



A comparative study for estimation of the parameters of the folded exponential power distribution

Şükrü Acıtaş


Eskişehir Technical University

Faculty of Science

Department of Statistics

26470 Eskişehir, TURKEY

sacitas@eskisehir.edu.tr

 0000-0002-4131-0086

Abstract

Folded distributions are commonly used for the data set which is obtained without regarding the algebraic signs of the measurements. Therefore, they have extensive applications in different fields such as engineering, finance, insurance and so on. Folded exponential power (FEP) distribution is a newly proposed distribution which has modeling flexibility and easy usage [1]. In this study, we therefore consider different parametric methods for estimating the unknown parameters of FEP distribution. Maximum likelihood (ML), ordinary and weighted least squares (LS and WLS), Cramer von Mises (CVM) and maximum product of spacings (MPS) methods are used during the estimation process. The performances of the considered estimators are compared in a Monte-Carlo simulation study via bias and mean squared error (MSE) criteria. Results show that MPS method outperforms its rivals. Two real life applications taken from the literature are also considered.

Keywords: Bias, Efficiency, Estimation, Folded exponential power distribution.

Öz

Katlanmış üstel güç dağılımının parametrelerinin tahmini için karşılaştırmalı bir çalışma

Katlanmış dağılımlar, ölçümlerin cebirsel işaretlerinin önemli olmadığı veri setleri için yaygın olarak kullanılırlar. Bu nedenle, mühendislik, finans, ekonomi v.b. bir çok alanda kapsamlı uygulamaları vardır. Modelleme esnekliği ve kolay kullanımı olan katlanmış üstel güç (FEP) dağılımı yeni önerilmiştir [1]. Bu çalışmada bu nedenle, FEP dağılımının bilinmeyen parametrelerinin farklı parametrik yöntemlerle tahmin edilmesi ele alınmıştır. En çok olasılık (ML), sıradan ve ağırlıklandırılmış en küçük kareler (LS ve WLS), Cramer von Mises (CVM) ve aralıkların çarpımının maksimumu (MPS) metotları tahmin sürecinde kullanılmıştır. Ele alınan tahmin edicilerin performansı, Monte-Carlo simülasyon çalışmasında yan ve hata kareler ortalaması (MSE) kriterleri kullanılarak karşılaştırılmıştır. Sonuçlar, MPS yönteminin rakiplerinden daha iyi bir performansa sahip olduğunu göstermiştir. Literatürden alınan iki gerçek hayat uygulaması ele alınmıştır.

Anahtar sözcükler: Yan, Etkinlik, Tahmin, Katlanmış üstel güç dağılımı.

1. Introduction

Folded distributions have many applications in different fields of science such as engineering, economics, medicine and so on [2 - 6]. Folded distributions are useful when the algebraic signs of the measurements are ignored. Some examples for these kind of measurements can be given as differences, deviations, lengths and angels. Therefore, the distribution of these measurements are mostly described by folded distributions which are defined as the distribution of absolute measurements.

Suppose that random variable X has probability density function (pdf) and cumulative distribution function (cdf) $g(\cdot)$ and $G(\cdot)$, respectively and define $Y = |X|$. The distribution of Y is a folded distribution and its pdf and cdf are then given by

$$f(y) = g(y) + g(-y) \quad \text{and} \quad F(y) = G(y) - G(-y). \quad (1)$$

respectively. If one take $g(\cdot)$ as the pdf of $N(\mu, \sigma^2)$, Y has folded normal (FN) distribution which is proposed by Leone et al. [5]. In the same manner, folded t distribution is suggested by Psarakis and Panareteos [7]. Cooray et al. [8] introduce folded logistic (FL) distribution. Nadarajah and Bakar [1] also consider some folded distributions to model the log-transformed Norwegian data. Among these, they originate three new distributions namely folded generalized t, folded Gumbel and folded exponential power (FEP) distributions and study some statistical properties and maximum likelihood (ML) estimation of the parameters.

