IMPLEMENTATION OF CONTINGENCY TABLES FOR THE EVALUATION OF ICT IN BOTH ENVIRONMENTS OF UNIVERSITIES IN TETOVO

Lazim Kamberi[1], RushadijeRamani[2] ,Shpëtim Rexhepi[3], Shpresim Kameri[4]

[1]*Faculty of Math-Natural Scuiences, State University of Tetova, Tetovo, Macedonia*
lazim.kamberi@unite.edu.mk
[2]*Faculty of Math-Natural Scuiences, State University of Tetova, Tetovo, Macedonia*
rushadie.ramani@unite.edu.mk
[3]*Faculty of Math-Natural Scuiences, State University of Tetova, Tetovo, Macedonia*
shpetim_math@hotmail.com
[4]*Faculty of Math-Natural Scuiences, State University of Tetova, Tetovo, Macedonia*
shpresim_kameri@hotmail.com

ABSTRACT

The paper describes a wide survey of the application of statistical model especially of log-linear models in the analysis of contingency tables with two entrances, which is based on the close connection multinomial distributions and Poisons on the use of Information and Communication Technologies(ICT), especially in the using computer and internet web for diplomic and post diplomic student of two universities in Tetovo (State University of Tetovo (SUT) and South East European University in Tetovo). A contingencytablesummarizes theconditionalfrequenciesof twovariables, and shows howthese variablesdepend on each other. So acontingency table can be seen as a link between the two variables according to the information data(Andersen EB, 1990; Agresti, 1996). The results of contingency tables may described by diagram (diagram with colons). The situation isstudiedto have alocal and regional viewofcomputer and Internetuseby studentsfortheirneeds. Italsopresentstheresultsofa studyof asurveymadebythestudentsoftheState University of Tetovo(SUT) and the South East European Universityin Tetovo, whichtry to measure the level of computer and Internet use an dindependence in this environment for students with application of statistical models in analysis of contingency tabels.Surveys conductedduringJune-November 2010 anddduringNovember-December 2011, showthatthespreadanduseofnewtechnologiesfeaturedseveral timesbetweenenvironmentsconditions where they study, faculties, studyprograms, generationandeconomicdevelopment country.

Keywords: ICT, contingency tables, data interpretation, digital local development, economic development.

1. INTRODUCTION

This paperdeals withthe analysis ofcategoricaldata, contingencytablesandstatisticalmodelstoexaminetheirparticularmodelinhi-square testandlogmodels-linearcontingencytableswithtwo entrances, which are based onthe close relationship ofdistributionmultinomial.The models aredescribed, and on them were appliedtodata obtainedfrom surveysconductedduring June-November 2010 overNovemberandDecember of2011 with students of two universities in Tetovo.These surveys have to do with in formation about the performance by the Information Technologies and Communications(ICT). Contingencytableswithtwo entranceshandlethe relationship betweentwoattributes(variables) according totheirfrequency, the results of whichcanbe presentedgraphicallyintheform ofabargraph.Alsoshowingstatisticalindependenceconditionsfor2x2contingencytables, thetables2xn.Resultsexpandcontingencytableswithmanyentriesanddiscussstatisticalindependence ofthematricesgrumpy.

2. CONTINGENCY TABLES WITH TWO ENTERANCE. APPLICATION

A2-entry tableoratableofrandomtwo-way hash valuesobservedsimultaneouslyprovidesbothstatisticalcategoricalvariables(qualitative, quantitative), [3].Let $(x_1,y_1),\ldots,(x_2,y_2)$dataforbothqualitativevariablesXandYrespectively. Asimple description ofthe couple (X,Y) isthe presentationofvalues $x_i$and$y_i$inthe form ofa2-input tablecalleda contingencytablewith2-entrances.Let $p_1,\ldots,p_k$ be modalities of variable X, and$q_1,\ldots,q_r$ be modalities of variable Y. Than matrix Moforder *kxr*, thatasfrequencyelementsx_{ij}, where$x_{ij}$the number of individualswhohavevalues respectively $p_i$andq_j, i=1,…,k, j=1,…,risa contingencytableof variablesXand Y, [5]. Formofacontingencytableis:

