International Journal of Advanced Biotechnology and Research (IJBR)
ISSN 0976-2612, Online ISSN 2278–599X,
Vol-8, Special Issue-3, 2017, pp263-270
http://www.bipublication.com

Research Article

Interpolating Data in Transition Probability Matrix of Markov Chain to Improvise Average Length of Stay

Rahela Abd Rahim[1] and Fadhilah Jamaluddin[2]

[1]School of Quantitative Sciences, College of Arts and Sciences,
Universiti Utara Malaysia, 06010 Sintok, Kedah; rahela@uum.edu.my
[2]School of Quantitative Sciences, College of Arts and Sciences,
Universiti Utara Malaysia, 06010 Sintok, Kedah; missfadhilah04@gmail.com

ABSTRACT

Data interpolation is proposed for estimating transition probability matrix (TPM) of Markov chain model. We showed that interpolated estimator was unbiased. To show its applicability the model on the manpower recruitment policy is developed and analyzed on Excel spreadsheet. Based on the model, the new estimation of the state transition matrix for each category of manpower driven by interpolation technique is devised. The revised transition matrix of Markov chain was substituted by embedding interpolation and can be used as an equation solver to calculate mean time estimation for each category of manpower. The model results were then compared to the classical Markov chain for both old and new policies by means of mean time estimation. Two scenarios were considered in the study; scenario 1 was based on historical data pattern in five years and scenario 2 was based on the new policy. The results showed the possibility average length of stay by position and probability of loss for both scenarios. The proposed data interpolation based TPM approach has shown a new way of recruitment projection for policy changes. The results have indicated better estimation of average length of stay for each category compared to the traditional Markov chain approach.

Keywords*:* Interpolation, Markov Chain Model, Transition Probability Matrix, Average Length of Stay

INTRODUCTION

Markov chain (MC) model is a popular mathematical model used not only to see the flow of data using the stochastic process but also to forecast the data for short-term period. It is being used in various fields especially in education. Many researches focus on the the use of MC model for the purpose of examining the data flow(Rahim, 2015; Rahim, Ibrahim Mat Kasim and Adnan, 2013). The forecast value can be obtained from the Transition Probability Matrix (TPM), the most important characteristic in MC model. TPM is defined as a matrix where the probability of the system being in a given state in a particular period depends only on its state in the preceding period and it is independent of all earlier periods. Therefore, the historical data are not taken into consideration in developing TPM. In order to predict the forecast value, we need to project the TPM. Normally, the projected TPM is obtained from the multiplication of the recent TPM by itself and this process is known as smoothing. This process is used to smooth a data set by creating an approximating function that attempts to capture important patterns in the data. Smoothing is the most commonly used time series techniques for removing noise from the underlying data to help reveal the important features and components such as trend and seasonality. In addition, it can be used to fit in missing values and to conduct a forecast. Interpolation is a mathematical smoothing technique widely used in various ways. It is a process of finding and evaluating a function whose graph goes through a set of given points. It is originally used in order to do interpolation using tables by defining common mathematical functions. Nowadays, interpolation is applied in the related problem of extending functions that are known only at a discrete set of points, and such problems occur frequently when numerically solving differential and integral equations. Interpolation is commonly used in estimating missing data. However, in this study, we have done a different approach. The concept of estimating the missing data by interpolation will be used in estimating the transition probability of states in Markov chain, in order to obtain a more accurate and unbiased estimator. The proposed technique will be discussed further in the following section. Recently, studies in Markov chain model have been on improving the forecasting ability by hybridising it with other potential method such as in Rahim, Ibrahim, Nadhar Khan and Saad, 2016; 2015; Dindarloo, Bagherieh, Hower, Calder, and Wagner, 2015; Fellows, Rodriguez-Cruz, Covelli, Droopad, Alexander, and Ramanathan, 2015; Jamal, Marco, Wolfgang, and Ali, 2013). The hybrid approaches are all meant to supplement the Markov chain capabilities in data flow forecasting. Therefore, in this study, we integrate the interpolation technique in the Markov chain model for modeling manpower flow in order to identify the recruitment and promotion behavior for academic staff in higher learning institutions.

