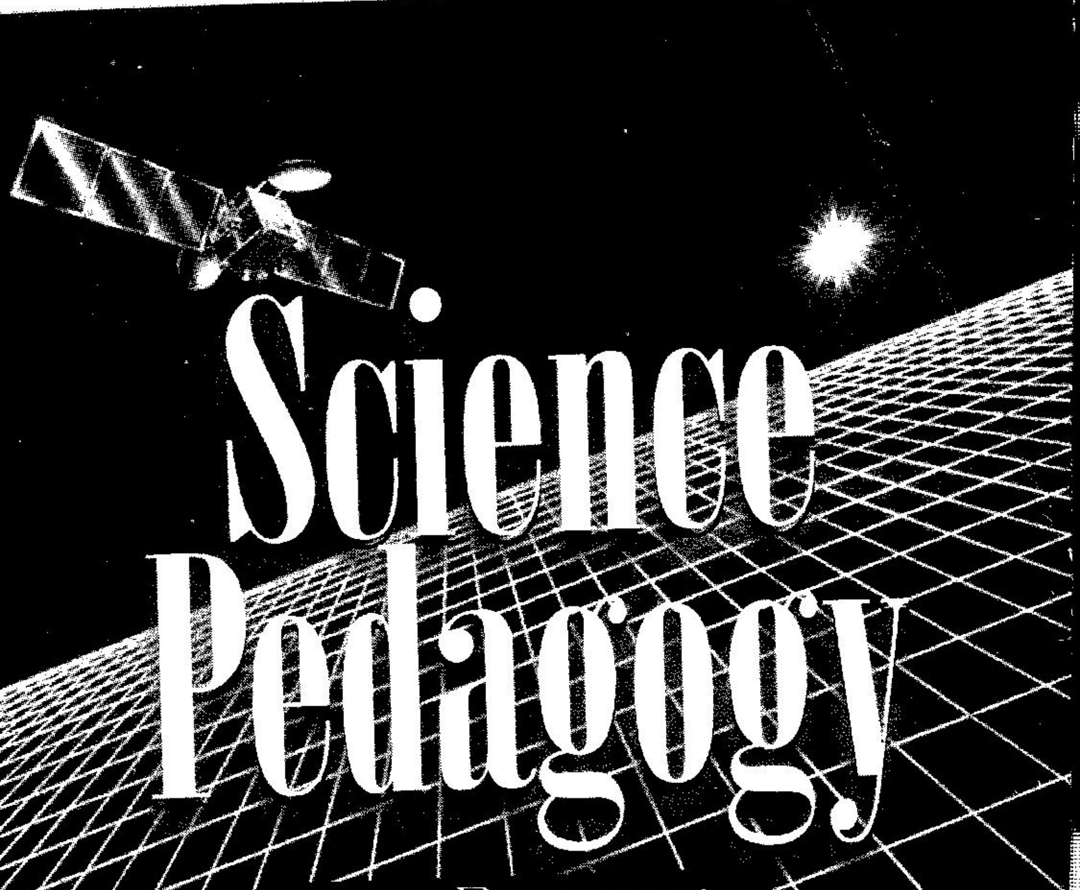
LAGOS STATE GOVERNMENT

Ministry of Education



Science Pedagogy

FOR BEGINNERS

Olayinka Oladunjoye

Benjamin Ayinde, Benjamin Etobro

Emmanuel Egho, Adigun Alamutu,

Stephens Oluyemi, Kehinde Olojede

8 DATA ANALYSIS AND INTERPRETATION OF RESEARCH FINDINGS

S. Adigun ALAMUTU
Senior Lecturer, Business Administration & Management Technology,
Faculty of Management Sciences, Lagos State University Ojo.

Introduction

After the required data have been gathered, the next step is to subject the arranged data to statistical analysis by the researcher with the intention of predicting or drawing inference on some preconceived notion about the population from which the data were drawn. At this stage, the researcher is faced with two options of using either parametric or non-parametric statistical method of analysis (statistical test).

Choosing the one to use is not a matter of gambling but a function of the objective to be achieved and the hypothesis to the tested amongst other factor.

