Statistical Analysis of Academic Level of Student in Quantitative Methods Courses by Using Chi-Square Test and Markov Chains - Case Study of Faculty of Sciences and Humanities (Thadiq) -Shaqra University-KSA

Dr. Mohamed Hassan Mahmoud Farg [1], Dr. Faiza Mohamed Hassan Khalil [2]

[1] Associate Professor, (Statistics), Faculty of Economics and Political Sciences-Omdurman Islamic University (Sudan) – Shaqra University- KSA (Secondment), mhmfaraj@gmail.com
[2] Associate Professor (Economics), Faculty of Economics and Political Sciences-Omdurman Islamic University (Sudan) – Shaqra University- KSA (Secondment), fayza09227@gmail.com

Abstract: This paper aims at predicting the trend of the student academic status, and to make improvements in the educational process, that can be obtained through the transition probabilities from one academic status to another. The paper also aims at determining whether actual frequency distribution is appropriate theoretical frequency distribution.
The study applied on population of size 98 student in faculty of science and humanities, at Shaqraa University, KSA. A Markov chain and chi-square were used for data analysis, the first was used to determine Stability of Markov matrix, and the second was used to test the association between actual and theoretical distributions. The most important results that the actual frequency distribution of transmission of student was significantly different from the theoretical one. The academic level of the student settles in semester V, and the probability for improvement in the academic level of the student is 64%, probability of the decline in the level of courses is 27 % finally; probability of stable level of courses is 9%.

[Mohamed Hassan Mahmoud Farg, Faiza Mohamed Hassan Khalil. **Statistical Analysis of Academic Level of Student in Quantitative Methods Courses by Using Chi-Square Test and Markov Chains - Case Study of Faculty of Sciences and Humanities (Thadiq)-Shaqra University-KSA. .** *Nat Sci* 2014;12(12):182-186]. (ISSN: 1545-0740). http://www.sciencepub.net/nature. 24

Keywords: Quantitative Methods, Chi-Square, Markov Chains, Shaqra University, KSA.

1- Introduction

The statistical analysis of the scores of the students is of great importance, since it is based upon making important decisions. Two methods were used in the analysis. The first one is the Chi -Square Test (CST), which was used to compare the actual frequency distribution of paper data with the theoretical distribution. The second one is Markov method for measuring of transmission probability matrix through, students' perfective "P", stability "S" and retreating "R" academic level. Also the Law of Large Numbers and the Central Limit Theorem (LNCLT) was used.

Modern probability theory studies chance processes for which the knowledge of previous outcomes influences predictions for future experiments. Basically, when we it is observe a sequence of chance experiments, all of the past outcomes could influence our predictions for the next experiment. For example, this should be the case in predicting a student's grades on a sequence of exams in a course. But to allow this much generally would make it very difficult to prove general results. In 1907, A. A. Markov began the study of an important new type of chance process. In this process, the outcome of a given experiment can affect the outcome of the next experiment. This type of process is called a Markov Chain' (Ephraim; 2002).

In this paper it is assumed that instability in levels of students in all courses in general, and in quantitative methods courses (statistics and mathematics) in particular. The phenomenon of student inconsistency in academic level is not new, and a lot of educational institutions experiencing of it. This leads to predict what the academic levels of these students would be in the future.

The importance of this paper is that, it addresses an important problem of education in general and the problems of the Shaqra University in particular, and it is summarized in volatility in students' levels among Retreating "R", Perfective "P", and Stable "S".

This paper aims at predicting the direction of the student academic status, and to make improvements in the educational process, that can be obtained through the transition probabilities from one academic status to another one. The paper also aims to determine whether actual frequency distribution is appropriate theoretical frequency distribution.

There are many studies that have used the analysis of Markov chains method. For example: First

Study of (Hussain, 2009), under the title "Using of Markov matrix in the estimation of student staying in the Faculty of Law at the University of Damascus", that reached to matrix to predict the numbers of graduates and the number of years that is expected to be spent by the student at each academic level. The second study of (Mohamed et al., 2010), entitled "The use of Markov chains to predict the prices indices of consumer in Iraq". The most important results showed that all the indices of consumer prices of basic commodities increase and then decrease, except the price index of the rent will stabilize later on. The third study is the study of (Belazzouz and Malika, 2014), under the title "The role of Markov chains in reducing the risks threatens economic institutions" and was focused on the theoretical framework. The fourth study, was of (Alkusso and Saleh, 2010), entitled "Estimating the rank of Markov chain for the case of the weather of Mosul city, using reverse spread network", it was in search of artificial neural network design a spread network reverse fault, which has been used in the estimation of the rank of Markov chain, case study of rainy months in the city of Nineveh (clear-cloudy-rainy).

