Studying the Scientific State Of Students Using the Adjusted Residuals

SAMEERA ABDULSALAM OTHMAN[1], SHELAN SAIED ISMAEEL [2]

1. Department of Mathematics, Faculty of Educational Science , School of Basic Education, University of Duhok, Kurdistan Region-Iraq.

E-mail: samira.a@hotmail.com

2. Department of Mathematics, faculty of science, University of ZAKHO, kurdistan Region-Iraq.

E-mail: sheelansaid@yahoo.com

Abstract

The aim of this article is to apply the adjusted residuals to analysis of (two-way) contingency tables to determine the cells which affected to the significance of chi-square statistic

Keywords: adjusted residuals, contingency tables ,significance, chi-square

1 - Introduction

We can defined the contingency table by as order rectangles frequency or matrix consist of number of variables divided by number of level, for example: patients case (bad- middle- good) or (successful case and failure case) and other of variables. This table may be containing two variables. Each variable divided into numbers of levels called (contingency table) for two ways on shape matrix (r x c) number of rows (r) denoted levels the first variable and number of column (c) denoted by number of levels of the second variable when we add the third variable by (s) by levels into two ways table , the shape of table will change into three ways table, consist of (s) from two ways tables (r x c), thus when add any variable will change the dimension of table. Generally we called contingency table for more than two variables by multi- ways. We assume that we have contingency table of two ways (r x c). The null hypothesis (independent) and Alternative hypothesis dependent as follow:-

$$H_0 : P_{ij} = (P_{i.})(P_{.j})$$

$$H_1 : P_{ij} \neq (P_{i.})(P_{.j})$$

Where Pij is a frequency in the cell (i,j)and

$$p_{i.} = \sum_{j=1}^{c} p_{ij} \qquad\qquad i=1,\ldots,r$$

$$p_{.j} = \sum_{i=1}^{r} p_{ij} \qquad\qquad j=1,2,\ldots,c$$

For testing hypotheses above we used chi-square test that has the formula as:

$$x_c^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} (n_{ij} - n_{ij}^{\wedge})^2 / n_{ij}^{\wedge} \qquad \ldots\ldots\ldots\ldots(1)$$

Where $n_{ij}, n_{ij}^{\wedge}$ are the observed and expected frequencies for cell(i,j) respectively ,and

$$n_{ij}^{\wedge} = (n_{i.})(n_{.j})/n \qquad\qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots..(2)$$

Where $n_{i.}, n_{.j}$ are observed marginal frequencies, n is sum of observed frequency of table.

After calculate the value of (x^2) from the formula (1) above we compare it with the tabulated value under significant level α and degrees of freedom (r-1)(c-1)

2- Definitions:-

 Contingency table: data arranged in table form for the chi-square independence test, with r rows and c columns.

Observed frequency : the actual frequency value obtained from a sample and used in the chi-square test.

Null hypothesis: a statistical hypothesis that states that there is no difference between a parameter and a specific value or that there is no difference between two parameters.

Alternative hypothesis: a statistical hypothesis that states a difference between a parameter and a specific value or states that there is a difference between two parameters.

Degrees of freedom: the number of value that is free to vary after a sample statistic has been computed; used when a distribution (such as the t distribution) consists of a family of curves.

Level of significance: the maximum probability of committing a type I error in hypothesis testing

Independence test: a chi-square test used to test the independence of two variables when data are tabulated in table form in terms of frequencies.

3-Adjust residuals method:

During the 1978, Haberman suggested method of test after the significant (x^2) in contingency table. We can summarize by the following steps:

-Compute the value of standard error cell (i,j) by following formula:

$$e_{ij} = (n_{ij} - \hat{n_{ij}})/\sqrt{\hat{n_{ij}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots..(3)$$

-Compute the value of asymptotic variance as:

$$\hat{v_{ij}} = (1 - n_{i.}/n)(1 - n_{.j}/n) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(4)$$

-Compute the value of adjusted residual as:

$$d_{ij} = e_{ij}/\sqrt{\hat{v_{ij}}} \quad \sim N(0,1)\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(5)$$

We compare the absolute value (d_{ij}) with tabulated value for standard normal distribution by level of significant (0.05) i.e., (1.96), if the value of (d_{ij}) is greater than (1.96) it is mean that significant, which means that the relation between level(i)from first variable and (j) level from second variable is significant, and value (d_{ij}) is less than (1.96) reference to the relation between level of (i) and (j) are insignificant. We can determine the responsible for the relation between variables are significant.

