A FOUR WAY CATEGORICAL DATA ANALYSIS OF ACADEMIC PERFORMANCE OF STUDENTS

Ugege, L. & Nwakuya, M. T.
Department of Mathematics/Statistics, University of Port Harcourt
Port Harcourt, Rivers State, **NIGERIA**

ABSTRACT

This paper is centered around a four way categorical data analysis of performance of students in a secondary school. It is primarily based on the Chi-Square Test of Independence and also Odds Ratio Test. The test showed that the performance of students was dependent on age, gender and region of origin. It also showed that the performance of Junior school class 1 students was also dependent on age. The odds ratio test was performed and it showed that females from the south in the 16 to 20 years age group have the least pass odds of 0.174 and females from the north between 10 to 15 years have the highest pass odds of 9.009 within the group respectively. We also found out that junior school class one student's odds ratio test showed that students of 12years had the highest pass odds ratio of 1.343.

Keywords: Odds ratio, Test of Independence, Categorical Data, Contingency Table, Four-way Analysis.

INTRODUCTION

Research works have been characterised with categorizing or grouping data into one or more bins, classes or groups. Data can be categorized into ordinal or nominal groups. The ordinal data are those groups where positionis paramount like grouping items into levels: low, moderate or high. While for the norminal category, position or order is irrelevant. A good example of nominal category is grouping data by gender: male or female or categorizing data by marital status: single, married or divorced. This research work is no different, in fact its whole essence is centered around grouping data and using frequency counts to analyse and draw up conclusions about the data being examined.

Chi-square test are used to analyse frequency data. This analysis can be used for the following tests: goodness of fit test; test of independence and the tests of homogeneity. Nduka (2015). The goodness of fit test is a statistical procedure used to test if an assumed distribution is correct. Hsiao-Mei (2009). When the experimenter is concerned with association between variables then the test of independence is carried out. The test of homogeneity is employed to check if two or more well-defined populations are equal. Nduka(2015).

This research is borne with the responsibility of finding out the association between age, gender and region of origin of students to performance. To achieve this, categorical data were assembled to this effect showing the frequencies of these groups (age, gender and region of origin) grouped into different performance levels (below 65% and above 65%) drawn into a contingency table. The chi-square test of independence uses the contingency (cross-classification) table to establish the nature of relationship between the variables of the table. The contingency table shows the nature in which the variables are either related or not related

to each other. Here the variables are partitioned into mutually exclusive (independent) categories.
The test begins with a null hypothesis of no association between the variables while the alternate agrees with association present between the variables. The test does not tell the nature of the relationship between the variables, all it does is to check the evidence of dependence between the variables.

The test statistic used is given below:
…………………………………………………(1)

Where:
– is the chi-square distribution with (r – 1)(c – 1) degree of freedoms
– Observed frequencies
– Expected frequencies

Nduka(2015).

The two major assumptions in the application of the chi-square test of independence are:

Sample size assumption: must be large as the p-values obtained are approximations.

Independence assumption: The data must be uncorrelated.

Aim and Objectives

The aim of this research work is to test the association between age, gender and region of origin on the performance of students in a secondary school by means of chi-square test of independence.
To achieve this aim, the following objectives are stated:

1. Build a four-way contingency table of age, gender, region of origin on the performance of students.
2. Design a simple computer program (php script) to coalate the frequencies into a contingency table and perform the computerations of the chi-square on the student data.
3. Test the data for complete independence among categories.
4. Run pass odds ratio test on the students data.
5. Draw up conclusions based on the results of the chi-square test statistic performed on the student records.

REVIEW OF LITERATURE

Intuitively, people often like to relate different factors to yeild or performance. This has led to various works on tests of association on different fields using several methods including chi-square test of indpendence. In this chapter we will reflect on the different literatures/works by other researchers.

The chi-square test more properly known as the Pearson's Chi-square test dates back to paper publications by Karl Pearson's in the 1900s. Howell (2009). This test is centered around categorical data; it compares the observed count to the count which would be expected under the assumption of no association between the different classifications (categories) of the data under study.