In this study, we consider FEP distribution because of its simplicity and modeling flexibility. The pdf and cdf of FEP distribution are given by

$$f(x; \mu, \sigma, p) = \frac{1}{2p^{1/p}\Gamma(1 + 1/p)\sigma} \left[\exp\left\{-\frac{|x - \mu|^p}{p\sigma^p}\right\} + \exp\left\{-\frac{|x + \mu|^p}{p\sigma^p}\right\} \right] \quad (2)$$

and

$$F(x; \mu, \sigma, p) = \frac{1}{2\Gamma(1/p)\sigma} \left\{ \text{sign}(x - \mu)\gamma\left(\frac{1}{p}, \frac{|x - \mu|^p}{p\sigma^p}\right) - \text{sign}(-x - \mu)\gamma\left(\frac{1}{p}, \frac{|x + \mu|^p}{p\sigma^p}\right) \right\}. \quad (3)$$

respectively. Here, $x > 0$, $\mu > 0$, $\sigma > 0$ and $p > 0$. It should be mentioned that p stands for the shape parameter and μ , σ have their standard meanings, i.e the former and the latter are location and scale parameters, respectively. $\Gamma(\cdot)$ is the well-known gamma function and $\gamma(\cdot)$ denotes the incomplete gamma function which is defined as $\gamma(a, x) = \int_0^x t^{a-1}\exp(-t)dt$, $a > 0$, $x > 0$.

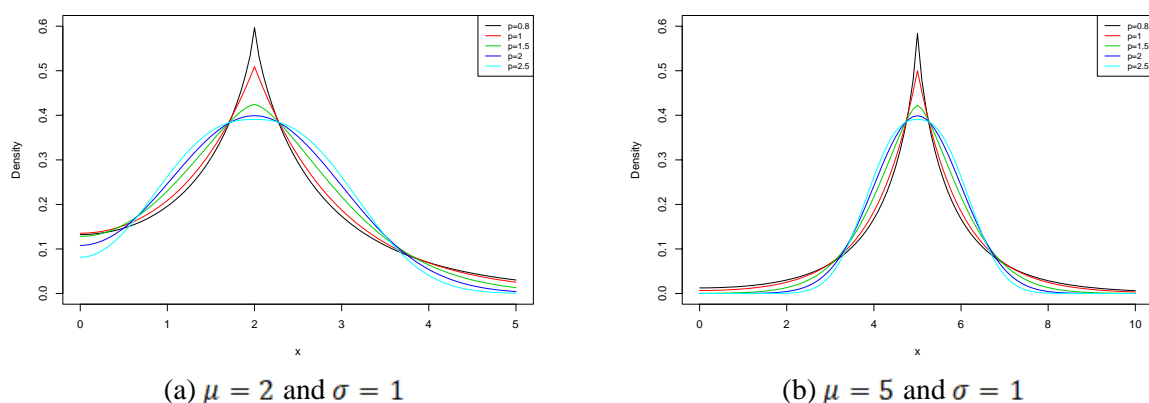


Figure 1. The pdf plots of FEP distribution for some certain values of the parameters.

FEP distribution is obtained using exponential power (EP), which is also known as generalized normal distribution [9], therefore it inherits properties of EP. For example, FEP distribution is platykurtic when $p > 2$ otherwise it is leptokurtic, see Figure 1 in which density plots of FEP distribution are provided for some representative values of the parameters. FEP reduces to folded Laplace [10] and folded normal distributions for $p = 1$ and $p = 2$, respectively. Furthermore, half exponential power distribution [11] is obtained when $\mu = 0$ and $\sigma = 1$. It is clear that half normal distribution is a special case of FEP distribution

if $\mu = 0$, $\sigma = 1$ and $p = 2$. We refer to Nadarajah and Bakar [1] for further information about FEP distribution.

As mentioned above, FEP distribution has attractive properties and modelling flexibility. However, there are no previous studies concerning different parametric estimation methods for the parameters of FEP distribution to the best of our knowledge. We therefore consider least squares (LS), weighted LS, Cramer von Mises (CVM) and maximum product of spacings (MPS) methods for estimating the parameters of FEP distribution. The performances of these estimation methods are compared using different settings of parameter values and sample sizes according to the bias, mean squared error (MSE) and deficiency (Def) criteria. We aim to evaluate the best method for estimating the parameters of FEP distribution. There are plenty of studies in which performances of the different estimation methods are compared in the context of different distributions, see for example [12 - 14].

