International Journal of Engineering

Journal Homepage: www.ije.ir

Data Consumption Analysis by Two Ordinal Multivariate Control Charts

A. Hakimi[a], H. Farughi*[a], A. Amiri[b], J. Arkat[a]

[a] *Department of Industrial Engineering, Faculty of Engineering, University of Kurdistan, Sanandaj, Iran*
[b] *Department of Industrial Engineering, Faculty of Engineering, Shahed University, Tehran, Iran*

PAPER INFO

Paper history:
Received 01 Febraury 2022
Received in revised form 01 August 2022
Accepted 28 August 2022

Keywords:
Statistical Process Monitoring
Ordinal Multivariate Processes
Data Consumption
Customer Behavior

ABSTRACT

The process quality is described by one or more important factors called multivariate processes. Contingency tables used to demonstrate the relevance between these factors and modeled by log-linear model. There are also two types of statistical variables that are nominal and ordinal. In this paper, the variables are ordinal and two new control charts have been used to monitor the process of analyzing subscribers' consumption. These two multivariate ordinal chart are the MR chart and the multivariate ordinal categorical (MOC) used to monitor processes based on the ordinal log-linear model in Phase II. In addition, with a real numerical example, about analyzing the internet usage of mobile subscribers, two control charts are drawn and compared with each other in terms of average run length. In this case, we focus on customer behavior and in real action, by marketing department, changing in data consumption has been seen and analyzed. The study of the two proposed charts was performed using simulation based on real example in different situation, and the MOC performed relatively better.

doi: *10.5829/ije.2022.35.11b.14*

1. INTRODUCTION

Multivariate processes have many applications in today's world, both in manufacturing organizations and in services. By increasing such processes, the importance of monitoring them also increased and different methods were suggested to them [1]. There are generally two forms of log-linear models, which are the nominal log-linear model (NMLLM) and ordinal log-linear model (OLLM), associates the expected frequency of multivariate categorical characteristic with two or more variables. The multivariate ordinal processes include more than one ordered factor, for which multivariate categorical control charts based on the OLLM are used. Statistical Process Monitoring (SPM) uses contingency tables to monitor multivariate category processes simultaneously [2]. As well as, OLLM used to demonstrate the relevance between ordinal variables and correlative observations in an ordinal contingency table (OCT) cell. Thus, Yamamoto and Morakami [3] proposed the square OCT model, this study also considered the assumption of an unbalanced normal distribution applied to dental caries.

Many researches have also been done on multivariate monitoring processes, for example: Maleki et al. [4] proposed chart to monitor multivariate process mean and variation simultaneous in the existence of measurement error with linearly increasing variance under additive covariate model. Rasay et al. [5] explained and extended multivariate control charts applications for condition-based maintenance. Khedmati et al. [6] proposed auto-regressive integrated moving average and machine learning approaches for forecasting Bitcoin price. A generalized likelihood ratio test (GLRT) proposed by Li et al. [7] for monitoring multivariate categorical processes, in phase II. They proposed the EWMA-GLRT to improve the GLRT chart performance for detecting small shifts. In another related research, an integrated multivariate spatial-sign test with EWMA procedure proposed to monitor the model parameters of the multivariate nonparametric processes in phase II [8]. Li et al. [9] proposed a new multivariate categorical statistic based on binomial/multinomial by considering the

*Corresponding Author Institutional Email: h.farughi@uok.ac.ir (H. Farughi)

Please cite this article as: A. Hakimi, H. Farughi, A. Amiri, J. Arkat, Data Consumption Analysis by Two Ordinal Multivariate Control Charts, ***International Journal of Engineering, Transactions B: Applications***, Vol. 35, No. 11, (2022) 2196-2204

correlation between categorical variables. In phase II, they showed that the proposed control chart was robust to detect different shifts. Afterwards, the GLT statistic proposed to monitor multivariate processes in phase-II by Kamranrad et al. [2]. As well as, GLT is mixed with an EWMA statistic to enhance performance in small and medium shifts.

For monitoring categorical ordinal processes, Li et al. [10] proposed a new ordinal categorical control chart to determine changes in univariate ordinal processes in phase-II. Furthermore, they comparison that method with two charts and they showed this new method had good performances. Wang et al. [11] proposed two control chart for monitoring multivariate ordinal process (MOP) in phase II, which one of them is MOC chart, based OLLM. On that research, it was shown the MOC chart performed better than the LMBM chart. New multivariate ordinal based method known as multivariate ordinal-normal statistic (MONS) became brought by Hakimi et al. [12]. Additionally, the performance of the MONS approach became in comparison with the MG-p technique and outcomes showed the superiority of the MONS method under the small and slight shifts in OLLM parameters. Furthermore, Hakimi et al. [13] proposed MR statistic which is develop R statistic by Li et al, [10] and they showed have good performance to detect shifts. In addition, in phase I, one research has been done by Hakimi et al. [14], which used MR and LRT methods for estimating OLLM parameters. In this article, they monitor drug dissolution process as case study.