The Method

To demonstrate how the proposed method is applied, secondary data from one university are obtained. The data are collected from the registrar office at the university, the detail of academic staff such age and status ranks are used in the model design. The data contains about 1033 individuals and for each of the user there are three categories of rank and five sub-categories of age interval. The data are categorized as A (Lecturer), B (Senior Lecturer), and C (Associate Professor),

$$M = \{s_1 = 22-27yrs, s_2 = 28-33yrs,$$
$$s_3 = 34-39yrs, s_4 = 40-45yrs, s_5 = 46-51yrs\}.$$

Let $\{C_t; t - 1,2,\dots\}$ be the state at time t taking values in the state space $M - \{s_1, s_2, s_3, s_4, s_K\}$,

$$P(C_t = s_i / C_{t-1} = s_j) = q_{ij} \qquad [1]$$

Where $Q = \{q(s_i \backslash s_j; i,j = 1,2,\dots,K$ and satisfy $q(s_i \backslash s_j) \geq 0, i,j = 1,\dots,K$ and $\sum_{i=1}^{K} q(s_i \backslash s_j) = 1,$ $\forall j - 1,\dots,K$.

Then the state vector

$$X^{(n+1)} = Q\, X^{(n)}$$

Our method used Lagrange interpolation to estimate Q,

that is the transition matrix of the state $\{X^n\}$. Given $\{X^n\}$, we can calculate the transition frequency f_{li} from state *l* to state *j*. Hence the transition matrix for the state $\{X^n\}$ can be constructed as follows

$$F = \begin{pmatrix} \widehat{f_{11}} & \cdots & \cdots & \widehat{f_{1m}} \\ \widehat{f_{21}} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{f_{m1}} & \cdots & \cdots & \widehat{f_{mm}} \end{pmatrix}$$ Let

From

$$F^i, Q_i = \left[q_{ij}^{(i)}\right]$$

Where $\widehat{f_{ij}} = \sum_{i=1}^{m}\sum_{i=1}^{m} L_{ij}(x) f_{ij}$ and $L_{ij}(x) - \prod_{j=0}^{n} \frac{x-x_j}{x_i-x_j}$, $n = mxm$

Where $\widehat{f_{ij}} = \sum_{j=1}^{m}\sum_{i=1}^{m} L_{ij}(x) f_{ij}$ and $L_{ij}(x) = \prod_{j=0}^{n} \frac{x-x_j}{x_i-x_j}$, $n = mxm$

Where $\widehat{f_{ij}} = \sum_{i=1}^{m}\sum_{i=1}^{m} L_{ij}(x) f_{ij}$ and $L_{ij}(x) = \prod_{j=0}^{n} \frac{x-x_j}{x_i-x_j}$, $n = mxm$

And transition probability matrix for the state $\{X^n\}$ is given by

$$Q = \begin{pmatrix} q_{11} & \cdots & \cdots & q_{1m} \\ q_{21} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ q_{m1} & \cdots & \cdots & q_{mm} \end{pmatrix}$$

where $q_{lj} = \begin{cases} \frac{\widehat{f_{lj}}}{\sum_{l=1}^{m} \widehat{f_{lj}}} & if \sum_{l=1}^{m} \widehat{f_{lj}} \neq 0 \\ 0 & \text{otherwise} \end{cases}$ (i)

[i] $Q = \begin{pmatrix} q_{11} & \cdots & \cdots & q_{1m} \\ q_{21} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ q_{m1} & \cdots & \cdots & q_{mm} \end{pmatrix}$

where

$$q_{lj} - \begin{cases} \frac{\widehat{f_{lj}}}{\sum_{l=1}^{m} \widehat{f_{lj}}} & if \sum_{l=1}^{m} \widehat{f_{lj}} \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

The following proposition shows that the proposed interpolated estimators are unbiased.