2- Material and Methods

Chi-Square Distribution (CSD), is one of the most important statistical distributions; it is used in the independence test, homogeneity test, goodness of fit, and in population variance inferred; its probability density function is given by equation (1).

$$f(\chi^2) = \frac{1}{2^{\frac{n}{2}}.\Gamma\frac{n}{2}}.e^{\frac{-\chi^2}{2}}.(\chi^2)^{\frac{n}{2}-1} \rightarrow (1)$$

Where: $0 \le \chi^2 < \infty$

CSD is a non-symmetric distribution, the arithmetic mean of the distribution is equal to its degrees of freedom, and the variance is equal to its degrees of freedom multiplied by two. CSD has summation property, so that the distribution of the total independent variables, each of which has CSD is CSD with average equal to the sum of all principle CSD averages and variance equals to the sum of the principle variances, (Awda, and Alqadi, 2002).

Often we have available statistical data in the form of iterations, that can be analyzed to reach to many important characteristics of the population. Perhaps the most important of chi square test (CST) is its use in the test of independence, and it is also used in parametric and non-parametric statistics, which there is no needed to test the distribution of natural data, Abu Zeid, 2005).

If we have the variables A, B, where A takes "c" levels, and B takes "r" levels, and if, it is required to test whether the two variables A, B are independent, in this case a random sample of n units selected; in such a case represent the data in a table called Contingency Table, which has a "r" row and "c" column. The" ij" cell includes the unit that has been classified in row number "i", and column number "j". The function used to find calculated value of CST is given in equation (2):

$$\chi_c^2 = \sum_{j=1}^{c}\sum_{i=1}^{r}\left(\frac{E_{ij}-O_{ij}}{E_{ij}}\right)^2 \rightarrow (2)$$

Where: E_{ij} = expected frequency, and O_{ij} = Observed value.

The test is based on the following hypothesis.

H_0: Variables A and B are independent

H_1: Variables A and B are dependent.

The calculated CS will be compared with tabulated CS, with (r-1) (c-1) degrees of freedom and "α" level of significance. If the null hypothesis accepted and the alternative hypothesis rejected, that means A and B are independent variables. In the case of the results of SPSS analysis, the null hypothesis rejected if the level of significance is less than 0.05, ie sig <0.05, (Abu Zeid, 2005).

According to (Al Azari, and Al Wakeil, 1991), and (Bhat and Johnson, 1997), the MC is a series of cases that can be experienced by a particle moving through different time periods, based on the laws of the transition probabilities. Also it is defined as a series of random variables, in which the future status X_{n+1} is independent of previous cases $X_1, X_2, \ldots.., X_n$

Transition matrix, substitution matrix, or Markov matrix (MC) is a matrix used to describe the transitions of a Markov chain. Each of its entries is a nonnegative real number representing a probability. It has used in probability theory, statistics and linear algebra, as well as computer science and population genetics. According to Asmussen there are several different definitions and types of stochastic matrices:

A right stochastic matrix is a real square matrix, with each row summing to 1.

A left stochastic matrix is a real square matrix, with each column summing to 1, (Asmussen; 2003).

The following conventions will be adopted: random variables will be denoted by capital letters X_t and Z_t, while realizations of random variables will be denoted by corresponding lowercase letters xt and z_t. A stochastic process will be a sequence of random variables, say $\{X_t\}$ and $\{Z_t\}$, while a sample path will be a sequence of realizations $\{x_t\}$and $\{z_t\}$.

Roughly speaking, a stochastic process $\{X_t\}$ has the Markov property if the probability distributions:

$\Pr\{X_{t+1} \leq x \mid x_t; x_{t-1}; \ldots, x_{t-k}\} = \Pr\{X_{t+1} \leq x_i \mid x_t = x_i\}$

For any $k \geq 2$. A Markov chain is a stochastic process with this property and takes values in a finite set. A Markov chain (x; P; π_0) is characterized by a triple of three objects : a status space identified with an n-vector x, an n-by-n transition matrix P, and an initial distribution, an n-vector π_0.