Parametric and non-parametric statistics

Parametric tests are based on interval measurement of the dependent variable. they differ markedly from non-parametric tests based on certain assumptions associated with them. The assumptions are:

1. The variables are measured on interval scale. In other words, the dependent variable must be a continuous equal interval measure.
2. The population distribution of variable should be normal.
3. The variance or speed within the groups must be equal (homogeneous) therefore, data that do not meet the above assumptions are analysed by the use of non parametric tests may be, and often are more powerful in detecting population differences when the above stated assumptions are not satisfied.

Some examples of parametric test of hypothesis are:

1. Student "t" test (or t-test)
2. Analysis of variance (ANOVA) F-test
3. Z test
4. Least square method of regression analysis
5. Pearson product moment correlation technique for ranking

Some examples of Non parametric statistics are

1. Chi square (X^2)
2. Spearman Rank order correlation coefficient
3. Kruskal- wallis test
4. Wilcoxon whitney Test

Now the test

1. Student "t" test or (t-test)

When there are two independent samples and specific experimental treatment is assigned to each groups and after the treatment the two independent samples and specific experimental treatment is assigned to each groups and after the treatment the two groups are compared with respect to certain characteristic in order to find the effect of the treatment. The statistics that is appropriate here is the t- test.

The "t-test" is in two main forms independent samples and dependent sample.

If for example, we desire to determine whether stress affect problem solving performance among B.SC (yr 1) students of the Management Technology Unit of Lagos State university the following procedure may be followed.

The first step is to divide the class into two randomly, designate or give a label to each of the groups. The group members must come from homogeneous group and selected ramdomly. That gives each member an equal chance of belonging to a group. Thus, the average performance of the two groups in a problem solving task should not significantly differ prior to treatment.

After the treatment however the mean performance of the two groups should differ significantly if stress is actually related to problem solving performance.

The formula for t-test independent samples is given as

$$t = \frac{X_1 \; X_2}{\sqrt{\left[\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 \; 2}\right]\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

Example: A researcher is carrying out a study on the effects of being well fed before an examination on the candidates' performance in an examination. In performing this research, the first step to take is to divide examination by the same lecturer into two by random selection. One group is denied the opportunity of having breakfast while the other is provided with all their favorites before the examination which starts 8.00 am in the morning. Assuming the two groups are lebelled A and B respectively.

The test of the null hypothesis is recorded below complete with all sample data.

Ho: There is no difference between the mean academic performance of students before and after meal.

H_A: There is difference between the mean academic performance students before and after meal.

Assuming the following marks were obtained by the students in the two groups after the examination.

Subject	Groups A	Group B
1	60	62
2	87	80
3	45	46
4	53	48
5	67	70
6	76	60
Total	388	366

From the table above, it can be deduced that

$n_1 = 6$ $n_2 = 6$

$X_1 = 64.67$ $X_2 = 61.00$

$S_1 = 15.34$ $S_2 = 12.95$

$S_1^2 = 235.47$ $S_2^2 = 167.6$

$$t = \frac{64.67 \; 61}{\sqrt{\left[\frac{(6-1)235.47 + (6-1)167.61}{6+6-2}\right]\left[\frac{1}{6} + \frac{1}{6}\right]}}$$

$$= \frac{3.67}{\sqrt{\left(\frac{2015.35}{10}\right)\frac{1}{3}}}$$

$$= \frac{3.67}{\sqrt{67.18}} = \frac{3.67}{8.196} = 0.4478 = 0.45$$

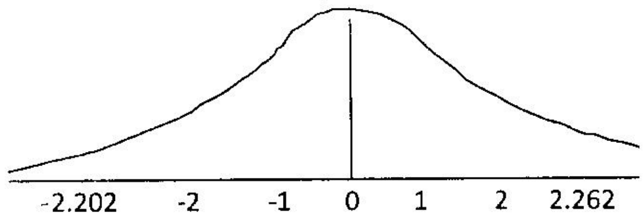
The "t" ratio tells the researcher that the observed difference is 0.45 times as great as the difference that would be expected under a true null hypothesis. To determine the statistical significance however, the researcher needs to consult the "t" test table.