According to reviews by the authors, there are few studies in the field of monitoring the MOP. This paper aims to use two statistics for monitoring ordinal process. Ramakrishnan and Pecht [15] proposed a methodology life consumption monitoring for electronic systems. Our chosen process is about data consumption and subscriber consumption analysis is one of the most important indicators [16]. On the other hand, some research have been done on behavior's customer in mobile network field. Mehrotra et al. [17] analyze power consumption via mobile applications by using fuzzy clustering approach. Sarker [18] a machine learning based robust prediction model proposed for real-life mobile phone data. Jiang et al. [19] did the analysis by big data based network behavior insight of cellular networks applications of industry 4.0. Zhao et al. [20] proposed a method for monitoring data based by bottom-up modeling and its application for energy consumption prediction of campus building.

According to the studies summarized in Table 1, some of which are mentioned above, few articles have been done in the field of MOP. On the other hand, examining a real process in an up-to-date field such as internet use by subscriber can arouse the desire among researchers to conduct more research in this field. Moreover, the following bullet points can be identified

TABLE 1. Related works in multivariate process

Reference No.	Phase	Univariate/ Multivariate	Ordinal/ Nominal
[2]	II	Multivariate	Nominal
[7], [8], [9]	I , II	Multivariate	Nominal
[10]	II	Univariate	Ordinal
[11], [12], [13]	II	Multivariate	Ordinal
[14]	I	Multivariate	Ordinal

for the innovation of this paper:

- Comparing two charts for an ordinal multivariate process (It should be noted that these two charts were not compared with each other until now and both have recently been generalized)
- Using a real and up-to-date case study to compare charts and analyze them

The rest of the article is organized as follows. In next section the MOP is described, then in section 3, control charts for that process described one by one. In section 4, we describe the case study and analyze those charts for that. And at the end, we have mentioned conclusions and suggestions for future research.

2. MULTIVARIATE ORDINAL PROCESS

The MOP has at least two factors with more than one ordered level known as OCT. The OLLM is used to analyze the OCT. The OLLM demonstrates the main and interaction effects between ordinal factors; hence, it is applied to propose multivariate ordinal control charts. In fact, the OCT displaying a simultaneous relationship between two or more ordinal variables. p variables such as $y_1, y_2, ..., y_p$ consider each with $h_i, i = 1, 2, ..., p$ conceivable levels. Hence, the table cells demonstrate $h_1 \times h_2 \times ... \times h_p$ the value of possible frequencies (for more information refered to cited literature [2]). OLLM has been introduced in the literature to model the relationship between the level of an ordinal factor and its corresponding number in each cell. Assume the contingency table with two ordinal variables (y_1, y_2) with h_1 and h_2 levels. Thus, the OLLM is defined as the Equation (1):

$$\log \mu_{ij} = \mu + \alpha_i + \beta_j + \varphi(u_i - \bar{u})(v_j - \bar{v}), \quad (1)$$

where μ_{ij} is the mean value of observations for cell (i,j) and calculated by $\mu_{ij} = N\pi_{ij}$, which N is count cells in contingency table and $\boldsymbol{\pi}_{ij}$ is probably of occurs cells (for more information refered to cited literature [12]). Therefore, μ is the constant effect, α_i and β_j are the main effect of the ith row and jth column, respectively.

In addition, φ is defined as linear by linear interaction parameter in OLLM, which can be estimated by the Equation (2):

$$\log\left(\frac{\mu_{ij}\ \mu_{i+1.j+1}}{\mu_{i.j+1}\mu_{i+1.j}}\right)=\varphi(u_i-u_{i+1})(v_j-v_{j+1}), \quad (2)$$

where, $(u_i-u_{i+1})=1$ and $(v_j-v_{j+1})=1$. Notice that, parameters at OLLM in phase I can be estimated by iterative Newton's single-dimensional algorithm (for more information refered to cited literature [18, 19]).

For more than two variables, the OLLM with p factors is defined as Equation (3):

$$Log\ \boldsymbol{\mu}=\beta_0+\beta_1 y_1+\cdots+\beta_p y_p+\beta_{12}(y_1-\overline{y_1})(y_2-\overline{y_2})+\cdots+\beta_{1p}(y_1-\overline{y_1})(y_p-\overline{y_p})+\beta_{2p}(y_2-\overline{y_2})(y_p-\overline{y_p})+\cdots+\beta_{p-1,p}(y_{p-1}-\bar{y}_{p-1})(y_p-\overline{y_p})+\cdots+\varphi(y_1-\overline{y_1})\ldots(y_{p-1}-\bar{y}_{p-1})(y_p-\overline{y_p}) \quad (3)$$

where $\bar{y}_i$; $(i=1,2,..,p)$ is the average of the ith ordinal variable.

3. MULTIVARIATE ORDINAL CONTROL CHARTS

Wang et al. [11] proposed two control charts for monitoring multivariate ordinal processes, one of which was the chart presented by Li et al. [10]. This control chart has been used and compared under the Log-linear Multivariate Binomial/Multinomial (LMBM) in their paper. In addition, they presented another multivariate control chart called the Multivariate Ordinal Categorical (MOC), which showed that the MOC chart performed better in all shifts. Hakimi et al. [12, 13] also generalized three multivariate control chart statistics to monitor this type of process. These three statistics proposed by Li et al. [10] for univariate processes. One of them was the generalized multivariate generalized-p, which was integrated with EWMA statistics. Another statistic called multivariate ordinal-normal statistic (MONS) was presented by Hakimi et al. [12] showed that this statistic performs better than the generalized-p chart in small and medium shifts. Furthermore, Hakimi et al. [13] proposed a new control chart and they called MR statistic, which has better performance in these three charts. In this article, we intend to compare the best methods from these two articles that means bt Wang et al. [11] and Hakimi et al. [13], namely MOC and MR chart. Therefore, we should first explain how to calculate each statistic in below and the flowchart of monitoring multivariate control chart with categorical ordinal data is illustrated in Figure 1.