5-Collection of data

We take the data during the (questionnaire) which is mentioned in the appendix (A) which is distributed the two college (Basic Education in Duhok and Education in Zakho city), we got (101) form (63) females and (38) males. We grouped the data on the two ways table by using icon (table) in the software of Minitab (13.2) as show that in the tables below:

Table(1) : relation between level study and sex of each student.

Sex / Level of student	Weak	Pass	Middle	Good	Very good	total
Male	21	12	17	10	5	63
Female	9	3	5	9	10	38
Total	30	15	22	19	15	101

Table(2) : relation between level study for student and Father's level education student.

Father's level education student / Level of student	Weak	Pass	Middle	Good	Very good	total
Ignorance	5	9	5	5	5	29
Primary	3	3	8	6	3	23
Middle	6	2	7	3	2	20
High school	6	0	1	2	2	11
University	10	1	1	3	3	18
Total	30	15	22	19	15	101

Table(3) : relation between level study for student and Mother's level education student.

Mother's level education student / Level of student	weak	Pass	Middle	Good	Very good	total
Ignorant	3	10	10	11	2	36
Primary	8	3	8	6	3	28
middle	4	1	2	0	4	11
High school	8	1	1	1	3	14
University	7	2	1	1	1	12
Total	30	17	22	19	13	101

Table(4) : relation between level study for student and economic case for the family.

Level of student / economic for the family.	weak	Pass	Middle	Good	Very good	total
Weak	6	1	2	5	6	20
Middle	7	9	14	6	4	40
Good	8	2	3	5	3	21
Very good	9	3	3	3	2	20
Total	30	15	22	19	15	101

Table (5): relation between levels studies for student and for student and is   mother still life.

Level of student / Mother still life	weak	Pass	Middle	Good	Very good	total
Yes	22	9	18	15	6	70
No	8	6	4	4	9	31
Total	30	15	22	19	15	101

Table (6): relation between levels studies for student and is father still life.

Level of student / father still life	weak	Pass	Middle	Good	Very good	total
Yes	24	10	13	14	8	69
No	6	5	9	5	7	32
Total	30	15	22	19	15	101

Table (7) : relation between level study for student and members of the family which student contain.

Level of student / members of the family	weak	Pass	Middle	Good	Very good	total
1-5	10	3	5	2	4	24
6-10	10	11	7	21	6	56
11 more	10	1	10	5	5	31
Total	30	15	22	19	15	101

Table (8): relation between level study for student and is the student work or not.

Level of student / the student work	weak	Pass	Middle	Good	Very good	total
Work	12	4	9	6	8	39
not work	17	11	13	14	7	62
Total	29	15	22	20	15	101

6-Analysis of data:

We calculate the (x^2)values for the tables above to know if there are significant relationships between levels two variables that have been studied or not, by formula (1) by using (chis) in the software Minitab and that have been sum marriages in the table (9) below.

Table (9): the chi-square values for the table (1-8)

Tables	1	2	3	4	5	6	7	8
χ_c^2	11.049*	26.304*	34.383*	18.006	9.345	4.585	13.145	3.043
χ_t^2	9.49	26.296	26.296	21.026	9.49	9.49	15.507	9.49

We note from the above table that the relation between the levels of variables in tables (1), (2), (3) significant. But the other tables are not significant. Therefore the method is explained in the formulas   from (3) to (5) will apply on these three tables (1-3) only, because we have not special icon to calculate (d_{ij}) contain software program in statistic, we write the program (1) in appendix (B) by using commands of software Minitab, to compute it . When we applied in these three tables above we obtained these tables.