To obtain the degree of freedom (d.f) the formula $n_1 + n_2 \; 2$ is used

The degree freedom $= 6 + 6 \; 2 = 10$

The critical value at degrees of freedom 10 and level of 0.05, using two tail table is 2. 262. Therefore, since the calculated value is less than the critical or table value, H_A is therefore rejected while Ho is embraced.

Diagram



Now, the t-Test for dependent sample: Recall that in the independent sample, the data are not from same source. In dependent sample however, the desire of the research is to compare the means obtained by the same group under two different experimental conditions. In such cases, the groups are no longer independent since the composition of one group is related to the other group. A typical example of t-test for dependent sample can be found in pre test and post-test control designs.

The formula is as follows:

$$t = \frac{\overline{D}}{\sqrt{\dfrac{\sum D^2 - \dfrac{(\sum D)^2}{N}}{N(N-1)}}}$$

Where t = t value of non independent means

D= The difference between the paired scores

$\overline{D}$ = The mean of the difference between the paired score

$\sum D^2$ = The sum of the square difference

N = The number of pairs

Here, the degrees of freedom is calculated using the formula

d.f = N - 1

Where N = number f pairs

2. ANALYSIS OF VARIANCE

Analysis of variance (ANOVA) is a statistical technique that is used to test the equality of more than two population means by analyzing the sample variance. It determines whether mean scores in one or more factors differ significantly from each other, and whether the various factors interact significantly.

The major differences between t-test and ANOVA are:-

(i) t- test considers the difference between two means while ANOVA considers differences between two or more means

(ii) ANOVA is a more versatile technique than the t-test

For a better understanding of ANOVA, consider the following illustration

A researcher is interested in finding out the effect of alcohol on the academic performance of students in higher institutions of learning. He therefore selected 20 students who do not take alcoholic drinks and randomly assigned them to one of the three groups he has formed. Groups 1,2 and 3. Group 1 is controlled, he administered moderate level of alcoholic drink to Group 2 and high level to Group 3. After the treatment, he subjected them all to a simple academic task and obtained the following result.

Groups 1	Groups 2	Groups 3
7	9	3
5	6	2
6	3	1
5	2	1
4	4	6
9	6	3
8	5	1
9	4	7
6	5	2
7	4	4

The question now is whether or not the differences among the means of the above scores are great enough to be statistically significant or is it likely that they occurred by chance? To answer this question, F-ratio is the best way out.

ANOVA involves eight basic steps as analysed below"

Step 1: Find the sum of the squared deviation of each of the grand mean the result of this step is known as the total sum of squares and is found by applying the formula

$$SS\,To = \Sigma X^2 \quad \frac{(\Sigma X)^2}{N}$$

$$\text{i.e.} \Sigma X^2_1 + \Sigma X^2_2 + \Sigma X^2_3 \quad \frac{(\Sigma X_1 + \Sigma X_2 + \Sigma X_3)}{N_1 + n_2 + n_3}$$

Step 2: Find the sum of squared deviations of the grand mean. This is known as the sum of square between groups. The formula is

$$SS_b = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \frac{(\Sigma X_3)^3}{n_3} + \frac{(\Sigma X)^2}{N}$$

Step 3: Find the sum of squared deviations of each score from its own group mean. This index is known as the sum of the squares within groups. This is determined by the formula:

$$SSn = (\Sigma X_1^2) \quad \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \Sigma X_2^2 + \frac{(\Sigma X_3)^2}{n_3}$$

Step 4: Determine the degree of freedom for between groups. This is determined by the formular:

$$d.f_b = G - 1$$

where G is the number of groups

Step 5: Determine the degree of freedom within groups. This is determined using the formula: $df_w (n_1 - n_2) + (n_2 - 1) + (n_3 - 1)$

Step 6: Find the sum between groups mean square and within groups mean square using the formula:

$Ms_b = SS_b / df_b$

and

$MS_w = SS_w / df_w$

Step 7: Calculate the F ratio using the formula:

$$F\text{-ratio} = \frac{Msb}{Msw}$$

Step 8: Present the summary of the analysis as shown below:-

Source	SS	df	Ms	F	significance (p)
Between	xx	xx	x		
Within groups	xx	xx	xx		
Total		xx			

Applying the above step to the table obtained, we have the following:

Groups 1		Groups 2		Groups 3	
X_1	X_2^2	X_2	X_2^2	X_3	X_3^3
7	49	9	81	3	9
5	25	6	36	2	4
6	36	3	9	1	1
5	25	2	4	1	1
4	16	4	16	6	36
9	81	6	36	3	9
8	64	5	25	1	1
9	81	4	16	7	49
6	36	5	25	2	4
7	49	4	16	4	16
66	**462**	**48**	**264**	**30**	**130**

$n_1 = 10$ $n_2 = 10$ $n_3 = 10$

$X_1 = 6.6$ $X_2 = 4.8$ $X_3 = 3.0$

$$SSTO = \Sigma x^2 \frac{(\Sigma x)^2}{N}$$

$$= 462 + 264 + 130 \frac{(66+48+30)^2}{30}$$

$$= 856 - \frac{(144)^2}{30}$$

$$= 856 - 691.2 = 164.8$$

$$SSb = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \frac{(\Sigma X_3)^2}{n_3} - \frac{(\Sigma X)^2}{n}$$

$$= \frac{(66)^2}{10} + \frac{(48)^2}{10} + \frac{(30)^2}{10} \frac{(144)^2}{30}$$

$$= 435.6 + 230.4 + 90 \quad 691.2$$

$$= 756 \quad 691.2 = 64.8$$

$$SSw = \Sigma X_1^2 - \frac{(\Sigma X_1)^2}{n_1} + \Sigma X_2^2 - \frac{(\Sigma X_2)^2}{n_2} + \Sigma X_3^2 - \frac{(\Sigma X_3)}{n_3}$$

$$= 462 - \frac{(66)^2}{10} + 264 \frac{(48)^2}{10} + 130 \frac{(30)^2}{10}$$

$$= (462 - 435.6) + (264 - 230.4) + (130 - 90) = 26.4 + 33.6 + 40$$

$$= 100$$

$$df_b = 6 - 1$$

$$= 3 - 1 = 2$$

$$dfw = (n_1 - 1) + (n_2 - 1) + (n_3 - 1)$$

$$= (10 - 1) + (10 - 1) + (10 - 1)$$

$$= 9 + 9 + 9 = 27$$

$$Msb = \frac{SSb}{dfb} = \frac{64.8}{2}$$

$$= 32.4$$

$$MSw = \frac{SSw}{dfw} = \frac{100}{27}$$

$$= 3.7$$

$$F\text{-}ratio = \frac{Msb}{Ms_w} = \frac{32.4}{3.7} = 8.76$$

The final analysis is hereby presented as follows:

Source	Ss	Df	Ms	F	Significance
Between group	64.8	2	32.4	8.76	
Within group	100.0	27	3.7		
Total	164.8	29			

To ascertain whether the F-ratio we have obtained is significant or not, the F-table has to be consulted. If the calculated F-Value is greater than the table value at 9 specific 'P' level, then H0 is rejected, otherwise, it is accepted.

In the example under consideration, the tabulated F-value with degree of freedom 2 and 27 is 3.35.

Since the tabulated F-value is less than the calculated F-value, it can be therefore be conducted that there is a significant difference among means of the three experimental groups.

Pearson Product moment correlation co-efficient (The Pearsonr)

The Pearson product moment is a parametric statistics that measures the relationship between two sets of variables. Two sets of scores are either positively related, negatively related, highly positively related, lowly negative related or no relationship between them.

When high; positives $r>0.5$
When lowly positive $r<0.5$
When highly negative $r>-0.5$
When there is no relationship, $r=0$

The formula for finding the Pearson product moment (Pearson r) is calculated using the formula

$$r = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{(N\Sigma X^2 \ (\Sigma X)^2][N\Sigma y^2 - (\Sigma y^2)]}}$$

Where:

r = Product Moment correlation co efficiency

Σx = sum of x scores

Σy = sum of Y scores

$(\Sigma x)(\Sigma y)$ = product of X score and its corresponding Y score

Illustration: Assuming a researcher is interested in finding our whether or not students who are good in mathematics are also good in physics. It simple academic tests are administer on 8m students who are mathematics and physics students and the following results are obtained.