Y / X	$q_1 \ldots q_j \ldots q_r$	sum
P_1	$x_{11} \ldots x_{1j} \ldots \quad x_{1r}$	$x_{1.}$
p_i	$x_{i1} \quad \ldots \quad x_{ij} \quad \ldots \quad x_{ir}$	$x_{i.}$
p_k	$x_{k1} \quad \ldots \quad x_{kj} \quad \ldots \quad x_{kr}$	$x_{k.}$
sum	$x_{.i} \quad \ldots \quad x_{.j} \quad \ldots \quad x_{.r}$	n

Members $x_i$and$y_j$can be calculatedwithequations:

$x_{i.} = \sum_{j=1}^{r} x_{ij}$, $x_{.j} = \sum_{i=1}^{k} x_{ij}$ and $x_i y_j$are called specific effective and $n = \sum_{i=1}^{k} x_{i.} = \sum_{j=1}^{r} x_{.j}$.

APPLICATION1 In theyear2010becamea1382studentsurveyinterviewed forinternetconnectionin the environmentandsutUniversitystudy[8], [6].From thisresearchreportwe givea table whichpresents: Groupofstudentsandlinkto the Internetintheir studyenvironments.Here isabrief description of thegroups: Group I-SUT (Students of theUniversity of Tetovo), GroupIIUniversity(Students of the University ofEasternEurope).

TABLE 1.1 Classificationshowstwoentriesundertwovariables: frequency ofInternetconnectivityintheir facilitiesandgroupstudystudents, 1382studentsforover 18years.

TABLE 1.1 Frequency ofInternetconnectivityinthestudyenvironment.

InternetConnection / Student group	Every day	One totwotimes	Rarely	Never	No answer	sum
Group I-SUT	266	200	170	62	28	726
Gruop II-SEEU	304	124	150	49	29	656
Sum	570	324	320	111	57	1382

Source: Data from thesurveyconductedin June2010

TABLE 1.2 Percentageaccordingtodatarowsin table 1. Frequency ofInternetconnectivityinthestudy Environment.

InternetConnection / Student group	Every day	One totwotimes aweek	Rarely	Never	No answer	Sum
Group I-SUT	36.64	27.55	23.42	8.54	3.85	100
Group II-SEEU	46.34	18.90	22.87	7.47	4.42	100
Sum	41.24	23.44	23.16	8.04	4.12	100

Source: Data from thesurveyconductedin June2010

TABLE 1.3 Percentageaccordingtodatacolumnsof Table1.Frequency ofInternetconnectivityinthestudyenvironment.

InternetConnection / **Student group**	Every day	One totwotimes aweek	Rarely	Never	No answer	Sum
Group I-SUT	46.67	61.73	53.13	55.86	49.12	52.53
Group II-SEEU	53.33	38.27	46.87	44.14	50.88	47.47

Sum	100	100	100	100	100	100

Source: Data from thesurveyconductedin June2010

Resultsofthecontingencytablecanbe displayedgraphicallyinthe formofabargraphwith, deciding onone of the axesoxoroyoneofthe variablesXorYaxisandonthe otherthe othervariableremains. Incouplescuttingpointvalues(p_i, q_j) raisedthebarheight ofn_{ij}orrespectivepercentage.Figuresbelowshowgraphsofthe results oftablecolumns tothe table1.1.2and1.1.3[7].