Proposition: The estimators in (i) satisfy

$$E(f_{lj}) = q_{lj} E \sum_{j=1}^{m} f_{lj}$$

Ot eps, we obtain a sequence of events $q_{i1} \rightarrow q_{i2} \rightarrow q_{i3} \rightarrow \cdots \rightarrow q_{in}$. Let T be the length of sequence, $[q_{ij}]$ be the transition probability matrix and $\hat{X}_l$ be the steady state probability that the process is in state *l*. Then we have

$$E(f_{lj}) - T.\hat{X}_l\, q_{lj}$$

$$E\left(\sum_{j=1}^{m} f_{lj}\right) = T.\hat{X}_l \sum_{j=1}^{m} q_{lj}$$

$$E\left(\sum_{l=1}^{n}\sum_{j=1}^{m} f_{lj}\right) = T.\hat{X}_l \sum_{l=1}^{n}\sum_{j=1}^{m} q_{lj}$$

Let $\sum_{j=1}^{m} f_{lj} = F_l$

$Q = \begin{pmatrix} q_{11} & \cdots & \cdots & q_{1m} \\ q_{21} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ q_{m1} & \cdots & \cdots & q_{mm} \end{pmatrix}$ where $q_{lj} = \begin{cases} \frac{\widehat{f_{lj}}}{\sum_{l=1}^{m} \widehat{f_{lj}}} & if \sum_{l=1}^{m} \widehat{f_{lj}} \neq 0 \\ 0 & \text{otherwise} \end{cases}$ (i)

Then from (iv)

$$E\left(\sum_{l=1}^{n} F_l\right) = T.\hat{X}_l \left(\sum_{l=1}^{n} 1\right)$$

$$n.\,E(F_l) - T.\hat{X}_l.n$$

$$E(F_l) = T.\hat{X}_l$$

$$(f_{ij}) = q_{ij}E(F_i)$$
$$E(f_{ij}) = q_{ij}E(\sum_{j=1}^{m} f_{ij})$$

Hence satisfy that the proposed interpolated estimators are unbiased.

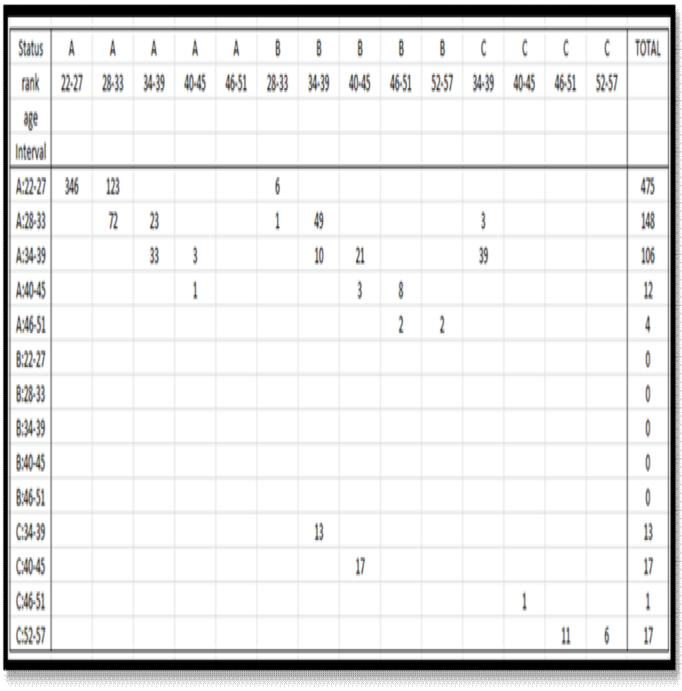
$$(f_{1j}) = q_{1j}E(F_1)$$
$$E(f_{1j}) = q_{1j}E(\sum_{i=1}^{m} f_{1j})$$

$$(f_{lj}) = q_{1j}E(F_i)$$
$$E(f_{1j}) = q_{1j}E(\sum_{i=1}^{m} f_{1j})$$

$$(f_{1j}) - q_{1j}E(F_1)$$
$$E(f_{1j}) = q_{1j}E(\sum_{i=1}^{m} f_{1j})$$

Data Collection and the Markov Chain Model

We have prepared the worksheet to demonstrate the proposed method. The worksheet displays the states transition matrix of 5 years transition of staff categorized by rank and age as shown in figure 1. Two alternative scenarios were proposed in this study. The first scenario considers the policy of promoting the staff remains the same. Therefore, the transition probabilities developed from the previous five years of data would hold. The second scenario considered the new policy suggested that the recruitment of staff type A is decreased by 10% while staff type B and C is increased by 70%. In this model, the transition probabilities would change according to the respective change in the number of staff recruited