Student	Mathematics (x)	Physics (Y)
Lateef	5	5
John	4	3
Micheal	2	1
Audu	3	4
Afusat	2	1
Funmi	4	3
Total	20	17

The question here is that, are these two sets of scores related? For yes, is it positively negative? If not, does that mean that there is no relationship at all now let's find out.

$$r = \frac{N\Sigma Y - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma x)^2][N\Sigma y^2 - (\Sigma y^2)]}}$$

The table above can be presented according to the formula as follows:

N	X	Y	X^2	Y^2	XY
1	5	25	25	25	25
2	4	16	9	9	12
3	2	14	1	1	2
4	3	9	16	16	12
5	2	4	1	1	2
6	4	16	9	9	12
Σ	20	74	61	61	65

Substituting the values obtained from the table

$$r = \frac{6(65)-(20)(17)}{\sqrt{[6(74)-(20)^2][6(61)-(17)^2]}}$$

$$r = \frac{390-(20)(17}{\sqrt{[6(74)-(20)^2][6(61)-(17)^2]}}$$

$$r = \frac{390-340}{\sqrt{[6(74)-(20)^2][6(61)-(17)^2]}}$$

$$r = \frac{50}{\sqrt{(444-400)(366-289)}}$$

$$r = \frac{50}{\sqrt{44 \times 77}}$$

$$r = \frac{50}{\sqrt{3388}}$$

$$r = \frac{58}{\sqrt{58.21}}$$

$$r = 0.85895$$

$$r \simeq 0.86$$

Interpretation: The above result shows that there is a positive (high) relationship between the students' scores in mathematics and physics. In other words, a student that is mathematically inclined has every tendency to be good in physics and vice versa.

In the other way round, there is a negative relationship when the value of r is negative. That is, a student that performs well in a subject is likely to perform poorly in another subject and vice versa.

If the value of r is nearly 0.00, then the scores are not related. This implies that a student's score in one subject is no indication of the student's score in the other subject.

REGRESSION ANALYSIS

If we have two sets of variables in such a way that one set depends on the other, regression analysis enables the researcher to express the relationship between the two and predict on one given the value of the other whether or not contained in the given sets.

The model for prediction in given as

Y	= a + bx
Where Y	= dependent variable
a	- Y - Intercept
b	= the gradient/slope/rate of change
X	= Independent variable

The condition is that

Y = f (x)

Which implies that:

Y is a function of X

Meaning that;

Y depends on X

In other words, Y cannot change until when X changes. X on the other hand can change on its own. Y can be likened to a working husband. In other words, the income of Y depends on X. Y's income has changed. If X's income reduces, Y's income will reduce.

Finally, if X is sacked, Y's income, all things being equal comes to zero level.

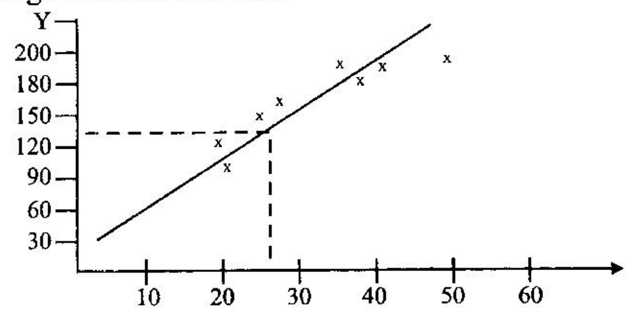
The regressing analysis can be obtained either through manual or scientific means. The manual method involves drawing the best fit line through scattered points on X and Y axis otherwise known as scatter grain or scatter diagram. This method has been found unscientific because one researcher's best fit line may not be another's. however the least square method gives an accurate and scientific best fit line.

Illustration: if the table below gives the sale of forms and the advertising expedition of Grace polytechnic for eight years as follows:

Year	Sales (Y) (N000) (N000)	Advertising expenditure (X) N000)
1999	140	25
2000	150	32
2001	107	20
2002	180	42
2003	166	38
2004	192	58
2005	174	39
2006	115	22

Sales has been denoted Y and advertising expenditure X because it is assumed that sales depends on advertisement. Thus, sales is the dependent variable while advertisement is the independent variable.