Let

$x = [x_1 \ x_2 \ldots.. \ x_n]/$

Then the transition matrix $P = [p_{ij}]$ has elements with the interpretation:

$p_{ij} = \Pr(X_{t+1} = x_j \mid X_t = x_i)$

So, fix a row i. Then the elements in each of the j columns give the conditional probabilities of transiting from status x_i to x_j . In order for these to be well-defined probabilities, we require

$0 \leq p_{ij} \leq 1$, $i; j = 1; 2; \ldots\ldots\ldots ; n$

$$\sum_{j=1}^{n} p_{ij} = 1, \quad , i = 1,2,..,n$$

Based on (Howrad, 1971), that if p_{ij} represents the possible movement of the phenomenon from situation "i" to situation "j" in a certain period of time, and the MC contains "N" cases (where "N" is a natural number), we can formulate the transition probability according to the "P" matrix in equation (3), which is a square matrix with an order nxn, all of its elements non- negative and the sum of each row is equal to one.

$$P = \begin{bmatrix} P_{11} & P_{12} & \ldots & . & P_{1n} \\ P_{21} & P_{22} & \ldots & . & P_{2n} \\ . & . & \ldots & . & . \\ . & . & \ldots & . & . \\ P_{n1} & P_{n2} & \ldots & . & P_{nn} \end{bmatrix} \rightarrow (3)$$

The value of the possible movement of the phenomenon from status (i) to the status (j) after "m" number of steps or time periods is calculated according to equation (4).

$$P_{ij}^{m} = \{X_{n+m} = j / X_i = i\} = P_{ij}^{m} \rightarrow (4)$$

Where: P_{ij}^{m} represents the transition probabilities through (m) steps or time, under study periods. According to relationship (5), we find that for each ($n,m \in N$) there will be:

$$P^n . P^m = P^{n+m} \longrightarrow (5)$$

Where: P_{ij}^{n+m} represents the transition probabilities of the Markov chain matrix after (m+n) steps, and the element in the (i^{th}) row and (j^{th}) column of the matrix P_{ij}^{n+m} is as shown in equation(6):

$$P^{(n+m)} = \sum_{h=0}^{\infty} P_{ih} P_{hj} \longrightarrow (6)$$

$$\forall n, m > 0$$

Equation (6) is (Chapman Kolommogrov) equation, which is the method for calculating the n steps of the transition probability (P_{ij}^{n}).

Stability and Steady Status of the Markov Matrix is considered in this paper. Stability means not change the statistical characteristics of the stochastic process over time, and MC with a discrete or continuous time will be homogeneous in time, if the transition probabilities do not depend on the time difference. Ie a Markov chain { X_t } is said to be time homogeneous if : $P\{X_{s+t} = j / X_s = i\}$is independent of s. When this holds, putting s=0, gives $P\{X_{s+t} = j / X_s = i\} = P\{X_t = j / X_0 = i\}$. Through the application we can get the transition probabilities, through n of moves by multiplying the matrix of transition probabilities itself n of times, and MC matrix can represented similar to the equation (7)[2],[11],[12].

$$\hat{P} = \begin{bmatrix} P_0 & P_1 & P_2 & . & . \\ P_0 & P_1 & P_2 & . & . \\ . & . & \ldots & . & . \\ . & . & \ldots & . & . \\ P_0 & P_1 & P_2 & . & . \end{bmatrix} \rightarrow (7)$$

It is noted that all the rows of the matrix P^ are identical, and P^ =P^m for all values of $m \geq 1$. And the matrix P^ represent the steady status of transition probability matrix, where the probability of movement settle each case at a certain value. This matrix appears when the stochastic process continues for a long time. Stable distribution of the stochastic process can be obtained, when the transition probability matrix of Markov chain satisfies that:

$$\lim_{m \to \infty} P^m = U = \begin{bmatrix} u \\ u \\ . \\ . \\ u \end{bmatrix} \rightarrow (8)$$

Where U is unique vector, that characterized by the following:

$$U = [u_1, u_2, ..., u_n], \quad \sum_{i-1}^{m} u_i = 1, \quad 0 \le u_i \le 1$$

$$UP = U \to (9)$$

In equation (9), it is found that U is stable Markov chain distribution, ie P is stable matrix.