Table (10) : the values of **Adjust residuals** for the table (1).

Level of student	Males	Female	$n_i^{\wedge}$	$n_j^{\wedge}$	$n_{ij}^{\wedge}$	$e_{i.}$	$e_{\cdot j}$	$v_i^{\wedge}$	$v_j^{\wedge}$	$d_{i.}$	$d_{j.}$
weak	21	9	30	19.3069	10.6931	0.38532	2.34570-	0.250564	0.452407	0.76977	3.48745-*
pass	12	3	15	9.6535	5.3465	0.75524	2.14144-	0.303500	0.547985	1.37090	2.89282*-
middle	17	5	22	14.1584	7.8416	0.75518	2.43396-	0.278796	0.503382	1.43024	3.43055*-
good	10	9	19	12.2277	6.7723	0.63707-	0.92305-	0.289383	0.522498	1.18427-	1.27697-
Very good	5	10	15	9.6535	5.3465	1.49773-	0.11153	0.303500	0.547985	2.71867*-	0.15067

**Table (11) :** the values of **adjust residuals** for the table (2).

Level of student / level of father	$\hat{v}_1$	$\hat{v}_2$	$\hat{v}_3$	$\hat{v}_4$	$\hat{v}_5$	Weak d1.	Pass d2.	Middle d3.	Good d4.	Very good d5.
Ignorant	0.501127	0.606999	0.557592	0.578767	0.606999	1.73940-	2.90254*	0.70165-	0.25631-	0.428646
Primary	0.542888	0.657583	0.604058	0.626997	0.657583	1.98962-*	0.27746-	1.71882	1.01590	0.277462-
Middle	0.563768	0.682874	0.627291	0.651113	0.682874	0.03246	0.68129-	1.59915	0.48709-	0.681294-
High school	0.626409	0.758749	0.696990	0.723458	0.758749	1.91013	1.46734-	1.08028-	0.05664-	0.329041
University	0.577688	0.699735	0.642780	0.667189	0.699735	2.64785*	1.22342-	1.83985-	0.25690-	0.238894

Table (12) : the value of **adjust residuals** for the table (3).

Level of student / level of mother	$\hat{v}_1$	$\hat{v}_2$	$\hat{v}_3$	$\hat{v}_4$	$\hat{v}_5$	Weak d1.	Pass d2.	Middle d3.	Good d4.	Very good d5.
Ignorant	0.452407	0.535242	0.503382	0.522498	0.560729	3.49771*-	2.18812*	1.08639	2.24749*	1.63388-
Primary	0.508087	0.601118	0.565337	0.586805	0.629742	0.15413-	1.01766-	1.02376	0.41674	0.40090-
Middle	0.626409	0.741104	0.696990	0.723458	0.776394	0.51214	0.72690-	0.30646-	1.69124-	2.46474*
High school	0.605529	0.716400	0.673757	0.699343	0.750515	2.42091*	1.04398-	1.42982-	1.20375-	1.03017
University	0.619449	0.732869	0.689246	0.715420	0.767768	2.31214*	0.01628-	1.20237-	0.98945-	0.50006-

Discussion and analysis the results:

1- Although the value significant (χ^2) between the scientific level of the student and his sex, there are four cells from total (10) cause cells for significant, all of them from negative direction as the table (10).

2- From the studding special table by relation between levels scientific for student and father education levels, we noticed there are three of cells responsible for the significant relation between the two variables, we see for example the father's education (primary down) there is(1) from cell responsible for significant. There is one cell their effected to minuses direct there are reaction this level from father's education with the level (pass) from student scientific level. This is also agree with scientific logical (notice table 11).

3- We notice through the table (12) the cells (or the reaction between the levels of student scientific level and the mother's scientific levels) that responsible for the relation significant between these two variables is more compare with the number of significant reaction between the levels student scientific level, and father's education levels and this mean the student scientific level the effective is more by mother's education level from father's education level.