Kevin (1983) in his paper *"The Use and Misuse of Chi-Square: Lewis and Burke Revisited"* outlined some of the major sources of errors in the application of the chi-square

ISSN 2059-9951

statistic to include: lack of independence among single events, small theoretical frequencies, neglect of frequencies or non-occurence, failure to equalize the sum of the observed frequencies and the sum of the theoretical frequencies, indeterminant theoretical frequencies, incorrect or questionable categorizing, use of nonfrequency data, incorrect determination of the number of degrees of freedom and incorrect computations. Lewis and Burke (1949).
On the issue of small theoretical frequencies, Lewis and Burke pointed out this as the *most common weakness* of the chi-square tests. They opined that an expected frequency value of 5 is the absolute lowest limit with a preference minimum of 10.Kevin (1983).

The issue of minimum frequency value have been sparsely opinated.Jeffreys, Kempthorne and Slakter recommended a value of 1, minimum expected value of 5 by Fisher, Cramer and Kendall both recommended 10 and 20 respectively. Kevin (1983).

Cochran taking a slightly different approach rather than pegging the expected minimum frequency on only a value of 1 to 5, gave a bench mark for the acceptance of this minimum expected value to be no more than 20% of the cells having this value. Sorana (2011).
Howel (2009) in his paper *Chi-Square Test: Analysis of Contingency Table* also pointed out the effect of small frequencies on the chi-square test. He pointed out that the chi-square statistic is only an approximation of the chi-square distribution and the approximation only worsens with small expected frequencies. He noted that with small expected values the chi-square statistic are quite discrete rather than continuous. He also refrenced Cochran as stating an arbitrary minimum value of 5 as a general guide to the implementation of the chi-square test.

Sorana (2011) in his publication "Pearson-Fisher Chi-Square Statistic Revisited" though still in line with Cochran on the issue of small expected frequencies added that the application of the chi-square tests to small frequencies leads to an unaccepted level of type II error (i.e. accepting the null hypothesis when it is false). He opined that there are three main problems which emanate in the application of the Chi-square tests in order to perform comparisons between theoretical and observed distributions: problem of defining the frequency class, calculating the chi-square statistic and the problem of applying the tests.

Sorana (2011) pointed out that the Fisher's Exact Test is prefered when small expected values are present. This test is based on calculation of marginal probabilities. One down side as stated by Sorana (2011) of the Fisher's Exact test is that it has a formula only for 2 x 2 contingencies.

Mehta and Patel (2000) in their book *Exact Tests* stated that in the asymptotic method, p values are computed based on the assumption that the data given is from a sufficiently large sample and conforms with a particular distribution. They added that the asymptotic method fails when the data set is small, sparse, contains many ties, is unbalanced or is poorly distributed. In such cases exact methods are preferable as they do not rely on the assumptions that may not be met by the small data set.

In a situation where the data are too large for the exact method but still do not meet the assumption of the asymptotic method, the Monte Carlo method provides an unbiased estimate of the expected p value. Mehta and Patel (2000).

Lunsford (2010) in his paper *Water Taste Test* presented data, analysis and results on experiments carried out at Longwood University in the Fall of 2008. The experiments were centered around the preference of bottled water over tap water. In particular it was conjecturered that there was a preference of bottled water over tap water. The experimental question of importance to this work is to find out if this preference is gender associated. This leads to the application of chi-square test of independence with a null hypotheses: *Gender and First Preference are independent.*

From the data collected in the experiment and considering the assumption of minimum expected value, it was observed that the cell for tap-male category had an expected count of less than 5 but this only represented about 12.5% of the cells which means the chi-square test could be safely applied. The data gave a chi-square test statistic of 1.336 with a degree of freedom of 3 and a p-value of 0.721. This led to the conclusion that the result is not significant and the data presented had no evidence that from the preference already drawn there is gender associated to that result.