Two quantitative methods tests were conducted to 98 students (all population) in the College of Science and Humanities Studies at Shaqraa University (Thadiq Branch) in KSA, during the academic year 1434/1435 AH. According to the scores of students in two tests, the students were classified. Classification of the initial situation of the academic status of the student was based on the arithmetic mean of the first test, a student who earned a degree equal to the average arithmetic, classified in stable "S" condition, student who received the higher of the arithmetic average classified status as perfect "P", finally a student who earned a degree lower than the arithmetic mean is classified as retreating "R". The student who earned a degree in the second test, greater than the degree that in the first test was classified as Perfective "P. The student who earned a degree in the second test equal to the degree in the first test was classified in a stable situation " S ". Finally the student who earned a degree in the second test less than that in the first test was classified in a retreating situation "R". Based on this classification, the transition matrix has been configured in table (1) as follow:

Table (1): Number of students who move from mode to another.

	The mode to which the student moved				
		P	R	S	TOTAL
Condition of the student	P	31	18	8	57
	R	29	7	0	36
	S	4	1	0	5
	TOTAL	64	26	8	98

Firstly: Testing the frequency distribution of the students' transition

From table (3), it is clear; there is variation between the actual distribution and the theoretical distribution of the mode of students' transition, through the value of the level of significance (sig < 0). According to table (2), it is found that the transition situations are concentrated in two cases, the first case for the stability with respect to the good levels, and the second is the improvement with respect to declining levels.

Table (2): Contingency table of observed and expected frequency distribution of the academic status of students

			Column			**Total**
			Perfe ctive "P"	**Sta ble "S"**	**Retre ating "R"**	
Row	"P"	Observed	31	**29**	**4**	**64**
		Expected	37.2	**23.5**	**3.3**	**64.0**
	"S"	Observed	18	**7**	**1**	**26**
		Expected	15.1	**9.6**	**1.3**	**26.0**
	"R"	Observed	8	**0**	**0**	**8**
		Expected	4.7	**2.9**	**.4**	**8.0**
Total		Observed	57	**36**	**5**	**98**
		Expected	57.0	**36.0**	**5.0**	**98.0**

Table 3: Chi-Square Test for Frequency Distribution of the academic status of the Student

Chi-Square Tests			
	Value	df	Asymp. Sig0. (2-sided)
Pearson Chi-Square	9.552[a]	4	.049
N of Valid Cases	98		
a. 5 cells (55.6%) have expected count less than 5. The minimum expected count is .41.			

Secondly: Determination of the probability of a student move from mode to another one.

- The primary condition of the system obtained by dividing the number of students in each mode to the total number of the population size, ie dividing the totals of rows by 98. To get the primary vector in the system $a^{(0)}$ as follows:

$a^{(0)}$ = (0.58163 0.36735 0.05102)

The transition probability matrix may be obtained from table (1) by dividing each number in each row of the matrix by the total of its row, this is shown in table (4).

Table (4): The probability of transition of student from mode to another one.

The mode to which the student moved				
	P	R	S	TOTAL
P	0.54385	0.31580	0.14035	1
R	0.80556	0.19444	0	1
S	0.8	0.2	0	1

The transition probability matrix of the Markov chain "P", can be written as follows:

$$P = \begin{bmatrix} 0.54365 & 0.31580 & 0.14035 \\ 0.80556 & 0.19444 & 0.00000 \\ 0.80000 & 0.20000 & 0.00000 \end{bmatrix}$$

Since the Markov chain does not depend upon the past, therefore the transition probability matrix of the Markov chain in the second semester (P^2) is as follows:

$$P^2 = \begin{bmatrix} 0.6624487 & 0.2612220 & 0.0763293 \\ 0.5947369 & 0.2922028 & 0.1130603 \\ 0.5961920 & 0.2915280 & 0.1122800 \end{bmatrix}$$

The transition probability matrix of the Markov chain in the 3rd semester (P^3) is as follows:

$$P^3 = \begin{bmatrix} 0.6397036 & 0.2716281 & 0.0886683 \\ 0.6351721 & 0.2737011 & 0.0911268 \\ 0.6352695 & 0.2736565 & 0.0910740 \end{bmatrix}$$