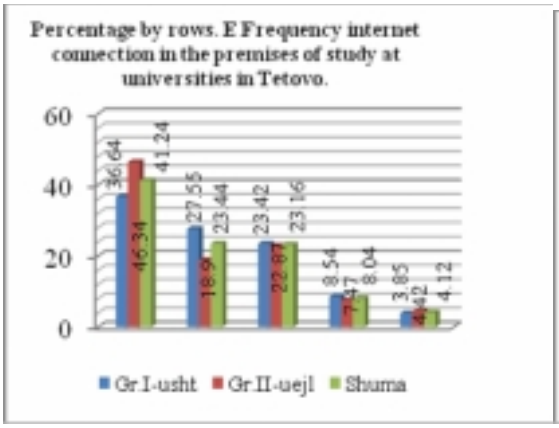

Figure1.1Diagramofbatsforthe resultstable1.2

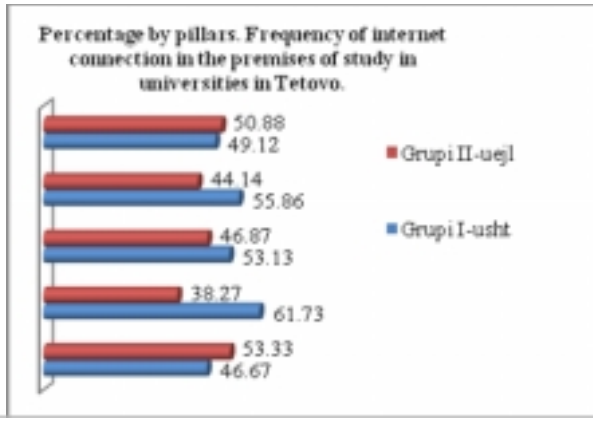

Figure1.2Diagramofbatsforthe resultstable1.3

2MODEELHI-SQUARETEST(INDEPENDENCE)

The mainhypothesisbetweentwoqualitativevariablesis whethertheyare independent. Resultsare used forquantitativevariables. LetXandYbetwoqualitativevariables.We suppose that the modalities of variable X are $p_1,\ldots,p_k$and modalities of variable Y are $q_1,\ldots,q_r$. By definition,XandYare independentif:

$$P(x=k,Y=r)=P(X=k)\cdot P(Y=r) \qquad (2.1)$$

foreach pairof valuesk andr, of XandY, respectively [5]. To checkthe hypothesis:

H_0: X and Y are independent against **H_1: X and Y are not independent**

willusehi-square testmarked withasymbolχ^2.Since:

$$P(X=k,Y=r)=\frac{x_{ij}}{n}\quad P(X=k)=\frac{x_{i.}}{n}\quad P(Y=r)=\frac{x_{.j}}{n} \qquad (2.2)$$

thenXandYare independentifwe have:

$$\frac{n_{ij}}{n}=\frac{n_{i.}}{n}\cdot\frac{n_{.j}}{n} \qquad (2.3)$$

Statisticsof the testwillbe:

$$i^2 = \sum_{i=1}^{k}\sum_{j=1}^{m}\frac{\left(n_{ij} - m_{ij}\right)^2}{m_{ij}} \qquad (2.4)$$

where,m_{ij}- isan estimateoftheexpectedfrequencyincellijandn-is the total number ofobservedvalues.

IfXandYare independent, thenχ^2hashi-square distributionwith (k-1)(r-1) degree of freedom.Hypothesis$\mathbf{H_0}$not acceptedforlargevaluesof the test$\chi^2$and converselyitis acceptedforsmall valuesof the testχ^2Inastatisticalsoftwareperformedallactionsforfindingthe valueof the coefficientχ^2, [3]. In the teststatistic(2.4) goalis to comparethe observedvalueswiththe expectedfrequencies, taking into account theindependence, and to do sobyχ^2.

- Ifχ^2it's greatwehave reason to believethattheindependencehypothesisisnottrue.
- If$\chi^2$is smallorhasamoderatevalue, wewouldconcludethattheattempt isactuallyindependent.

Followingthis applicationlet's looktosee whya comparisonbetween$x_{ij}$and$m_{ij}$is so important.

APPLICATION 2(Independencebetweenvariables**)**.Frequencyofreceptioninternetconnectionsat homeandstudentsgroupsclearly showsthatthere is a differencebetween thefrequencywith whichpeopleinthe groupI andgroupIIpersonsperformingInternetconnectionsat home. This caseis illustratedin Table2.1, wherethereceptionfrequenciesare calculated.

TABLE 2.1 Frequencyof internet connectionat home. Frequencyof waiting forinternet connectionat home.

InternetConnection / Student group	Every day	One totwotimes	Rarely	Never	No answer	Sum
Group I-SUT	430 (409.75)	125 (123.45)	91 (97.18)	50 (65.15)	30 (13.47)	726
Group II-SEEU	350 (356.02)	110 (111.55)	94 (75.73)	74 (58.07)	28 (27.53)	656
Sum	780	235	185	124	58	1382

Source: Data from thesurveyconductedin Juneof 2010.