The rest of the paper is organized as follows. Brief descriptions of the estimation methods are given in Section 2. Section 3 consists of a Monte-Carlo simulation study and its results. Real life data applications are considered in Section 4. The paper is finalized with some concluding remarks.

2. Methods of estimation

In this section, brief descriptions of the ML, LS, WLS, CVM and MPS methods are provided. We first provide a common notation which will be used in descriptions of the methods. Let x_1, x_2, \dots, x_n be a random sample drawn from FEP distribution. The order statistics of this sample are then denoted by $x_{(1)}, x_{(2)}, \dots, x_{(n)}$.

2.1. ML method

The ML estimators of the parameters μ , σ and p are obtained by maximizing the following loglikelihood function:

$$\log L = -n \log 2 - \frac{n}{p} \log p - n \log \sigma - n \log \Gamma \left(1 + \frac{1}{p} \right) + \sum_{i=1}^n \log \left(\exp \left\{ -\frac{|x_i - \mu|^p}{p\sigma^p} \right\} + \exp \left\{ -\frac{|x_i + \mu|^p}{p\sigma^p} \right\} \right). \tag{4}$$

After taking partial derivatives of the $\log L$ function with respect to the parameters of interest and setting them equal to zero, we obtain the following likelihood equations [1]:

$$\frac{\partial \log L}{\partial \mu} = \sum_{i=1}^n \frac{|x_i - \mu|^{p-1} \text{sign}(x_i - \mu) \exp \left\{ -\frac{|x_i - \mu|^p}{p\sigma^p} \right\}}{\exp \left\{ -\frac{|x_i - \mu|^p}{p\sigma^p} \right\} + \exp \left\{ -\frac{|x_i + \mu|^p}{p\sigma^p} \right\}} - \sum_{i=1}^n \frac{|x_i + \mu|^{p-1} \text{sign}(x_i + \mu) \exp \left\{ -\frac{|x_i + \mu|^p}{p\sigma^p} \right\}}{\exp \left\{ -\frac{|x_i - \mu|^p}{p\sigma^p} \right\} + \exp \left\{ -\frac{|x_i + \mu|^p}{p\sigma^p} \right\}} = 0, \tag{5}$$

$$\frac{\partial \log L}{\partial \sigma} = -n\sigma^p + \sum_{i=1}^n \frac{|x_i - \mu|^p \exp \left\{ -\frac{|x_i - \mu|^p}{p\sigma^p} \right\} + |x_i + \mu|^p \exp \left\{ -\frac{|x_i + \mu|^p}{p\sigma^p} \right\}}{p^2 \sigma^p \left[\exp \left\{ -\frac{|x_i - \mu|^p}{p\sigma^p} \right\} + \exp \left\{ -\frac{|x_i + \mu|^p}{p\sigma^p} \right\} \right]} = 0, \tag{6}$$

$$\frac{\partial \log L}{\partial p} = \frac{n}{p^2} \log p - \frac{n}{p^2} + \frac{n}{p^2} \psi \left(1 + \frac{1}{p} \right) + \sum_{i=1}^n \frac{|x_i - \mu|^p \exp \left\{ -\frac{|x_i - \mu|^p}{p\sigma^p} \right\} \left(1 - p \log |x_i - \mu| + p \log \sigma \right)}{\exp \left\{ -\frac{|x_i - \mu|^p}{p\sigma^p} \right\} + \exp \left\{ -\frac{|x_i + \mu|^p}{p\sigma^p} \right\}} + \sum_{i=1}^n \frac{|x_i + \mu|^p \exp \left\{ -\frac{|x_i + \mu|^p}{p\sigma^p} \right\} \left(1 - p \log |x_i + \mu| + p \log \sigma \right)}{p^2 \sigma^p \left[\exp \left\{ -\frac{|x_i - \mu|^p}{p\sigma^p} \right\} + \exp \left\{ -\frac{|x_i + \mu|^p}{p\sigma^p} \right\} \right]} = 0 \tag{7}$$

where $\psi(\cdot)$ is the digamma function.

It is clear that equations (5)-(7) cannot be solved explicitly since they involve nonlinear functions of the parameters. Therefore, some numerical methods should be employed to obtain the ML estimates.