The transition probability matrix of the Markov chain in the 4th semester (P^4) is as follows:

$$P^4 = \begin{bmatrix} 0.6380796 & 0.2723710 & 0.0895494 \\ 0.6380593 & 0.2723803 & 0.0895604 \\ 0.6380597 & 0.2723801 & 0.0895602 \end{bmatrix}$$

The transition probability matrix of the Markov chain in the 5th semester (P^5) is as follows:

$$P^5 = \begin{bmatrix} 0.6380722 & 0.2723744 & 0.0895534 \\ 0.6380722 & 0.2723744 & 0.0895534 \\ 0.6380722 & 0.2723744 & 0.0895534 \end{bmatrix}$$

Matrix P^5 is the steady status of the Markov matrix, i.e., in semester 5 and above, the transition probability matrix will be Stable. Therefore the vector " U_j " can be written as:

U_j =[0.6380722 0.2723744 0.0895534]

According to the last vector " U_j ", the probability of improvement of student's status is equal to 0.6380722, the probability of declining of student's status is equal to 0.2723744, and the probability of the student's status to be remaining in stable condition is equal to 0.0895534.

3- Results:

There are three important finding:

The first one, there is significant difference between the theoretical and actual distributions of the students' academic transition mode.

The second one is the academic level of the students will settle in the fifth semester.

The third one is the possibility of improving the level of student in quantitative methods courses is a 65%, probability of the decline in the same courses is a 27%, and the probability of being in stable level is 9%.

4- Discussions

In consequence of the above mentioned results, the following points discussed:

To conduct similar studies to other courses.

To take the advantages of this study in the planning, and improvement of education process.

Economically it needs budget for 27 % of the student to improve their declined academic level.

Acknowledgement:

Foundation item: Saqraa University, KSA. Authors are grateful to the faculty of science and humanities (Thadiq branch) for data support to carry out this work.

References

1. Abu Zeid, Mohamed Khair Saleim, "Using Statistical Analysis Methods of SPSS Software ", Dar Al Safa for printing, publishing and distribution, Amman, Jordan, ed1, p. 195, 2005.
2. Al Azari, Faris Muslim, and Al Wakeil, Ali Abdul-Hussein, "Stochastic Process", University Press of Mosul, Iraq, 1991.
3. Alkusso, Ibtihaj Abdul Hamied, and Saleh, Israa Abdel-Gawad, " Markov Chain Order Estimation to the Weather Status of Mosul City by Using Backpropagation Network " Iraqi Journal of Statistical Science, No. 18, Baghdad, Iraq , 2010, pp. 133-154.
4. Asmussen, S. R. "Applied Probability and Queues". Stochastic Modelling and Applied Probability", 51. pp. 3–8, 2003.
5. Awda, Ahmad A. Abdul Majied and Alqadi, Abdul Rahman Mansour, "Statistical Inference", Vol. 2, Falah Library for Printing and Distribution, Kuwait, p. 45, 2002.
6. Bhat Tachagya, GK, and Johnson, RA, "Statistical Concepts and Methods", John Wiley and Sons, New York- USA. 1997.
7. Ephraim,Y., Merhav, N., Hidden Markov processes. IEEE Trans. Inf. Theory 48 (6), 1518–1569, 2002.
8. Howard, R. A., Dynamic Probabilistic Systems, Volume I: Markov Models, Wiley, New York, (1971).
9. Hussain, Abdul Karim Mohammed, " Using of Markovian matrix in estimating students studying time in the Faculty of Law at the University of Damascus,", Damascus University, Journal of Economics and Law sciences, Vol 25, No. 1, 2009 .
10. Kemeny, John and Snell, Laurie, " Finite Markov Chains", New *York:* Van Nostrand Reinhold, 1960.
11. Mohammed, Siham Kamil, et al., "Predict the indices of consumer prices in Iraq by Markov Chains", Diyala Journal of Pure Science, Vol. 6, No. 1, Iraq, June 2010.
12. Sirl, D. "Markov Chain, An introduction review", University of Queensland, Australia, htt//www.math.uq.edu.au/dsirl.15/12/2014.
13. www. Usr / hosting_files4 / main / previews / doc378089652 / preview/ Belazzouz and Malika, "The Role of Markov Chains in Reducing the Risks that Threaten the Economic Institutions 1/12/2014.

12/21/2014