Marie (2008) in a lecture on *Use of the Chi-Square Statistic in a Test of Association Between a Risk Factor and a Disease* at the John Hopkins Bloomberg School of Public Health designed a 2 x 2 contingency table with categories exposure (yes and no) and diseased (yes and no) found out that the chi-square obtained was 29.1 with a degree of freedom of 1. From the chi-square table, in order to obtain discrepancies between observed and expected frequencies of this magnitude less than 0.001 that is unlikely to occur by chance. Thus the findings is unlikely to occur if there is no association between exposure and disease, hence there appears to be an association.

METHODOLOGY

This chapter talks about the methods used in the analysis of the test of association between age, gender, region of origin on the performance of students in a secondary school. This research work is to test the independence of age, gender and region of origin on the performance of students in a secondary school. This work would only be applicable to the school in view as the data collected is particular to it. A computer program would be designed in the course of the project to coalate the frequencies of data from the database and perform the appropriate computations on the data drawn.

Method of Data Analysis

1. Contingency Table Design
2. Bar chart of the pefromance level of the different categories
3. Chi-Square Value Calculations
4. Interpretation of the chi-square values gotten with the theoretical values.
5. Odds ratio test.

Research Design

For a data forming the contingency table, there are a few ways the research could be designed. One could follow the Fisher's famous *'tea-tasting'* experiment. This design had 8 cups of tea in total with four having milk poured first and the other half had milk added second. The tasters where required to assign four cups each to the order in which the milk was added. In this design all marginal total are fixed. (Lunsford, 2010). Another approach could be designed in such a way as to have only one set of marginals fixed. Having a fixed

ISSN 2059-9951

total with varying marginal totals is also another way to design the experiment. The last permutation would be to have unfixed marginal totals and sample size.
For this research, the data used was secondary data gotten from the archives (database) of the school under study. It comprised of records for about 12 school terms spanning 2012/2013 session through 2015/2016 session.

The research design approach was modeled to a double-blinded test: where the experimenter did not know which of the unique (independent) students records was choosen. The computer script written randomly selected one record from a list of each students records to meet the assumption of independence as stipulated in the application of the chi-square test. The program also coalated the various frequencies for the different groups and computed the chi-square value.

Contingency Table

A contingency or cross-classification table corresponds to two variables with disjoint categories represented by the rows and columns of the table. The *i-j*thentry of such table represents the count (frequency) when the *i*throw category and the *j*th column category co-occur.

Formally put, the counted observation constitute a set O of N entries. This cross-classification can be partitioned to I = {i_1, i_2, ... , i_m} and J = {j_1, j_2, ... , j_n} according to the categories of each variable. That is one variable has m-categories and the other n-categories. To correspond to the partitions, categories of each of the variables must be disjunctive and defined for all N entities in O. The table of the co-occurence members, Nij is refered to as the continency of cross-classification table. The cardinalities of classes *i*of **I** and *j* of**J** ususally are called marginals denoted by N_{i+} and N_{+j}, which are sums of co-occurence entries N_{ij}, in rows *i* and column *j* respectively.

Table 1: A simple design of a 2 x 2 contingency table.

Criterion 1	Criterion 2					
	1	2	3	. . .	C	**Total**
1	N_{11}	N_{12}	N_{13}	. . .	N_{1c}	**R_{1+}**
2	N_{21}	N_{22}	N_{23}	. . .	N_{2c}	**R_{2+}**
. . .	. . .	. . .	. . .	. . . . .	. . .	**. . .**
R	N_{R1}	N_{R2}	N_{R3}	. . .	N_{Rc}	**R_R**
Total	**C_{+1}**	**C_{+2}**	**C_{+2}**	**. . .**	**C_{+c}**	**N**

Chi-Square Statistics

The chi-square statistics is defined based on the observed frequencies (gotten from the data) and the expected frequeny of each classification. In this test the null hypothesis makes a case of the expected values for each cell (cross categorical groups) to be true. It is based on the difference between the observed and expected frequencies as earlier defined in equation 1
The chi-square statistic for a four-way table of four variables A, B, C, D is given below. Let