The scattergram looks like this:



This is drawn by being fair to all the parts plotted. The line must not slope too much to the left or right the question here is that to what extent can it be fair?

To obtain the regression equation

To obtain the regression equation, first, we compute the regression coefficient, 6

Using the formula we have

$$b = \frac{\text{Sum}(xy) - [\text{sum } x \text{ X sum } y]/n}{\text{Sum } X^2 - [(\text{sum } X)^2]/n} \quad \frac{\Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}$$

$$= \frac{45{,}016 - (276 \times 1232)/8}{10626 - (276)^2/8}$$

$$= \frac{45{,}016 - 42.504}{10{,}626 - 9{,}522}$$

$$= \frac{2{,}512}{1{,}104}$$

$$= 2.27536$$

Next, we compute the regression constant, thus;

$$A = \frac{\text{sum } y}{N} - b \frac{\text{sum } x}{n} = y - bx$$

$$= \frac{1{,}232}{8} - 2.27536 \times \frac{276}{8}$$

$$= 154 - 78.5 = 75.50$$

Therefore, the regression equation, $y = a + bx$, is

$Y = 75.5 + 2.27536x$

$= 75.5 + 2.275x$ (to three decimal places)

To get sales when advertising expenditure is N35,000 is to obtain y when x =35. Remember that the figures are in thousands. Thus:

$Y = 75.5 + 2.27536\ (35) = 75.5 + 79.6376 = 155.1376$

Therefore, sales, Y is N155,138.60, i.e. 155.1376 x 1000 (the value of 6 used in computing sales, y, is 2.27536 for a more accurate result).

Sales Y was found to be N120,000 in a certain year. The corresponding advertising expenditure, x, for that month will be obtained if we substitute for Y in the regression equation.

Thus:

$120 = 75.5 + 2.27536x$

$2.27536x - 120 - 75.5 = 44.5$

$x = 19.55734$

Therefore, advertising expenditure = N19,557.34 that is 19.55734 x 1000

The correlation coefficient is obtained as follows:

$$R = \frac{n\text{Sum}(xy) - \text{Sum } x \text{ X Sum } y}{\text{Sqrt.}[n(\text{sum}X)^2 \text{ X Sqrt}(n(\text{Sum } y)^2]}$$

$= \quad 8 \times 45{,}016 \quad 276 \times 1.232$

Sqrt. $[8 \times 10{,}262 - 276) \times$ Sqrt. $(8 \times 196.334 - 1{,}232)$

$$= \frac{20{,}096}{21{,}604.5}$$

$= 0.930177$

Thus the correlation coefficient between advertising expenditure (x) and sales (y) is:

$r = 0.930177$

Some examples of Non-parametric statistics are:

1. Chi square (x^2)
2. Spearman rank order correlation coefficient
3. Wilcoxon Mann Whitney Test
4. Kruskal-Wallis Test

The chi-square test

Chi-square is pronounced as "kai" square and notation X^2. It is used in comparing the observed set of date to expected ones in one or more variables.

The assumption of X^2 are:

1. The subjects of each group are randomly and independently selected from the population.
2. The groups are independent.
3. Each observation must qualify for one and only one category.
4. The sample size must be fairly large.

The formula for computing X^2 is as follows:

$$X^2 = \frac{(O-E)^2}{\epsilon}$$

Where:
X^2= chi-square calculated
O= observed (or actual) frequency
E= Expected frequency
ϵ = Summation

The computation requires the determination of expected frequency for each of the box in the table, subtracting each expected frequency from the actual frequency in the box, squaring it and dividing by the expected frequency in each box and then summing up all the quotients.

Illustration:
Suppose a researcher is interested in finding out whether or not, there is difference between male and female residents of Lagos State in their preference for two of the Yoruba traditional foods, "Amala" and "Iyan". He samples 550 people in the state (300 men and 250 women). Assuming he obtains the following result.

Responses

Sex of respondent	**Amala**	**Iyan**	**Total**
Male	100 (1)	200(b)	300
Female	126©	124(a)	250
Total	226	324	550

The most appropriate test for the above research is chi-square.

Hypothesis:
H_0: There is no difference between male and female Lagosians in their preference for Yoruba traditional foods.
H_i: There is difference between male and female Lagosians in their preference for Yoruba traditional foods.