Through the results from these searches from scientific application results to use straighten reminder method in table analysis. The researcher that is very important make this test and don't desisting from significant (χ^2).

References

Agresti, A, Lipstiz and LANG, J.B. (1992), Comparing marginal distributions Large, sparse tables, Computational Statistics and Data Analysis, 14, 55, 73.

Agresti, A, Liu, I. (1999), Modeling a categorical variable allowing arbitrarily Category choices Submitted to Biometrics.

Bishop, Y.M.M, Fienberg, S.Y. and Holland, P.W. (1975), Discrete

Multivariate Analysis: Theory and Practice', the MIT Press Cambridge Massachusetts and London.
Everitt, B.S. (1977), The Analysis of Contingency Tables, Chapin and Hall Ltd., London.
Haberman,S.J. (1978), Analysis of qualitative data vol. 1, Academic Press INC, London.

Appendix(A):

The (questionnaire) which affected the level study of student:

1.   sex of student: Male   female

2.   Father's level education student :   ignorance   Primary   Middle   High school   University

3.   Mother's level education student :   ignorance   Primary   Middle   High school   University

4.   economic case: Weak   Middle   Good   very good

5.   is father still life: yes   no

6.   is mother still life: yes   no

7.   number individual family:

8.   is student have the work :   yes   no

9.   level of the student: Weak   pass   Middle   Good   very good

Appendix (B)

The software Minitab(13.2) is used in the following macro programs.

Program(1):

gmacro

aa.macro

let k1=101

let k3= sum(c1)

```
let k4= sum( c2)

print k3,k4

do k2=1:5

let c3(k2)=c1(k2)+c2(k2)

let c4(k2)= k3*c3(k2)/k1

let c5(k2)= k4*c3(k2)/k1

let c6(k2)= (c1(k2) -c4(k2))/(c4(k2)**0.5)

let c7(k2)=(c2(k2)-c4(k2))/(c4(k2)**0.5)

let c8(k2)= (1-k3/k1)*(1-c3(k2)/k1)

let c9(k2)=(1-k4/k1)*(1-c3(k2)/k1)

let c10(k2)= c6(k2)/c8(k2)**0.5

let c11(k2)=c7(k2)/c9(k2)**0.5

enddo

endmacro

program(2):

gmacro

aa.macro

let k1=101

let k3= sum( c1)

let k4= sum( c2)

let k5=sum(c3)

let k6=sum(c4)

let k7=sum(c5)

print k3,k4,k5,k6,k7

do k2=1:5

let c6(k2)=c1(k2)+c2(k2)+c3(k2)+c4(k2)+c5(k2)

let c7(k2)= k3*c6(k2)/k1

let c8(k2)= k4*c6(k2)/k1

let c9(k2)= k5*c6(k2)/k1

let c10(k2)=k6*c6(k2)/k1

let c11(k2)=k7*c6(k2)/k1

let c12(k2)= (c1(k2) -c7(k2))/(c7(k2)**0.5)

let c13(k2)=(c2(k2)-c8(k2))/(c8(k2)**0.5)

let c14(k2)=(c3(k2)-c9(k2))/(c9(k2)**0.5)
```

let c15(k2)=(c4(k2)-c10(k2))/(c10(k2)**0.5)

let c16(k2)=(c5(k2)-c11(k2))/(c11(k2)**0.5)

let c17(k2)= (1-k3/k1)*(1-c6(k2)/k1)

let c18(k2)=(1-k4/k1)*(1-c6(k2)/k1)

let c19(k2)=(1-k5/k1)*(1-c6(k2)/k1)

let c20(k2)=(1-k6/k1)*(1-c6(k2)/k1)

let c21(k2)=(1-k7/k1)*(1-c6(k2)/k1)

let c22(k2)=c12(k2)/c17(k2)**0.5

let c23(k2)=c13(k2)/c18(k2)**0.5

let c24(k2)=c14(k2)/c19(k2)**0.5

let c25(k2)=c15(k2)/c20(k2)**0.5

let c26(k2)=c16(k2)/c21(k2)**0.5

enddo

endmacro