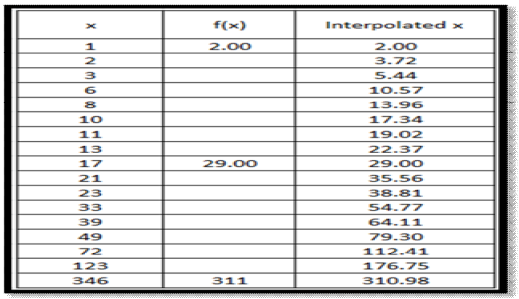


Status	A	A	A	A	A	B	B	B	B	B	C	C	C	C	TOTAL
rank	22-27	28-33	34-39	40-45	46-51	28-33	34-39	40-45	46-51	52-57	34-39	40-45	46-51	52-57	
age															
Interval															
A:22-27	346	123				6									475
A:28-33		72	23			1	49				3				148
A:34-39			33	3			10	21			39				106
A:40-45				1				3	8						12
A:46-51									2	2					4
B:22-27															0
B:28-33															0
B:34-39															0
B:40-45															0
B:46-51															0
C:34-39							13								13
C:40-45								17							17
C:46-51												1			1
C:52-57													11	6	17

Fig. 1: States transition diagram of scenario 1

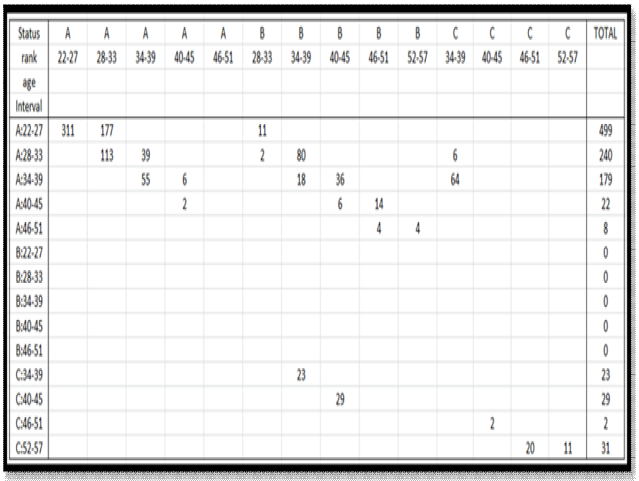
The data frequency in the transition probability matrix of scenario 1 is arranged ascendingly and the minimum, median and maximum data are selected. The data is then refered to which category of staff they are in and later assigned to a new frequency as stated by the rule in scenario 2. For example, the median is 17 and falls on state B(40-45). The new policy indicated that the category should be raised by 70%. Therefore the median is assigned to new data frequency, 29. Based on the observed frequency and the three assigned frequency in Table 1, a new set of interpolated data are generated using quadratic interpolation formula as in Figure 2. Later, the obtained interpolated data are arranged accordingly to their states and a new states transition matrix is developed as shown in Figure 3.

Table 1: Interpolated state of scenario 2



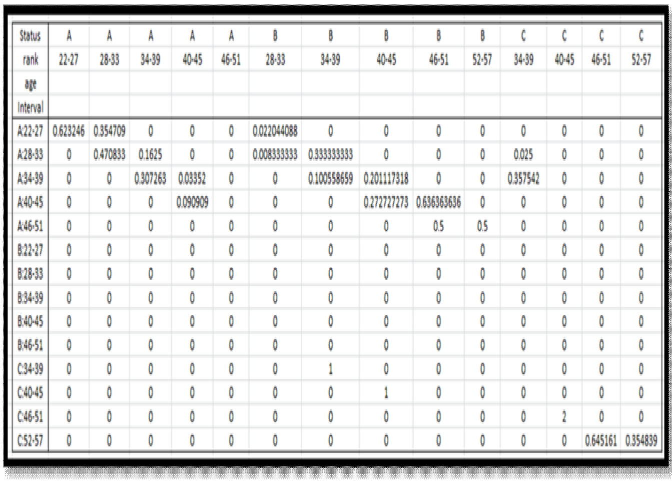
x	f(x)	Interpolated x
1	2.00	2.00
2		3.72
3		5.44
6		10.57
8		13.96
10		17.34
11		19.02
13		22.37
17	29.00	29.00
21		35.56
23		38.81
33		54.77
39		64.11
49		79.30
72		112.41
123		176.75
346	311	310.98