Calculatingthe valueofhi-square statistics, foundthat:

$$i^2 = \sum_{i=1}^{k}\sum_{j=1}^{m}\frac{\left(n_{ij} - m_{ij}\right)^2}{m_{ij}} = \frac{(430 - 409.35)^2}{409.35} + \ldots + \frac{(350 - 356.12)^2}{356.12} + \ldots = 34.21$$

It seemsclearthat the frequencyof internet connectionvariesaccording togroupsofstudents. Studentsin Group IIare less likelyto haveinternet connectionathome thanstudentsin Group I, [8], [9]. Theobservedvalueof$\chi^2$is 34.21 thatisaverylargenumber.Percentage of 95 % in a distribution$\chi^2$with 5 free degree is 11.07, so value ofχ^2thatweobservedis very difficult tobefallthereforebe rejectedhypothesisofindependenceat homeinternetconnectionto be independentfromGroups IandIIstudents.

APPLICATION3 Respondentstorecordtheirquestionwasmadeforthe device witha computer at homeandgender. 1382respondentsclassifiedaccording totwocriteria: thehome computerdevicewiththreemodes: 1-Yes, 2-No: 3-NoresponseandGenderintwomodes: 1-Female: 2-Man.The resultssummarized inthe formofacontingencytable, produced withSPSS programforvariables, [10], with thehome computerdeviceandgroupsare:

TABLE3.1 The devicewiththe home computer. Frequenciesof waitinghome computer.

Device with home computer / **Groups I, II**	**yes**	**no**	**No answer**	**Sum**
SUT	407 29.45	282 20.40	37 2.68	726 52.53
SEEU	425 30.75	196 14.18	35 2.53	656 47.47
Sum	832 60.20	478 34.59	72 5.21	1382 100.00

Source: Data from thesurveyconductedin2011larger state-owned.

Here, thetable, the percentages are calculatedas a percentageof the total. But it is possibleto calculatethepercentagesbyrowsorbycolumns. The results ofthis tableis presentedgraphicallyas follows, [6], [8]

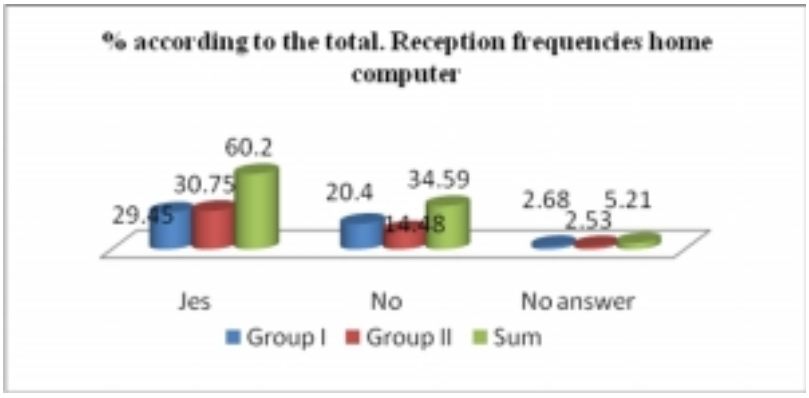

Figure3.1 Diagramofbatsforthe resultstable3.1

From Table3.1, these valuesyieldcalculations:

$$i^2 = \sum_{i=1}^{k}\sum_{j=1}^{m}\frac{\left(n_{ij}-m_{ij}\right)^2}{m_{ij}} = \frac{(407-29.45)^2}{29.45} + \frac{(282-20.40)^2}{20.40} + ... + \frac{(35-2.53)^2}{2.53} = ... = 16.67 \cdot$$

Withthehi-square valueisgreaterinthiscase, weconclude that theattempt isinfact noindependence, thenthe hypothesisH0is rejected. Statisticaldecisiontakeninthisway.Ifthe observedvalueofhi-square is such that$\chi^2 > \chi^2_{\alpha}(k)$, then thenull hypothesis is rejected. Otherwiseit is accepted.Since$\chi^2_{0,05}(2)$=5.99anditissmaller thanthe observedvalue thatis16,67, then itdoes notacceptthe null hypothesisthatthehome computerdeviceto be independentfromGroups IandII of universities.

3. RESULTSANDDISCUSSIONS

- From Table2.1, thefrequencyseems clearthatInternetconnectivityvariesaccording togroupsofstudents. Studentsin Group IIare less likelyto haveinternet connectionathome thanstudentsingroupI. Theobservedvalueofχ^2 is 34.21 which is a big number. Percentage 95 % inadistribution$\chi^2$with 5 freedom degree is 11.07, so value ofχ^2thatweobservedis very difficult tobefallthereforebe rejectedhypothesisofindependenceat homeinternetconnectionto be independentfromGroups IandII of students.