2.2. LS method

The LS estimators of the parameters μ , σ and p are obtained by minimizing the following function with respect to the parameters of interest [15]:

$$S = \sum_{i=1}^n \left(F(x_{(i)}; \mu, \sigma, p) - \frac{i}{n+1} \right)^2. \quad (8)$$

Here, $F(\cdot)$ is the cdf of FEP distribution. LS estimators can also be obtained by solving the following equations:

$$\frac{\partial S}{\partial \mu} = \sum_{i=1}^n \left(F(x_{(i)}; \mu, \sigma, p) - \frac{i}{n+1} \right) \Delta_1(x_{(i)}; \mu, \sigma, p) = 0, \quad (9)$$

$$\frac{\partial S}{\partial \sigma} = \sum_{i=1}^n \left(F(x_{(i)}; \mu, \sigma, p) - \frac{i}{n+1} \right) \Delta_2(x_{(i)}; \mu, \sigma, p) = 0, \quad (10)$$

$$\frac{\partial S}{\partial p} = \sum_{i=1}^n \left(F(x_{(i)}; \mu, \sigma, p) - \frac{i}{n+1} \right) \Delta_3(x_{(i)}; \mu, \sigma, p) = 0 \quad (11)$$

where

$$\Delta_1(x_{(i)}; \mu, \sigma, p) = \frac{1}{2\Gamma(1/p)\sigma} \left[\text{sign}(x_{(i)} - \mu) \gamma'_\mu \left(\frac{1}{p}, \frac{|x_{(i)} - \mu|^p}{p\sigma^p} \right) - \text{sign}(-x_{(i)} - \mu) \gamma'_\mu \left(\frac{1}{p}, \frac{|x_{(i)} + \mu|^p}{p\sigma^p} \right) \right], \quad (12)$$

$$\Delta_2(x_{(i)}; \mu, \sigma, p) = -\frac{1}{\sigma} F(x_{(i)}; \mu, \sigma, p) + \frac{1}{2\Gamma(1/p)\sigma} \left[\text{sign}(x_{(i)} - \mu) \gamma'_\sigma \left(\frac{1}{p}, \frac{|x_{(i)} - \mu|^p}{p\sigma^p} \right) - \text{sign}(-x_{(i)} - \mu) \gamma'_\sigma \left(\frac{1}{p}, \frac{|x_{(i)} + \mu|^p}{p\sigma^p} \right) \right], \quad (13)$$

and

$$\Delta_3(x_{(i)}; \mu, \sigma, p) = \frac{\psi \left(\frac{1}{p} \right)}{p^2} F(x_{(i)}; \mu, \sigma, p) + \frac{1}{2\Gamma(1/p)\sigma} \left[\text{sign}(x_{(i)} - \mu) \gamma'_p \left(\frac{1}{p}, \frac{|x_{(i)} - \mu|^p}{p\sigma^p} \right) - \text{sign}(-x_{(i)} - \mu) \gamma'_p \left(\frac{1}{p}, \frac{|x_{(i)} + \mu|^p}{p\sigma^p} \right) \right]. \quad (14)$$

Here, $\gamma'_\mu(\cdot)$, $\gamma'_\sigma(\cdot)$ and $\gamma'_p(\cdot)$ denote the partial derivatives of incomplete gamma function with respect to the parameters μ , σ and p , respectively.

Equations (9)-(11) include nonlinear functions of the parameters and thus numerical methods should be performed.

2.3. WLS method

The WLS estimators of the parameters μ , σ and p are obtained by minimizing the following function:

$$S_w = \sum_{i=1}^n w_i \left(F(x_{(i)}; \mu, \sigma, p) - \frac{i}{n+1} \right)^2 \quad (15)$$

where w_i denotes the weights which are computed by

$$w_i = \frac{1}{\text{Var}(F(x_{(i)}))} = \frac{(n+1)^2(n+2)}{i(n-i+1)}, \quad i = 1, 2, \dots, n.$$