Based on Quadratic interpolation for the assigned data, the following equation is obtained, $f(x) = -0.002407x^2 + 1.730826x + 0.27158$ This equation is used to obtain all other interpolated data and a new TPM is formed using these interpolated data as shown in Figure 2.



Status	A	A	A	A	A	B	B	B	B	B	C	C	C	C	TOTAL
rank	22-27	28-33	34-39	40-45	46-51	28-33	34-39	40-45	46-51	52-57	34-39	40-45	46-51	52-57	
age															
Interval															
A:22-27	311	177				11									499
A:28-33		113	39			2	80				6				240
A:34-39			55	6			18	36			64				179
A:40-45				2				6	14						22
A:46-51									4	4					8
B:22-27															0
B:28-33															0
B:34-39															0
B:40-45															0
B:46-51															0
C:34-39							23								23
C:40-45								29							29
C:46-51												2			2
C:52-57													20	11	31

Fig. 2: Interpolated states transition matrix categorized by state such as age interval and status



Status	A	A	A	A	A	B	B	B	B	B	C	C	C	C
rank	22-27	28-33	34-39	40-45	46-51	28-33	34-39	40-45	46-51	52-57	34-39	40-45	46-51	52-57
age														
Interval														
A:22-27	0.623246	0.354709	0	0	0	0.022044088	0	0	0	0	0	0	0	0
A:28-33	0	0.470833	0.1625	0	0	0.008333333	0.333333333	0	0	0	0.025	0	0	0
A:34-39	0	0	0.307263	0.03352	0	0	0.100558659	0.201117318	0	0	0.357542	0	0	0
A:40-45	0	0	0	0.090909	0	0	0	0.272727273	0.636363636	0	0	0	0	0
A:46-51	0	0	0	0	0	0	0	0	0.5	0.5	0	0	0	0
B:22-27	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B:28-33	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B:34-39	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B:40-45	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B:46-51	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C:34-39	0	0	0	0	0	0	1	0	0	0	0	0	0	0
C:40-45	0	0	0	0	0	0	0	1	0	0	0	0	0	0
C:46-51	0	0	0	0	0	0	0	0	0	0	0	2	0	0
C:52-57	0	0	0	0	0	0	0	0	0	0	0	0	0.645161	0.354839

Fig. 3: Transition probability matrix of interpolated data categorized by state such as age interval and status

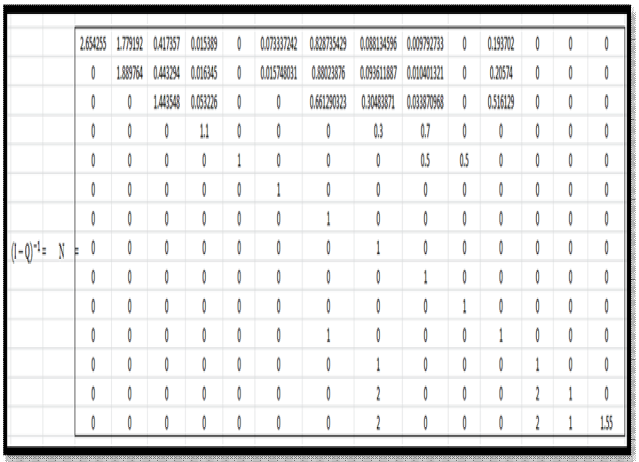
The following Markov chain model was developed using TPM of interpolated data; $n(t) = Z(t) + D(t)$ where, $n(t)$: k dimensional column vector whose i^{th} element represents the number of l staff in state i at time t. $Z(t)$: k x k square matrix whose ij^{th} element represents the number staff were in state i at time t but are in state- j at time $t+1$. $D(t)$:$k \times 2$ matrix whose ij^{th} element represents the number of manpower in the i^{th} transient state at time t. k : the number of transient states in the model. Matrix of transition probabilities, $Q(t) = \tilde{n}^{-1}(t)Z(t)$ where, Q (t) : ($k \times k$) matrix whose ij^{th} element represents the probability of a manpower making the transition from state- i at time t to state- j at time $t+1$. $\tilde{n}(t)$: diagonal matrix whose diagonal elements are the elements of $\tilde{n}(t)$.Average Length of Stay for a staff is given by $N = (I-Q)^{-1}$ where, I : the identity matrix. N : the mean time the staff stay in the same state or move to the next state.