- From figure3.1 we see that:
  60,20% of Thestudentsare provided withcomputer
  30,75% of the group IIare provided withcomputer
  29,45% % of the group Iare provided withcomputer
  2.53% of the group IIarenotequippedwith computersversus2.68% ingroupIarenotequippedwith computers.
- The calculationsin Table 3.1givethese values. Withthehi-square valueisgreaterinthiscase, there is no independence, H_0hypothesisis rejectedthenaccepted. Statisticaldecision istakeninthisway. Ifthe observedvalueofhi-square is such that$\chi^2 > \chi^2_{\alpha}(k)$, then thenull hypothesis is rejected. Otherwiseit is accepted.Since,$\chi^2_{0,05}(2)$=5.99anditissmaller thanthe observedvalueis16.67, thenthenull hypothesisis rejectedbyhome computerdeviceto be independentfromGroups I, II.

4. CONCLUSIONS

On theapplication oflog-linear modelsforcontingencytablescan drawthe following conclusions:

- Contingencytablesinterpreted in terms ofthe relationship betweentwovariablesasfrequencyandstatisticalindependence
- The mainhypothesisbetweentwovariablesis whethertheyare independent. Tocheckthis usehi-square test. The hypothesisis rejectedforlargevaluesand vice versasheadmittedtosmall values.
  a)Table2.1showsthat the frequencyofreceptioninternet connectionat homeandstudent groupsshowsno differencebetween thefrequencywith whichpeopleinthe groupI andgroupIIpersonsperforminginternet connectionat home. Group IIstudentshave lessinternet connectionathome thanstudentsinGroup I.

b)thatthe observedvalueofhi-square is34.21anditisgreater thanthe value ofhi-square with5degrees of freedomwithsignificancelevelof0.05is11:07, thenrejectedthe hypothesis ofindependence, theInternetconnectionhometo be independentfromgroups IandIIstudents.
c)Table3.1showsthat the hypothesisofindependencewithhome computerdeviceandgroupsI, IInotrecognizedasthe observedvalueofhi-square is16.67which isgreater than5.99which is the valueofhi-square withtwodegreessignificancelevelofdiscretionand0:05.

5. REFERENCES

[1] A. Agresti and A. Gottard. Independence in multi-way contingency tables: S. N. Roy's breakthroughs and later developments, *Journal of Statistical Planning and Inference*, (2007).
[2] Agresti, A. (1996). An Introduction to Categorical Data Analysis.John Wiley and Sons, Inc. New York, New York, USA
[3] Andersen. Introduction to the Statistical Analysis of Categorical Data, Spring 1997.
[4] Brich, M.W. (1963). Maximum likelihood in three-way contingency tables.Jour. Royal Statist. Soc. B. 25, 220-233
[5] LlPuka. ProbabilitetetdheStatistika e Zbatuar.KoncepteThemelore.Tiranë 2010.
[6] L. Kamberi, S. Alija, Ll. Puka, ICT Diffusion and Use in student Environment in Tetovo, European Journal of Scientific Research, ISSN 1450-216X, Volume 56, No 4 (2011), pp 556-561
[7] L. Kamberi, Sh. Kameri, Sh. Ismaili, An Evaluaton of the ICT Use by the Students of Tetovo Universities in Macedonia, International Journal of Digital Content Technology and Applications, ISSN 1975-9339, Volume 5, No 8 (2011), page 101-107
[8] L. Kamberi, S. Alija, Ll. Puka. ICT diffusion and use in student environment in Tetovo, PROCEEDINGS OF 4th International Conference for Entreprenurship, Inovation and Regional Development (iEIRD'2011), Ohrid, Macedonia
[9] L. Kamberi, Sh. Kameri, Sh. Ismaili. Përdorimii TIK ngastudentët e universitetevenëTetovëpërqëllimestudimi, TakimiiPestëVjetorNdërkombëtariInstitutitAlb-Shkenca, Konferenca e SeksionittëShkencaveInxhinierikedhetëTeknologjisësëInformacionit (TPVN-IASH-KSSHITI'2011), Prishtinë, Kosovë