□ = P (A = i), i = 1, 2, ...,I
□ = P (B = j), j = 1, 2, ...,J
□ = P (C = k), k = 1, 2, ...,K
□ = P (D = l), l = 1, 2, ...,L

So that, □$_{ijkl}$ = □$_i$ □$_j$ □$_k$ □$_l$

$\square \quad = \quad (\square_1, \square_2, \ldots, \square_I)^T$
$\square \quad = \quad (\square_1, \square_2, \ldots, \square_J)^T$
$\square \quad = \quad (\square_1, \square_2, \ldots, \square_K)^T$
$\square \quad = \quad (\square_1, \square_2, \ldots, \square_L)^T$

The maximum likelihood estimates are given by

$\square_i \quad = \quad y_{i+++}/n, \; i = 1, 2, \ldots, I$
$\square_j \quad = \quad y_{+j++}/n, \; j = 1, 2, \ldots, J$
$\square_k \quad = \quad y_{++k+}/n, \; k = 1, 2, \ldots, K$
$\square_l \quad = \quad y_{+++l}/n, \; l = 1, 2, \ldots, L$

Under complete independence, the maximum likelihood estimates of the expected frequencies are mutually independent as
;
…………………………………………….(2)

Therefore the chi square is given below
………………………………………….(3)
Lin(2006)

The degree of freedom for the four-way test is given as

$df \quad = \quad (I - 1)(J - 1)(K - 1)(L - 1)$

Odds Ratio Test

Odds are another way of comparing two probabilites: probability that an event occurs to the probability that it doesnt occur. The odds ratio is the ratio of odds for two probabilities. Given the frequency table:

Table 2: Odds Ratio

CATEGORIES		Category Y		TOTAL
		Y1	Y2	
Category X	X1	A	B	X_{t1}
	X2	C	D	X_{t2}
TOTAL		Y_{t1}	y_{t2}	

………………………………….(4)

This gives the odds of outcome X_1 given that Y_1 versus Y_2. Ivo(2006)

Data Presentation and Analysis

This chapter covers the presentation and analysis of data, to statistically confirm the existence of dependence between: age, gender, region of origin and performance of students in a secondary school using bar charts and chi-square test of independence.

The students' data gotten from the database was coalated into a categorical data and drawn into a cotingency table. The contingency table categories used were:

age (10-14 years, 15-20 years), gender (male, female), region of origin (north, east, west and south) and the performance of (below 65% average, above 65% average).

A breakdown for the regions is as follows:

North: Adamawa, Bauchi, Benue, Borno, Gombe, Jigawa, Kaduna, Kano, Katsina, Kebbi, Kogi, Kwara, Nasarawa, Niger, Plateau, Sokoto; Taraba, Yobe, Zamfara
East: Abia, Anambra, Ebonyi, Enugu, Imo
West: Ekiti, Lagos, Ogun, Ondo, Osun, Oyo
South: Akwa Ibom, Bayelsa, Cross River, Delta, Edo, Rivers

The performance category (above 65% and below 65%) was adopted for this work to meet the observed frequency limitations of the chi-square implementation explained in previous chapters.

A double blind approach was taken to pick independent data into the contingency table to meet the requirement of independence. This approach involves selecting students with complete data from the database, randomly choosing one out of the numerous (termly) performance enteries in the database per student.

Furthermore, to achieve all these a php script was written to traverse the database and coalate the different frequencies, compute the expected counts per, percentages per row, display the result in a contingency table and compute the chi-square value.

Table 3: Contingency Table for student data.