3. 1. MR Control Chart

Hakimi et al. [13] proposed this statistic and showed that it has better performance rather than others. p-way OCT consider (p-ordinal variables) each with $h_1, h_2, \ldots, h_p$ levels. Hence,

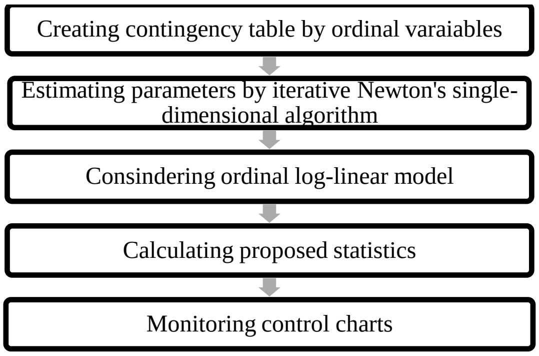

Figure 1. The flowchart of monitoring ordinal multivariate control chart

the known in-control (IC) probabilities for the ordinal cell is calculate by following equation:

$$\pi^{(0)}_{ijk\ldots p}=\frac{f(i,j,k,\ldots,p)}{\sum_{i=1}^{h_1}\sum_{j=1}^{h_2}\sum_{k=1}^{h_3}\ldots\sum_{p=1}^{h_p} f(i,j,k,\ldots,p)} \quad (4)$$

where, $f(i,j,k,\ldots,p)$ is the abundance in the cell $(i,j,k,\ldots,p)$. Thus, the MR statistic for each time is defined by Equation (5):

$$MR_t=\left|\sum_{i=1}^{h_1}\sum_{j=1}^{h_2}\sum_{k=1}^{h_3}\ldots\sum_{p=1}^{h_p}(F^{(0)}_{ijk\ldots p-1}+F^{(0)}_{ijk\ldots p}-1)z_{ijk\ldots p}\right|, \quad (5)$$

where, $F^{(0)}_{ijk\ldots p}=\sum_{i=1}^{h_1}\sum_{j=1}^{h_2}\sum_{k=1}^{h_3}\ldots\sum_{p=1}^{h_p}\pi^{(0)}_{ijk\ldots p}$. In addition, $\mathbf{z}_t$ is:

$$\mathbf{z}_t=a^{-1}_{0,t,\lambda}\sum_{s=1}^{t}(1-\lambda)^{t-s}\mathbf{n}_t, \quad (6)$$

where, $a^{-1}_{t_1,t_2,\lambda}=\sum_{t=t_1+1}^{t_2}(1-\lambda)^{t_2-t_1}$ and

$\boldsymbol{n}_t=[n_{111\ldots1t}, n_{112\ldots1t}, \ldots, n_{11h_3\ldots h_pt}, n_{121\ldots1t}, \ldots, n_{12h_3\ldots h_pt}, \ldots, n_{h_1h_21\ldots1t}, \ldots, n_{h_1h_2h_3\ldots h_pt}]$

Pseudo-code for MR_t calculated:

Step1: Consider a p way matrix which each cell has a number called f(i,j,..,p).
Step2: Calculate $\pi^{(0)}$ for each cell by Equation (4), then calculate $F^{(0)}$.
Step3: Consider $\mathbf{n_t}$ and calculate a^{-1} by specified λ, then calculate $\mathbf{z_t}$ by Equation (6).
Step4: Finally, compute MR_t by Equation (5).

Also, if UCL_1 indicates the upper limit of MR chart, we set $MR_t > UCL_1$ and this is done by simulation.

3. 2. MOC Control Chart

As mentioned, Wang et al. [11] presented this control chart for monitoring multivariate ordinal processes, which we want to describe here. Assuming a ordinal contingency table that has p ordinal variables and each levels $h_1, h_2, \ldots, h_p$, the statistics they use are calculated according to the following equation:

$$R_t = \frac{1}{N}\left(z_t - Np^{(0)}\right)^T Y\left(Y^T \Lambda^{(0)} Y\right)^{-1} Y^T\left(z_t - Np^{(0)}\right) \quad (7)$$

In equation (7), N represents the total number of samples and z_t is actually a vector based on EWMA, which is calculated as follows:

$$\boldsymbol{z}_t = (1-\lambda)z_{t-1} + \lambda n_t \quad (8)$$

We also know that $0 < \lambda \leq 1$. In Equation (8) the vector n_t represents the sample vector, same as MR statistic.

In Equation (7), $p^{(0)}$ the probability vector is in the controlled state which has dimensions of $1 \times h$. Y is also a vector for contingency table elements (refer to literature[11] for more information). The variance matrix $\Lambda^{(0)}$ is also calculated using the reference [21] IC state in Equation (9) as follows:

$$\Lambda^{(0)} = diag\left(p^{(0)}\right) - p^{(0)}p^{T(0)} \quad (9)$$

In the above equation, daig(x) means a square matrix with x values on its original diameter, and finally it is necessary to calculate the value of the upper limit of control of this chart.