After incorporating the cdf of FEP distribution into the function S_w and taking the partial derivatives with respect to the parameters of interest, we obtain the following nonlinear equations:

$$\frac{\partial S_w}{\partial \mu} = \sum_{i=1}^n w_i \left(F(x_{(i)}; \mu, \sigma, p) - \frac{i}{n+1} \right) \Delta_1(x_{(i)}; \mu, \sigma, p) = 0, \quad (16)$$

$$\frac{\partial S_w}{\partial \sigma} = \sum_{i=1}^n w_i \left(F(x_{(i)}; \mu, \sigma, p) - \frac{i}{n+1} \right) \Delta_2(x_{(i)}; \mu, \sigma, p) = 0, \quad (17)$$

$$\frac{\partial S_w}{\partial p} = \sum_{i=1}^n w_i \left(F(x_{(i)}; \mu, \sigma, p) - \frac{i}{n+1} \right) \Delta_3(x_{(i)}; \mu, \sigma, p) = 0 \quad (18)$$

where Δ_1 , Δ_2 and Δ_3 are given in equations (12), (13) and (14), respectively. Numerical methods should be utilized since equations (16)-(18) cannot be solved explicitly.

2.4. CVM method

CVM method is based on minimizing the distance between the estimated cdf and empirical distribution function. Therefore, CVM estimators are in the class of minimum distance estimators which are also known as maximum goodness of fit test, see i.e. [14] and references therein.

The CVM estimators of the parameters of FEP distribution are obtained by minimizing the following function with respect to the parameters μ , σ and p :

$$C = \frac{1}{12n} + \sum_{i=1}^n \left(F(x_{(i)}; \mu, \sigma, p) - \frac{2i-1}{2n} \right)^2 \quad (19)$$

where $F(\cdot)$ is the cdf of FEP distribution. It is clear that this minimization problem yields following nonlinear equations:

$$\frac{\partial C}{\partial \mu} = \sum_{i=1}^n \left(F(x_{(i)}; \mu, \sigma, p) - \frac{2i-1}{2n} \right) \Delta_1(x_{(i)}; \mu, \sigma, p) = 0, \quad (20)$$

$$\frac{\partial C}{\partial \sigma} = \sum_{i=1}^n \left(F(x_{(i)}; \mu, \sigma, p) - \frac{2i-1}{2n} \right) \Delta_2(x_{(i)}; \mu, \sigma, p) = 0, \quad (21)$$

$$\frac{\partial C}{\partial p} = \sum_{i=1}^n \left(F(x_{(i)}; \mu, \sigma, p) - \frac{2i-1}{2n} \right) \Delta_3(x_{(i)}; \mu, \sigma, p) = 0. \quad (22)$$

Here, Δ_1 , Δ_2 and Δ_3 are same as given in equations (12), (13) and (14), respectively. As it is clear from equations (20)-(22), the explicit solutions cannot be obtained and thus numerical methods should be employed.

2.5. MPS method

The MPS method is originated by Cheng and Amin [16] and Ranneby [17]. The MPS estimator is obtained by maximizing the geometric mean of spacings. The spacings are defined as follows:

$$D_i(\mu, \sigma, p) = F(x_{(i)}; \mu, \sigma, p) - F(x_{(i-1)}; \mu, \sigma, p), \quad i = 1, 2, \dots, n + 1 \quad (23)$$

where $F(\cdot)$ is the cdf of FEP distribution, $x_{(0)} = -\infty$ and $x_{(n+1)} = +\infty$. Then, the MPS estimators of μ , σ and p are defined as maximizers of the following function:

$$D = \frac{1}{n+1} \sum_{i=1}^n \log D_i(\mu, \sigma, p). \quad (24)$$

The corresponding partial derivatives are obtained as follows:

$$\frac{\partial D}{\partial \mu} = \frac{1}{n+1} \sum_{i=1}^n \frac{\Delta_1(x_{(i)}; \mu, \sigma, p) - \Delta_1(x_{(i-1)}; \mu, \sigma, p)}{D_i(\mu, \sigma, p)} = 0, \quad (25)$$

$$\frac{\partial D}{\partial \sigma} = \frac{1}{n+1} \sum_{i=1}^n \frac{\Delta_2(x_{(i)}; \mu, \sigma, p) - \Delta_2(x_{(i-1)}; \mu, \sigma, p)}{D_i(\mu, \sigma, p)} = 0, \quad (26)$$

$$\frac{\partial D}{\partial p} = \frac{1}{n+1} \sum_{i=1}^n \frac{\Delta_3(x_{(i)}; \mu, \sigma, p) - \Delta_3(x_{(i-1)}; \mu, \sigma, p)}{D_i(\mu, \sigma, p)} = 0. \quad (27)$$

See equations (12), (13) and (14) for definitions of Δ_1 , Δ_2 and Δ_3 , respectively. It is clear that MPS estimators should also be obtained using numerical methods.

