45 years or over is wage earners.

Answer: We have given all the figures in percentages. Hence we take N = 100. Let us denote male by A, wage earners by B and 45 years of age or over by C.

Hence we have given N = 100, (A) = 50, (B) = 60 and (C) = 50.

Further $(A\beta) = 50 \times 10\% = 5$ **and** $(A\gamma) = 50 \times 40\% = 20$

Therefore $(AB) = (A) - (A\beta) = 45$ **and** $(AC) = (A) - (A\gamma) = 30$.

We are required to find the limits for (BC).

Using the conditions of consistency,

i. $(AB)+(BC)+AC) \geq (A)+(B)+(C)-N$ Implies $(BC) \geq (A)+(B)+(C)-N-(AB)-(AC)$

Implies $(BC) \geq 50 + 60 + 50 - 100 - 45 - 30 = -15$.

ii. $(AB) + (AC) - (BC) \leq A$ Implies $(BC) \geq (AB) + (AC) - (A) = 45 + 30 - 50 = 25$.

iii. $(AB) + (BC) - (AC) \leq (B)$ Implies $(BC) \leq (B) + (AC)-(AB) = 60+30- 45 = 45$.

iv. $(AC) + (BC) - (AB) \leq C$ Implies $(BC) \leq (C) + (AB)-(AC) = 50+45-30 = 65$.

Therefore from (i) to (IV), we have $25 \leq (BC) \leq 45$.

Hence the percentage of wage earning population of 45 years or over must lies between 25 and 45.