Pseudo-code for R_t calculated:

Step1: Consider a *p* way matrix and N is the total number of samples.
Step2: Consider $\mathbf{n_t}$ and calculate $\mathbf{z_t}$ by specified λ in equation (8), as well as, z_0 is a vector with values of one.
Step3: Consider $p^{(0)}$ the probability vector and calculate Y and $\Lambda^{(0)}$ by equation (9).
Step4: At the end, calculate R_t by equation (7).

If UCL_2 indicates the upper limit of this chart, we set $R_t > UCL_2$ and this is done by simulation.

4. DATA CONSUMPTION ANALYSIS

In this section, we have presented a real example which is based on given examples by Agresti [21]. In companies offering data SIM cards, an important characteristic is the amount of data consumption (that mean the volume of Internet consumption) because the amount of repurchase of data packages is directly related to it. Some research focus on usage in mobile networks and analyze that by new methods like machine learning and big data (For additionl information see literature [17, 18]). On the other hand, marketing department in that companies try to increase their customers and offer plans to retain subscribers. As a case study, we set up a mobile virtual network operator (MVNO) company to monitor data usage and we will analyze the effects of marketing department offers on customer behavior. The company, which has more than half a million subscribers, wants to analyze the behavior of its subscribers. To do this, in the next section, a sample of subscribers is taken and analyzed. In this analysis, we want to discover the changes made and measure the changes using the control charts proposed in the previous section.

4. 1. Select and Categorize Data

Here we are dealing with a large number of subscribers and we have to choose a good example of them. The number of samples with stratified sampling method and calculated by Equation (10) (based on Zhao et al. [23]):

$$n = \frac{\left(\sum_{l=1}^{L} N_l.\sigma_l\right)^2}{\sigma^2 + \sum_{l=1}^{L} N_l.\sigma_l{}^2} \quad (10)$$

where N_l and σ_l are the number and standard deviation of each stratum, respectively and σ^2 is the variance of the total data set. After that calculation and with roundup numbers received to 2,000 subscribers to monitor the data consumption behavior who have entered the network within a period of one month. Here are three variables: Ages, education level, and data usage over a period for all of these subscribers. For the subscribers' age, three levels are considered young (less than 24 years), middle-aged (between 24 and 35 years), and old (more than 35 years). Also, for the education level, two levels of diploma (which includes diploma and lower) and higher education (which includes students and university graduates) and for data consumption in a period of time, three levels of low consumption (less than 1 GB), Medium consumption (between 1 to 3 GB) and high consumption (more than 3 GB) are considered. The following table shows the number of subscribers for each level:

TABLE 2. OCT for data consumption behavior of 2,000 subscribers

	Ages							
	Young		Middle-Aged		Older			
	Education Level							
Data Usage (GB)	Diploma	Higher education	Diploma	Higher education	Diploma	Higher education	Total	Percentage
Low consumption	106	127	101	99	90	102	625	31.3%
Medium consumption	121	133	131	125	66	121	697	34.9%
High consumption	139	155	125	112	57	90	678	33.9%
Total	366	415	357	336	213	313	2000	100%

Also, the covariance matrix used for the ordinal log-linear model estimates is as follows:

$$\text{cov}(\boldsymbol{\beta}) = \{\mathbf{X}'[\mathbf{diag}(\boldsymbol{\mu}) - \boldsymbol{\mu}\boldsymbol{\mu}'/N]\mathbf{X}\}^{-1}, \tag{11}$$

where $\boldsymbol{\mu}$ and $\mathbf{X}$ are the mean counts' vector and matrix values of the contingency table cells, respectively. As well as, $\mathbf{diag}(\boldsymbol{\mu})$ is a matrix that diagonal and show mean of the contingency table cells [21]. Moreover, the IC parameter vector of the OLLM is assumed (based on Zafar [22]):

$$\boldsymbol{\beta} = [1, -0.5, -0.5, -0.5, 0.15, 0.15, 0.15, -0.025]$$

Besides, the IC standard deviations for this example are as follows:

$$\hat{\boldsymbol{\sigma}}_{\mathbf{0}\boldsymbol{\beta}} = [2.09, 1.34, 1.31, 1.33, 0.79, 0.75, 0.77, 0.28]$$

The OLLM for OCT with three ordinal factors, including ages (A), education level (E), and data usage (D) under the IC condition is defined as Equation (12):

$$\text{Log}\,\boldsymbol{\mu} = 1 - 0.5\,A - 0.5E - 0.5D + 0.15(A - \bar{A})(E - \bar{E}) + 0.15(A - \bar{A})(D - \bar{D}) + 0.15(D - \bar{D})(E - \bar{E}) - 0.025(A - \bar{A})(E - \bar{E})(D - \bar{D});\ A = 1,2,3\,, E = 1,2\,, D = 1,2,3. \tag{12}$$

The UCLs of the MR and MOC charts in this contingency table are calculated equal to 37.2095 and 36.7201, respectively, that set through 10,000 simulation when ARL_0=370. All the simulations in this article and figures have been done and drawn by MATLAB software. Next, we want to be able to examine these control charts for this issue. This review is displayed in two modes: by changing the parameters and based on reality.