THE RESULTS

Matrix transition probability for data distribution of five consecutive years was used to project the manpower distribution. The assumption was the historical patterns for that five years will continue if policy of manpower recruitment remains the same. The transition probability matrix indicates the probability that a manpower will move from one state to another within one period (5 years). The important use of Markov Chain is to predict future manpower distributions if there is policy changes in the current policy. As stated earlier, regarding the new policy, which is the recruitments of type tutor and lecturer will be reduced by 10% and the appointment of senior lecturer, associate professor and professor will be increased by 70%, a new formation of matrix transition diagram is realized. We consider this policy change as scenario 2 and the old policy as scenario 1.

Additionally, we proposed quadratic interpolation to predict the expected probability for each transition value so that it can be used as a guideline to monitor the transition of each state as given in Figure 4. Estimation of average length of stay, N for each category of staff is calculated using inverse matrix operation given by

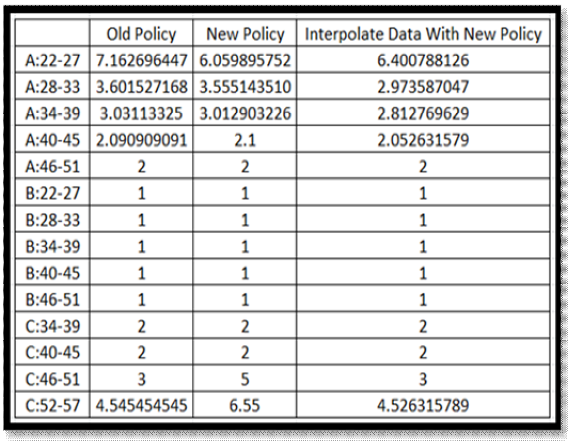
$$N = (I-Q)^{-1}$$



	2.654255	1.779192	0.417357	0.015389	0	0.073337242	0.828735429	0.088134596	0.009792733	0	0.193702	0	0	0
	0	1.889764	0.443294	0.016345	0	0.015748031	0.88023876	0.093611887	0.010401321	0	0.20574	0	0	0
	0	0	1.443548	0.053226	0	0	0.661290323	0.30483871	0.033870968	0	0.516129	0	0	0
	0	0	0	1.1	0	0	0	0.3	0.7	0	0	0	0	0
	0	0	0	0	1	0	0	0	0.5	0.5	0	0	0	0
	0	0	0	0	0	1	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	1	0	0	0	0	0	0	0
$(I-Q)^{-1}=$ N =	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	1	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	0	0	0	0	0	0	1	0	0	0	1	0	0	0
	0	0	0	0	0	0	0	1	0	0	0	1	0	0
	0	0	0	0	0	0	0	2	0	0	0	2	1	0
	0	0	0	0	0	0	0	2	0	0	0	2	1	1.55

Fig. 4: Average length of stay of interpolated data states for each level of position

Finally, the total estimation of average length of stay, for each state category is obtained by matrix multiplication NB, where B is the total frequency of each state in the state transition matrix. The average length of stay for each category of staff is calculated for the three scenario, which is scenario 1: the distribution of staff faculties remain as the past 5 years and scenario 2: the percentage of certain staff faculties has been changed with respect to the new policy while others remain, scenario 3: the percentage of certain staff faculties has been changed with respect to the new policy while others are interpolated with respect to minimum, median and maximum data of scenario 2. The analysis is done by comparing the average length of stay between the three categories of states.