Expected Values
To compute the expected value for each of the boxes, the row total is multiplied by the column total and the result obtained is divided by the grand total.

i.e.

$$E = \frac{RT \times CT}{GT}$$

Where E = Expected value
RT: Row Total
CT: Column Total
GT: Grand Total

It is computed as follows:

For box a, $E = \frac{300 \times 226}{550} = 123.27$

b, $E = \frac{300 \times 324}{550} = 176.73$

c, $E = \frac{250 \times 226}{550} = 102.73$

d, $E = \frac{250 \times 324}{550} = 147.27$

The expected value table can then be presented alongside the observed as follows:

Responses

Sex of respondent	Amala	Iyan	Total
O	E	O	E
Male 100	123.27	200	176.73
O	E	O	E
Female 126	102.73	124	147.27

The X^2 value can then be computed as follows:

Respondents	O	E	O-E	$(O-E)^2$	$(O-E)^{2/E}$
Male (Amala)	100	123.27	-23.27	541.49	4.39
Male (Iyan)	200	176.73	23.27	541.49	3.06
Female (Amala)	126	102.73	23.27	541.49	5.27
Female (Iyan)	124	147.27	-23.27	541.49	3.68
Total	550	550	O	2165.96	16.40

X^2 (cal) = 16.40

To determine the level of significance of the computed value of chi-square, the distribution of the chi-square values is referred. The number of degrees of freedom associated with the observed date is also determined as follows:

d.f = (R-DCC-1)

Where

d.f = degree of freedom

R = number of rows

C = number of columns

From the above

d.f = (2-1) (2-1) = 1 x 1 = 1

The critical value for 1 degree of freedom and an of 0.05 (5%) is 3.84. The null hypothesis (H_0) is therefore rejected since the computed value is greater than the critical value. This is attainable at 5% level of confidence certainty and 5% uncertainty.

H_A is accepted meaning that there is difference between male and female Lagosians in their preference for Yoruba foods.

Decision Rule: Calculated < critical / table value:
Rejected H_A or H_1 and Accept H_0
Calculated > critical / table value:
Reject H_0 and Accept H_A or H_1

Spearman (rho)

This is a non-parametric test and opposite to Pearson product moment mark correlation. In this case, the highest score is ranked 1 while the least is ranked accordingly depending on the number of scores under consideration. Where two scores are the same, they are ranked equal. It is most appropriate when our intention is to determine the relationship between two ranked quantifiable variables.

Using our illustration on Pearson product moment rank correlation, the following table is obtainable.

S/N	Student	Score in mathematics (x)	Rank	Scores in physics (y)	Rank
1.	Lateef	5	1	5	1
2.	John	4	2	3	3
3.	Micheal	2	5	1	5
4.	Andy	3	4	4	2
5.	Afusat	2	5	1	5
6.	Funmi	4	2	3	3

The formula is given as:

$$rho = 1 - \frac{\sum 6d^2}{N(N^2-1)}$$

Where:

$\sum d^2$ = sum of the squared differences

D = difference between the ranks for each student

N = number of students in a class

d^2 = the square of the difference.

This is calculated as follows:

Students	**Rank X**	**Rank Y**	**d (X - Y)**	**d^2**
Lateef	1	1	0	0
John	2	3	-1	1
Micheal	5	5	0	0
Andy	4	2	-2	4
Afusat	5	5	0	0
Funmi	2	3	-1	1
Σ			0	6

Substituting the values obtained from the table, we have

$$rho = \frac{1-6(6)}{6(6^2-1)}$$

$$rho = \frac{1 - 36}{6(36-1)}$$

$$rho = \frac{1 - 36}{6(35)}$$

$$rho = \frac{1 - 36}{210}$$

rho: 1-0.1714
rho = 0.82857
rho = 0.83

From the above calculation, it is obvious that there is a high positive relationship existing between the two scores, that is, a high score in Mathematics attracts a relatively high score in physics.

Conclusion

I hope you have enjoyed reading this chapter. Always remember that correlation measures relationships between variables while regression is used for predictions or forecasts. While both Pearson product movement (r) and spearman (rho) values ranges from -1 to 1