4. 2. The Charts Performance When Parameters Change

In this subsection, we want to compare the two control charts presented in the previous section under the same conditions by simulating them. This comparison is made in phase II, so the measure of the average run length (ARL_1) in OC conditions.

Some comparisons, by change in some parameters, of these two charts is shown in the following tables:

In Tables 3 to 6 you can see, the ARL_1 values of the MOC chart are slightly smaller than MR's chart under small shifts. However, in medium and large shifts the performance is similar and sometimes the MR chart is better. Hence, the MOC chart works a little better compared to the MR in these mentioned shifts. Also, the value of λ has a relatively better performance in finding larger changes with a value of 0.2 compared to 0.1, but for small shifts, the average value of 0.1 has a shorter run length and can be detected sooner. Also, the parameter β_1 is more sensitive to shifts than other parameters and the width parameter from the origin is relatively less sensitive. Note that, the $\boldsymbol{\delta}$ value has been changed by parameter in the tables to better identify the evaluation rate in each one. Accordingly, the rest of the simulations have been by various simultaneous shifts of OLLM parameters based on λ = 0.15, some of which are shown in Figures 2 to 5.

TABLE 3. ARL_1 values and standard deviations in parentheses under the different shifts in the intercept $(\beta_0 + \delta.\sigma_{\beta_0})$

λ	δ	-1	-0.75	-0.5	-0.25	-0.1	+0.1	+0.25	+0.5	+0.75	+1
0.1	MR	1.08 (0.04)	24.03 (1.22)	75.90 (1.61)	188.72 (2.78)	242.81 (3.01)	240.04 (3.02)	189.23 (2.71)	79.24 (1.61)	22.67 (1.21)	1.05 (0.04)
	MOC	1.03 (0.02)	26.11 (1.18)	76.82 (1.56)	187.79 (2.68)	224.06 (2.99)	223.95 (2.99)	185.56 (2.53)	77.31 (1.62)	23.10 (1.24)	1.00 (0.02)
0.2	MR	1.00 (0.01)	25.39 (1.19)	76.65 (1.63)	193.11 (2.57)	249.75 (3.01)	249.01 (3.03)	192.09 (2.82)	78.33 (1.59)	20.22 (1.21)	1.00 (0.01)
	MOC	1.00 (0.01)	23.27 (1.09)	75.44 (1.53)	191.65 (2.48)	232.79 (3.01)	231.22 (3.00)	190.17 (2.69)	77.02 (1.57)	20.53 (1.32)	1.00 (0.00)

TABLE 4. ARL_1 values and standard deviations in parentheses under the different shifts in A coefficient $(\beta_1 + \delta.\sigma_{\beta_1})$

λ	δ	-0.75	-0.5	-0.25	-0.1	-0.05	+0.05	+0.1	+0.25	+0.5	+0.75
0.1	MR	1.00 (0.00)	9.72 (1.00)	37.12 (1.32)	125.07 (2.21)	202.61 (3.05)	201.47 (3.00)	122.45 (2.33)	35.32 (1.29)	7.09 (0.97)	1.00 (0.00)
	MOC	1.00 (0.00)	9.91 (0.89)	36.27 (1.27)	121.45 (2.35)	205.05 (3.03)	203.23 (3.00)	118.29 (2.15)	33.92 (1.29)	7.02 (0.89)	1.00 (0.00)
0.2	MR	1.00 (0.00)	8.04 (0.68)	35.81 (1.19)	120.77 (2.26)	204.79 (3.06)	203.29 (3.04)	122.99 (2.61)	37.82 (1.26)	8.00 (0.77)	1.00 (0.00)
	MOC	1.00 (0.00)	8.05 (0.66)	35.92 (1.25)	119.93 (2.22)	206.09 (3.03)	204.97 (3.02)	119.87 (2.56)	35.73 (1.31)	8.03 (0.69)	1.00 (0.00)

TABLE 5. ARL_1 values and standard deviations in parentheses under the different shifts in E,D coefficient $(\beta_{23} + \delta.\sigma_{\beta_{23}})$

λ	δ	-0.75	-0.5	-0.25	-0.1	-0.05	+0.05	+0.1	+0.25	+0.5	+0.75
0.1	MR	1.00 (0.00)	9.04 (0.99)	41.02 (1.40)	129.34 (2.31)	211.79 (3.03)	209.54 (3.01)	128.09 (2.24)	40.35 (1.31)	8.81 (0.83)	1.00 (0.00)
	MOC	1.00 (0.01)	9.31 (0.87)	39.95 (1.36)	129.56 (2.36)	208.00 (3.01)	207.89 (3.01)	124.99 (2.27)	39.91 (1.30)	8.37 (0.81)	1.00 (0.00)
0.2	MR	1.00 (0.00)	8.95 (0.99)	39.25 (1.39)	130.27 (2.41)	209.23 (3.03)	210.08 (3.02)	129.37 (2.31)	38.39 (1.38)	8.91 (0.79)	1.00 (0.00)
	MOC	1.00 (0.00)	8.91 (0.93)	39.07 (1.36)	131.51 (2.38)	211.79 (3.00)	210.46 (3.01)	128.88 (2.30)	37.06 (1.32)	8.73 (0.68)	1.00 (0.00)

TABLE 6. ARL_1 values and standard deviations in parentheses under the different shifts in A,E,D coefficient $(\varphi + \delta.\sigma_{\varphi})$