	Old Policy	New Policy	Interpolate Data With New Policy
A:22-27	7.162696447	6.059895752	6.400788126
A:28-33	3.601527168	3.555143510	2.973587047
A:34-39	3.03113325	3.012903226	2.812769629
A:40-45	2.090909091	2.1	2.052631579
A:46-51	2	2	2
B:22-27	1	1	1
B:28-33	1	1	1
B:34-39	1	1	1
B:40-45	1	1	1
B:46-51	1	1	1
C:34-39	2	2	2
C:40-45	2	2	2
C:46-51	3	5	3
C:52-57	4.545454545	6.55	4.526315789

Fig. 6: Comparative average length of stay between scenario 1: Old Policy, scenario 2: New policy and scenario 3: Interpolated data with new policy

Figure 6 shows the comparative results between the average length of stay value for each category using classical TPM of Markov chain and the proposed interpolated TPM of Markov chain for the old and new policy. The results for interpolated TPM has shown a better estimates of average mean queue length of stays by status for states towards a higher range of age and status.

CONCLUSIONS

Based on the results obtained, the Markov chain model developed in this study is an appropriate evaluation tool for policy change concerning the appointment of faculty. This paper demonstrates that if new policy is implemented, there will be a high impact on the number of academic staff by diverse rank especially towards more senior faculty members. Mean time for the faculty remains in current state does not show much difference between old and new policy. Based on the results, it is predicted the proposed policy will not have much changes if it were to be implemented. Otherwise a modified predicted approach is required such as interpolated data estimators approach as proposed in this study. This paper presents the potential use of interpolation technique for predicting better estimate for projecting transition data in Markov chains. The states transition of staff faculties categories can easily be constructed using spreadsheet and the calculation of matrix performance can be done simply using excel built-in function.

ACKNOWLEDGEMENT

First and foremost, we would like to extend our sincere thanks to Ministry of Science, Technology and Innovation, and UUM for providing this research grant (s/o code: 12905).

REFERENCES

1. S. R. Dindarloo, Bagherieh, J.C.A. Hower, J. H. Calder, and N. J. Wagner. (2015). Coal Modeling Using Markov Chain and Monte Carlo Simulation: Analysis of Microlithotype and Lithotype Succession. Sedimentary Geology, 329: 1-11.
2. K. Fellows, V. Rodriguez-Cruz, J. Covelli, A. Droopad, S. Alexander, and M. Ramanathan. (2015). Hybrid Markov Chain-von Mises Density Model for Drug Holiday Distributions. The AAPS Journal, 17(2): 427-437.
3. J. A. Jamal, H. Marco, K. Wolfgang, and D. B. Ali. (2013). Integration of Logistic Regression, Markov Chain and Cellular Automata Models to Simulate Urban Expansion. International Journal of Applied Earth Observation and Geoinformation, 21: 265–275.
4. R. Rahim, (2015). Analyzing Manpower Data of Higher Learning Institution: A Markov Chain Approach, International

Journal of Human Resource Studies, 5(3), 2162-3058.

5. R. Rahim, H. Ibrahim, S. A. Nadhar Khan, S. Saad, (2016), Evaluation of Faculty Employment Policies using Hybrid Markov Chain Model, The Social Sciences Journal., 11(10), 2590-2595.
6. R. Rahim, H. Ibrahim, S. A. Nadhar Khan, S. Saad,. (2015). A Hybrid Markov Chain Model of Manpower Data. Global Journal of Pure and Applied Mathematics, 11(6), 5075-5088.
7. R. Rahim, H. Ibrahim, M. Mat Kasim, F. A. Adnan, (2013). Projection Model of Postgraduate Students Flow. Journal of Applied Mathematics and Information Sciences, Natural Science Publishing7(2L), 383-387