Categes			Regions															
			North				East				West				South			
Sex			Male		Female		Male		Female		Male		Female		Male		Female	
Age			A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
Perfo-rmance	>65 %	Observed Expected %within P(Row|col)	88 88.71 14.31% 1.158	35 34.08 5.69% 1.250	18 12.98 2.93% 3.00	15 13.52 2.44% 1.500	74 95.20 12.03% 0.725	29 39.49 4.72% 0.659	6 9.20 0.98% 0.545	11 12.98 1.79 0.846	119 91.41 19.35% 2.380	62 49.22 10.08% 2.138	22 16.77 3.58% 2.444	16 14.06 2.60% 1.600	66 79.51 10.73% 0.815	41 37.86 6.67% 1.414	8 10.82 1.30% 0.667	5 9.20 0.89% 0.417
	<65 %	Observed Expected %within P(Row|col)	76 75.29 14.56% 0.864	28 28.92 5.36% 0.800	6 11.02 1.15% 0.333	10 11.48 1.92% 0.667	102 80.80 19.54% 1.378	44 33.51 8.43% 1.517	11 7.80 2.11% 1.833	13 11.02 2.49% 1.182	50 77.59 9.58% 0.420	29 41.78 5.56% 0.468	9 14.23 1.72% 0.409	10 11.94 1.92% 0.625	81 67.49 15.52% 1.227	29 32.14 5.56% 0.707	12 9.18 2.30% 1.500	12 7.80 2.30% 2.400
Pass Odds Ratio			1.340	1.562	9.009	2.249	0.526	0.434	0.297	0.716	5.667	4.568	5.976	2.560	0.664	2.000	0.445	0.174

NOTE: 'A' represents age interval 10 – 15 and 'B' represents age interval 16 – 20.

The data gave a significant chi-square test statistic of 64.934with a 3 degree of freedom. The chi-square tabulated value at 3 degree of freedom and 0.5% level of significance is 12.838. This result shows that at 0.5% level of significance the Students' performance is found to be dependent of age, gender and region of origin.

The records for junior secondary school one entry into the school was also coalated into a contingency table to test the independence of student performance on age at this class level.

> $\mathbf{H_0}$**:** Performance of students in junior secondary school one are independent of age.
>
> $\mathbf{H_1}$**:** Performance of students in junior secondary school one are not independent of age.

The result of this coalation is fully represented in the contingency table presented below.

Table 4: Contingency Table for junior secondary school one student data.

DATA OF OBSERVED/EXPECTED FREQUENCIES OF JUNIOR SECONDARY SCHOOL ONE STUDENTS							
CATEGORIES		Ages					TOTAL
		10 yrs	11 yrs	12 yrs	13 yrs	14+ yrs	
Performan~	**Above 65%** Observed Expected P(col|row)	 12 12.71 0.041	 102 97.65 0.498	 160 152.50 1.088	 28 36.79 0.100	 5 7.36 0.017	 307

	Below 65% Observed Expected P(col|row)	7 6.29 0.048	44 48.35 0.407	68 75.50 0.810	27 18.21 0.216	6 3.64 0.041	152
TOTAL		19	146	228	55	11	459
Pass Odds		0.854	1.224	1.343	0.463	0.415	

The chi-square statistic calculated was 10.449 with a 4 degree of freedom. At a 5% level of significance gives a tabulated chi-square value of 9.488, while at a significance level less that 2.5% with a chi-square tabulated value greater than 11.143. At 5% level of significance we reject the null hypothesis, but at a significance level less than or equal to2.5% the null hypothesis is accepted. For the nature of this test the 5% level of significance is acceptable. So we reject the null hypothesis and accept the alternate hypothesis that performance of student in junior secondary school one is not independent of age.

ODDs RATIO ANALYSIS

Table 5: Pass Odds Table for student data.

PASS ODD RATIO OF THE STUDENTS			
Region	**Sex**	**Age**	**Pass Odd**
North	Male	10-15	1.340
		16-20	1.562
	Female	10-15	**9.009**
		16-20	2.249
East	Male	10-15	0.526
		16-20	0.434
	Female	10-15	0.297
		16-20	0.716
West	Male	10-15	5.667
		16-20	4.568
	Female	10-15	**5.976**
		16-20	2.560
South	Male	10-15	0.664
		16-20	2.000
	Female	10-15	0.445
		16-20	**0.174**

The odds ratio in table 3 above gives the odds of someone from a region of a particular gender and age group passing to someone from the same category failing. The odds ratio extracted, shows that Northern female between 10 and 15 years is 9 times more likely to pass than a girl from the same region and age group failing. This value region holds the record for having the greatest odds ratio.