λ	δ	-0.5	-0.25	-0.1	-0.05	-0.02	+0.02	+0.05	+0.1	+0.25	+0.5
0.1	MR	1.00 (0.02)	11.00 (1.01)	42.26 (1.41)	131.15 (2.39)	211.06 (3.04)	212.92 (3.03)	130.67 (2.32)	41.90 (1.32)	10.03 (0.88)	1.00 (0.00)
	MOC	1.00 (0.02)	11.02 (0.99)	40.93 (1.35)	130.87 (2.33)	211.77 (3.03)	209.80 (3.01)	129.92 (2.25)	40.07 (1.31)	10.11 (0.89)	1.00 (0.00)
0.2	MR	1.00 (0.00)	10.13 (0.98)	38.21 (1.28)	129.03 (2.32)	212.65 (3.05)	215.71 (3.03)	128.64 (2.20)	39.79 (1.30)	9.12 (0.71)	1.00 (0.00)
	MOC	1.00 (0.00)	10.21 (0.91)	39.24 (1.31)	129.05 (2.30)	212.09 (3.04)	211.68 (3.03)	128.05 (2.24)	39.44 (1.30)	9.09 (0.70)	1.00 (0.00)

Figures 2 to 5 show the MR and MOC control charts efficiency under shifts in β_2 and φ. The OC alarms by the MR and MOC control charts happen at the 36^{th} and 34^{th} observation under $-0.25\sigma_{\beta_2}$ shift in β_2 and 51^{st} and 47^{th} observation under $0.1\sigma_\varphi$ shift in φ, respectively. Note that we chose different values for two separate parameters to show a fair comparison for these two charts. Nevertheless, the results demonstrate that the MOC control chart find the OC situation sooner.

4. 3. The Charts Performance Based on Actual Change Following this review, the company decided to design and send promotions to subscribers who belong to the low-consumption category. After two

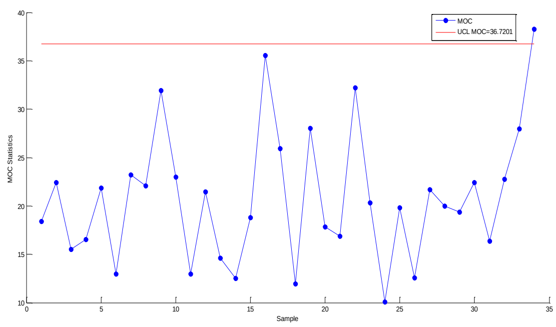

Figure 3. Performance of MOC control chart under $-0.1\sigma_{\beta_2}$ shift in β_2

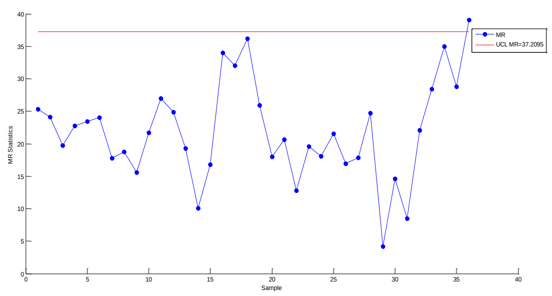

Figure 2. Performance of MR control chart under $-0.1\sigma_{\beta_2}$ shift in β_2

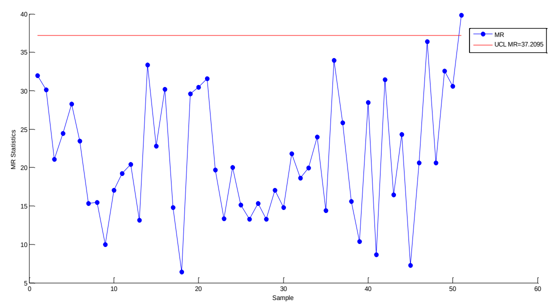

Figure 4. Performance of MR control chart under $0.1\sigma_\varphi$ shift in φ

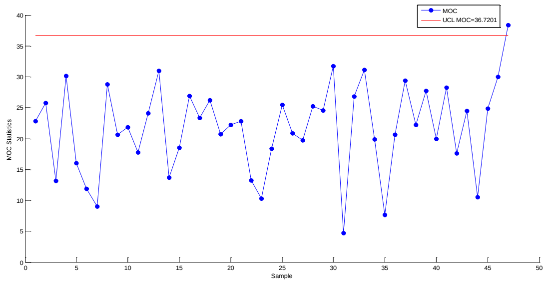

Figure 5. Performance of MOC control chart under $0.1\sigma_{\varphi}$ shift in φ

months, the above survey was conducted again for those 2000 subscribers, during which 1973 subscribers remained in the network, and the number and classification of these subscribers were as Table 7:

The above table shows that the company's promotions have responded, and the percentage of low-consumption subscribers has decreased compared to two months ago. However, the important thing for us is how the proposed control charts show the state of being out of control in a few samples, because in the second table, the average of the cells in the contingency table has changed and process is OC. Therefore, we want to examine the ARL_1 between the two proposed charts. The ARL_1 of these two charts shows the OC signals happen at the 7th and 6th sample in the MR and MOC control charts. Since the information in Table 1 is updated every week, a point in the chart is drawn for each week, which can be seen in Figures 6 and 7. The results again demonstrate the MOC chart finds the OC situation faster than the MR chart. It is noteworthy that both charts reveal the changes before the 8th week (two month), which is related to Table 7. In other words, the marketing department can detect changes in customer usage sooner. Also, as mentioned, these changes (campaign) are real and the results clearly show that this campaign has had a relatively good effect on subscriber consumption.