On the other hand the Southern female between the ages of 16 and 20 years holds the lowest pass odd ratio of 0.174. This implies an approximate ratio of 1:6 (pass:failure) within them. A category worthy of mention from the odds ratio table, are the Western females between the ages of 10 and 15 years of age. They have the second highest odds ratio of 5.976, showing that the odds of passing is about 6 times greater than for failing within this group.
From table 2 the pass odds of JS 1 students of 12years is the greatest among the other age groups of 1.343. This implies that the odds of a 12year old student passing in JS 1 is 1.3 times greater than a child failing of the same age.

Bar Charts

The stacked bar graphs in figure 1 and table 6 below show that the percentage of males and females in each region are almost cut at the 50% mark. The table 7 below gives the percentage gender composition of the school population across the various regions.

Table 6: Regional gender percentage composition of students in the school.

CATEGORIES	North	East	West	South
Male (%)	51.83	45.78	44.57	53.26
Female (%)	48.17	54.22	55.43	46.74

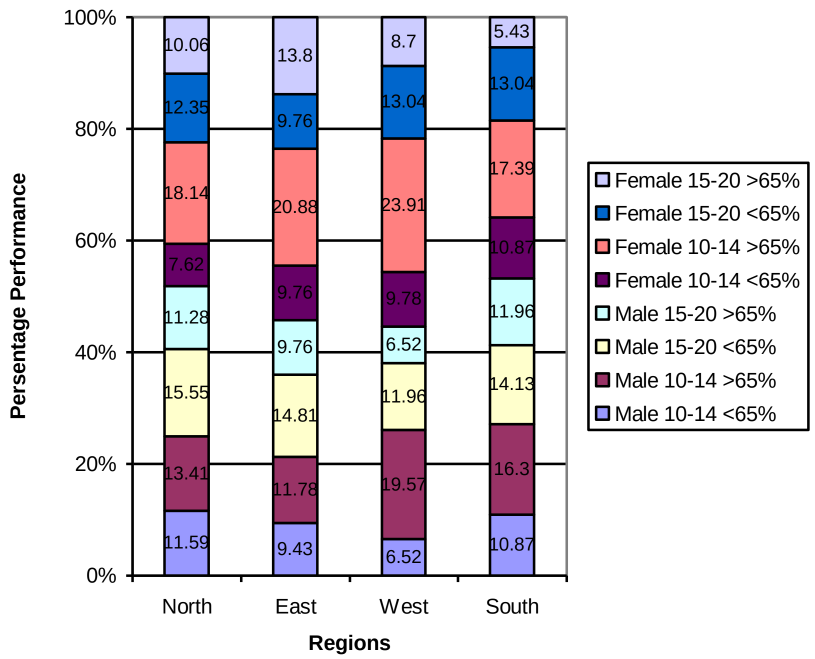
Table 7: Data of frequencies and regional percentage performance of the students' records

CATEGORIES		Sex								Total
		Male				Female				
Age		10-14yrs		15-20yrs		10-14yrs		15-20yrs		
Performance		□65%	□65%	□65%	□65%	□65%	□65%	□65%	□65%	
Region	**North** Observed % within	76 11.59%	88 13.41%	102 15.55%	74 11.28%	50 7.62%	119 18.14%	81 12.35%	66 10.06%	**656**
	East Observed % within	28 9.43%	35 11.78%	44 14.81%	29 9.76%	29 9.76%	62 20.88%	29 9.76%	41 13.80%	**297**
	West Observed % within	6 6.52%	18 19.57%	11 11.96%	6 6.52%	9 9.78%	22 23.91%	12 13.04%	8 8.70%	**92**
	South Observed % within	10 10.87%	15 16.30%	13 14.13%	11 11.96%	10 10.87%	16 17.39%	12 13.04%	5 5.43%	**92**
Total		120	156	170	120	98	219	134	120	1137

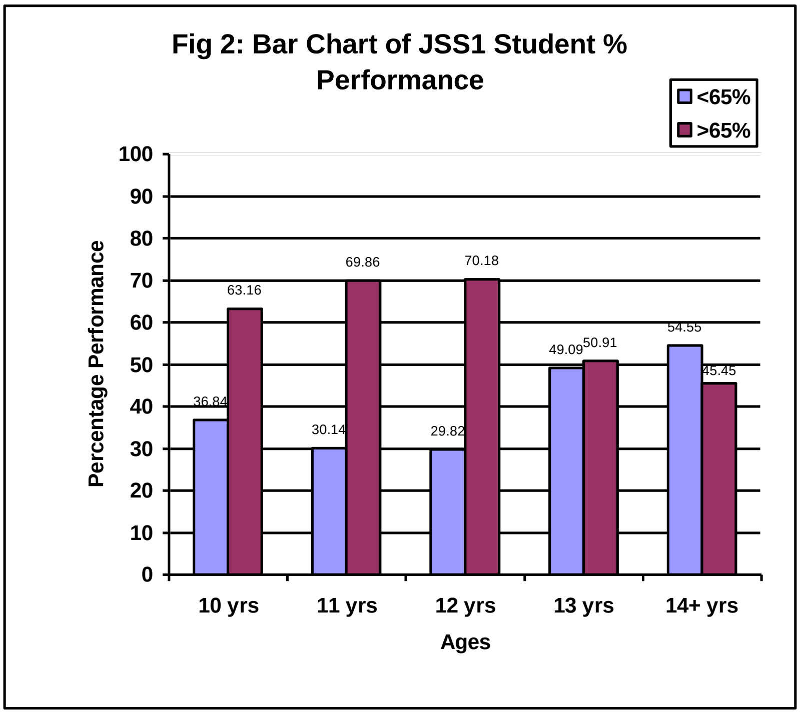
From table 7 and fig 1 below, it is pertinent to note that the best performers across all the regions were females between the ages of 10 and 14 years. The West lead this group with 23.91% pass, followed by the East 20.88%, North 18.14% and 17.39% for the south. The males within this same age groups followed suit with West having a percentage pass of 19.57%, South 16.30%, North 13.41% and East 11.78%.

The worst performers were males between 15 and 20 years of age across all regions except for the west with 13.04% below performance of females. The North lead this group with 15.55%, South 14.13% and East with 14%.

Fig 1: Bar Chart of % Student Performance of Regions



The lowest failure rates were males between 10 to 14 years of age from the West 6.52%, East 9.43%, South 10.87% with the males from the north coming second in this group with 7.62% below par performance.
This breakdown suggests that the West perform better overall than other regions as they have the lowest failure percentage (6.52%) and highest pass percentage (23.91%).



From the bar chart of the junior secondary school one students performance (fig 2), the worst performers are the students of 14 years and above with 54.55% failure. The 13 years age group follows with 49.09% failure, 10 years age group had 36.84% failure then 11 years with 30.14% and the lowest failure was the 12years age group with 29.82%.

ISSN 2059-9951

SUMMARY AND CONCLUSION

This research work was borne from the intuitive assumptions of different groups of thoughts attributing performance of students to various classifications. This work had to statistically affirm or disclaim this assumption based records gathered over a span of four years (2012/2013 academic session through to 2015/2016 academic session). The categories examined by this work were age, gender, region of origin and performance.

The chi-square analysis performed on the data gave a value of 64.934 with a degree of freedom of 3 (tabulated chi-square value at 0.5% level of significance = 12.838). This showed that performance of student was actually not independent of age, gender, region of origin of students in the school under study.

The bar chart analysis showed that there was a trend of performance of students based on region, where students between the ages of 10-14 years of age had the greater percentage performance. It also suggest that the West perform better overall than other regions as they have the lowest failure percentage (6.52%) and highest pass percentage (23.91%).