TABLE 7. OCT for data consumption behavior after promotions

	Ages							
	Young		Middle-Aged		Older			
	Education Level							
Data Usage (GB)	Diploma	Higher education	Diploma	Higher education	Diploma	Higher education	Total	Percentage
Low consumption	93	94	89	80	84	89	529	26.8%
Medium consumption	124	151	137	140	70	122	744	37.7%
High consumption	141	164	129	113	56	97	700	35.5%
Total	358	409	355	333	210	308	1973	100%

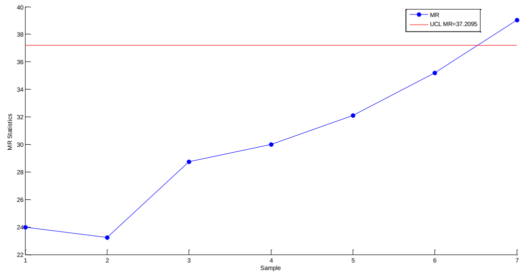

Figure 6. MR control chart performance in actual change

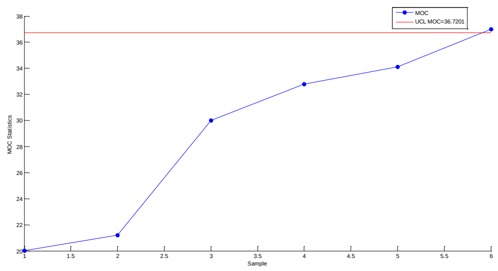

Figure 7. MOC control chart performance in actual change

5. CONCLUSION REMARKS

In this paper, the intention was to introduce statistics and charts for monitoring multivariate ordinal processes. This was done by using two charts from literature [11, 13], each of which was mentioned as the best control chart in their articles. For a better understanding, brief explanations of how the two statistics were calculated and explained; then, by simulation operations, both of them were set to 370 points with the ARL in IC state. In addition, a real example of customer behavior analysis was provided and with shifts in some parameters, these two control charts were compared. That case study, analyze the effects of marketing department offers on customer’s usage data internet behavior. After that analysis and shifts in the values and coefficients of the log-linear relationship, the ARL_1 was compared for both charts. The results showed that the MOC chart performs better in finding changes, especially in smaller shifts. Moreover, using the actual changes that occurred within 8 weeks on the selected subscribers, we performed an analysis on these two charts. As future research, other ordinal multivariate statistics can be used as further research, and new statistics can also be presented. Another research that could be interesting is to use charts to examine nominal and ordinal variables simultaneously.

6. ACKNOWLEDGMENT

Special thanks to the staff of Shatel Mobile for their cooperation with the authors. We are also grateful for Ms. Vafaie's efforts in collecting data.