A further analysis was carried out to find out if students performance in JSS 1 was independent of age. Again the chi-square value was 10.449 and this lead to a rejection of the null hypothesis at a 5% level of significance (tabulated chi-square value = 9.488). Thus students performance was yet dependent on age for students in JSS 1.

The bar chart analysis of the JSS 1 student records showed that the best performers were students within the age range of 11 and 12 while the worst performers were students within the age group of 14 years and above.

The data was coalated with the aid of a program written by the student. It was written in php. The raw data from the database was sorted and the complete one records collected into a new database table having the required fields (date of birth, age, gender, state of origin, average performance, session and student ID) for this work. The program compiled the various counts for the different categories and computed the expected counts for each cell and the chi-square value.

RECOMMENDATIONS

From the results obtained, we would like to recommend that the school pay more attention to students above 14years of age, and to also restrict her entry age to 11 or 12 years of age for better performance records.

Though the statistical analysis (chi-square test of independence) applied in this research work does not point to the nature of dependence of the various categories. It actually sheds light on grey areas that need attention in order to decipher the reasons why students perform poorly/excellently based on category.

We would also like to suggest that further statistical analysis be performed to know the extent to which the various categories affect the performance of students in the school for informed decision making by school management. In order to boost the overall performance of the students, it will be of benefit to the school's management board to take a close look at the poor performance of the different categories considered.

REFERENCES

Delucchi, K. (1983) *The Use and Misuse of Chi-Square: Lewis and Burke Revisited*. Available at: http://img2.timg.co.il/CommunaFiles/51165628.pdf

Hsiao-Mei, wang (2009) *Comparison of the Goodness-of-Fit Tests: The Pearson Chi-square and Kolmogorov-Smirnov Tests*. Available at: http://joqm.ctu.edu.tw/Download/joqm/QM6-1/QM-0601-05--paper_proof.pdf (Accessed: 27 November 2016).

Howel D. C., "chi-square test- analysis of contingency tables, "women, vol 35, no 3, 2009, pp 28-83

Ivo, D. (2006) *Chi-Square Test Relative Risk/Odds Ratios*. Available at: http://www.stat.ucla.edu/~dinov/courses_students.dir/06/Fall/STAT13.1.dir/STAT13_notes.dir/lecture10.pdf(Accessed: 27 November 2016)

Lin, J. (2006) Analysis of a three way contingency table. Available at: http://web.ntpu.edu.tw/~cflin/Teach/Cate/06CateUEN05ThreeWayPPT.pdf (Accessed: 27 November 2016)

Lunsford M. L. & Fink A. D. D. (2010). Water Taste Test Data. *Journal of Statistics Education* Volume 18, Number 1, 1-8,12-14. Available at www.amstat.org/publications/jse/v18n1/lunsford.pdf(Accessed: 27 November 2016)

Marie D. W. (2008). Use of Chi-Square Statistic, Retrieved from http://ocw.jhsph.edu/courses/FundEpiII/PDFs/Lecture17.pdf(Accessed: 27 November 2016)

Mehta, C.R. and Patel, N.R. (2000) *Exact Tests*. Available at: http://www.sussex.ac.uk/its/pdfs/PASW_Exact_Tests.pdf (Accessed: 27 November 2016).

Minhaz, F.Z. (2004) *CHI-Squared Test of Independence*. Available at: http://pages.cpsc.ucalgary.ca/~saul/wiki/uploads/CPSC681/topic-fahim-CHI-Square.

Nduka, E.C. (2015) *Statistics Concepts and Methods*. 3rd edn. Nigeria: Clara's Prints.In-line Citation:(Nduka, 2015)

Paul Gingrich (2004). Introductory Statistics for the Social Sciences, Parts 1 and 2 (Chi-Square Test). Retrieved from http://uregina.ca/~gingrich/ch10.pdf(Accessed: 27 November 2016)

Gingrich P. (2004). Introductory Statistics for the Social Sciences, (Association Between Variables). Retrieved from http://uregina.ca/~gingrich/ch11a.pdf(Accessed: 27 November 2016).