7. REFERENCES

1. MacGregor, J. F. and Kourti, T. "Statistical process control of multivariate processes" ***Control Engineering Practice***, Vol. 3, No. 3, (1995), 403-414. doi: 10.1016/0967-0661(95)00014-L
2. Kamranrad, R., Amiri, A. and Niaki, S.T.A. "New Approaches in Monitoring Multivariate Categorical Processes based on Contingency Tables in Phase II" ***Quality and Reliability Engineering International***, Vol. 33, No. 5, (2017), 1105-1129. doi: 10.1002/qre.2103
3. Yamamoto, K., Murakami, H. "Model based on skew normal distribution for square contingency tables with ordinal categories" ***Computational Statistics & Data Analysis***, Vol. 78 (2014), 135-140. doi: 10.1016/j.csda.2014.04.007
4. Maleki, M. R., Ghashghaei, R. and Amiri, A. "Simultaneous monitoring of multivariate process mean and variability in the presence of measurement error with linearly increasing variance under additive covariate model" ***International Journal of Engineering, Transactions A: Basics***, Vol. 29, No. 4, (2016), 514-523. doi: 10.5829/idosi.ije.2016.29.04a.10
5. Rasay, H., Fallahnezhad, M. S. and ZareMehrjardi, Y. "Application of multivariate control charts for condition-based maintenance" ***International Journal of Engineering, Transactions A: Basics***, Vol. 31, No. 4, (2018), 597-604. doi: 10.5829/ije.2018.31.04a.11
6. Khedmati, M., Seifi, F. and Azizi, M. "Time Series Forecasting of Bitcoin Price Based on Autoregressive Integrated Moving Average and Machine Learning Approaches." ***International Journal of Engineering, Transactions A: Basics***, Vol. 33, No. 7, (2020), 1293-1303. doi: 10.5829/ije.2020.33.07a.16
7. Li, J., Tsung, F. and Zou, C. "Directional control schemes for multivariate categorical processes" ***Journal of Quality Technology***, Vol. 44, No. 2, (2012), 136-154, doi: 10.1080/00224065.2012.11917889
8. Li, Z., Zou, C., Wang, Z. and Huwang, L. "A multivariate sign chart for monitoring process shape parameters" ***Journal of Quality Technology***, Vol. 45, No. 2, (2013), 149-165, doi: 10.1080/00224065.2013.11917923
9. Li, J., Tsung, F. and Zou, C. "Multivariate binomial/multinomial control char" ***IIE Transactions***, Vol. 46, No. 5, (2014), 526-542, doi: 10.1080/0740817X.2013.849830
10. Li, J., Tsung, F. and Zou, C. "A simple categorical chart for detecting location shifts with ordinal information" ***International Journal of Production Research***, Vol. 52, No. 2, (2014), 550-562. doi: 10.1080/00207543.2013.838329
11. Wang, J., Li, J. and Su, Q. "Multivariate ordinal categorical process control based on log-linear modeling" ***Journal of Quality Technology***, Vol. 49, No. 2, (2017), 108-122. DOI: 10.1080/00224065.2017.11917983
12. Hakimi, A., Farughi, H., Amiri, A. and Arkat, J. "New phase II control chart for monitoring ordinal contingency table-based processes" ***Journal of Industrial and Systems Engineering***, Vol. 12, (2019), 15-34.
13. Hakimi, A., Farughi, H., Amiri, A. and Arkat, J. "Phase II Monitoring of the Ordinal Multivariate Categorical Processes" ***Advances in Industrial Engineering***, Vol. 55, No. 3, (2021), 249-267. doi: 10.22059/aie.2021.329719.1802
14. Hakimi, A., Farughi, H., Amiri, A. and Arkat, J. "Phase I Monitoring of Multivariate Ordinal based Processes: The MR and LRT Approaches (A real case study in drug dissolution process)" ***International Journal of Industrial Engineering & Production Research***, Vol. 33, No. 1, (2022), 1-18. doi: 10.22068/ijiepr.33.1.7
15. Ramakrishnan, A. and Pecht, M. G. "A life consumption monitoring methodology for electronic systems". ***IEEE Transactions on Components and Packaging Technologies***, Vol. 26, No. 3, (2003), 625-634. doi: 10.1109/TCAPT.2003.817654
16. Cooper, J. O., Heron, T. E. and Heward, W. L. "Applied behavior analysis" (2020). Pearson UK.
17. Mehrotra, D., Nagpal, D., Srivastava, R. and Nagpal, R. "Analyse Power Consumption by Mobile Applications Using Fuzzy Clustering Approach" ***International Journal of Engineering, Transactions C: Aspects***, Vol. 31, No. 12, (2018), 2037-2043. doi: 10.5829/ije.2018.31.12c.07
18. Sarker, I. H. "A machine learning based robust prediction model for real-life mobile phone data." ***Internet of Things***, Vol. 5, (2019), 180-193. doi: 10.1016/j.iot.2019.01.007
19. Jiang, D., Wang, Y., Lv, Z., Qi, S. and Singh, S. "Big data analysis-based network behavior insight of cellular networks for industry 4.0 applications" ***IEEE Transactions on Industrial Informatics***, Vol. 16, No. 2, (2019). 1310-1320. doi: 10.1109/TII.2019.2930226
20. Zhao, T., Xu, J., Zhang, C. and Wang, P. "A monitoring data-based bottom-up modeling method and its application for energy consumption prediction of campus building" ***Journal of Building Engineering***, Vol. 35, (2021), 101962. https://doi.org/10.1016/j.jobe.2020.101962
21. Agresti, A. "Analysis of ordinal categorical data" Department of Statistics University of Florida Gainesville, Florida: John Wiley & Sons, Inc., Hoboken, New Jersey. (2010).
22. Zafar, S. "Non-iterative Estimation Methods for Ordinal Log-linear Models" (2017), Doctoral dissertation, The University of Newcastle.
23. Zhao, X., Liang, J. and Dang, C. "A stratified sampling based clustering algorithm for large-scale data." ***Knowledge-Based Systems***, Vol. 163, (2019), 416-428. doi: 10.1016/j.knosys.2018.09.007

Persian Abstract

چکیده

در برخی از کاربردهای پایش فرآیند آماری، کیفیت یک فرآیند یا محصول توسط یک یا بیش از یک عامل مهم به نام فرآیند چند متغیره توصیف می‌شود. برای نشان دادن رابطه‌ی بین این عوامل، از جداول توافقی استفاده می‌شود و با مدل لُگ‌خطی مدل‌سازی می‌شوند. همچنین دو نوع متغیر آماری وجود دارد که اسمی و ترتیبی می باشند. در این مقاله متغیرها از جنس ترتیبی بوده و از دو نمودار کنترل جدید برای پایش فرآیند بررسی مصرف مشترکین استفاده شده است. این دو نمودار ترتیبی چندمتغیره عبارتند از: نمودار MR و نمودار طبقه‌بندی شده چند متغیره (MOC) که برای نظارت بر فرآیندهای مبتنی بر مدل لُگ‌خطی ترتیبی در فاز ۲ استفاده شده است. همچنین با یک مثال عددی واقعی این دو نمودار کنترل رسم شده و با یکدیگر از لحاظ متوسط طول دنباله مقایسه می‌شوند. این مثال واقعی در مورد تحلیل مصرف اینترنت مشترکین در شبکه همراه می باشد. بررسی دو نمودار پیشنهادی با استفاده از شبیه سازی و مثال واقعی انجام شده و نمودار MOC عملکرد نسبتاً بهتری داشته است.