It should also be noted that if there are ties in the data set, i.e. $x_{(i)} = x_{(i-1)}$, $D_i(\mu, \sigma, p)$ is replaced by the corresponding log density [16].

3. Simulation study

This section consists of a Monte-Carlo simulation study to determine the estimator having the best performance. The bias and MSE criteria are used to compare the performances of the ML, LS, WLS, CVM and MPS estimators. The bias and MSE are respectively formulated as follows:

$$\text{Bias}(\hat{\theta}) = E(\theta - \hat{\theta}) \quad \text{and} \quad \text{MSE}(\hat{\theta}) = E(\theta - \hat{\theta})^2.$$

Here, θ shows true parameter value and stands for μ , σ or p . $\hat{\theta}$ denotes the estimator of θ obtained based on one of the methods given in the previous section. We also compute the joint efficiencies of the estimators using deficiency criterion which is formulated by

$$\text{Def} = \text{MSE}(\hat{\mu}) + \text{MSE}(\hat{\sigma}) + \text{MSE}(\hat{p}).$$

For further information on Def, see Kantar and Senoglu [13].

All computations are done using R software. Since the estimators cannot be obtained explicitly, we use “optim” function with “Nelder-Mead” algorithm. Random numbers are generated from FEP distribution using the “rpe” function in the “LaplacesDemon” package [18]. Without loss of generality, the true values of the parameters are taken as follows:

- (i) $\mu = 2$ or $\mu = 5$, (ii) $\sigma = 1$ and (iii) $p = 0.8, 1$,
1.5, 2 and 2.5.

The sample size is taken to be $n = 100$ and 200. Therefore, we have 20 different simulation schemes. The results are tabulated in Table 1 – Table 4.

Conclusions obtained from Table 1:

All estimators of μ and σ have negligible biases. In other words, they are almost unbiased. However, estimators of p , except MPS, have bias. MPS estimator of p have lower bias values.

The MSE values of ML, LS, WLS, CVM and MPS estimators of μ are low and more or less the same. This conclusion is also true for those of σ . Therefore, all of the estimators can be preferred for estimating μ and σ . For p , we have similar conclusions given for bias. In other words, the MSEs of the estimators of p are relatively large compared to those of μ and σ . Furthermore, WLS has the worst performance. MPS estimator of p is the best since it has the minimum value of MSE among the others.

As indicated previously, Def is a measure of joint efficiency. Therefore, it can be seen as total efficiency measure for ML, LS, WLS, CVM and MPS methods. From this point of view, MPS outperforms its rivals. ML method is following MPS and also giving promising results.

Conclusions obtained from Table 2:

It is clear from bias values of μ given in second column of Table 2 that all estimators are almost unbiased. However, it should be noticed that ML method has larger bias values. All estimators of σ have relatively smaller biases. It should also be mentioned that CVM and MPS underestimates σ in most of the cases. We have comments similar to those obtained from Table 1 for the estimators of p in terms of bias criterion. They have larger biases and LS and WL are more preferable in most of the cases.

The MSEs of all estimators of μ are close to each other and all of them are preferable. Similarly, all estimators of σ can also be used confidentially. However, we cannot say the same for estimating p since the corresponding MSEs are large. Among all estimators, MPS estimator of p is more preferable.

According to the Def criterion, MPS is the best estimator and it is followed by ML. There are no remarkable efficiencies of WLS and CVM but the performance of LS method is promising in some cases, i.e. $p = 0.8$.

Conclusions obtained from Table 3 and Table 4 :

The conclusions obtained from Table 3 and Table 4 are similar those obtained from Table 1 and Table 2. In Table 3 and Table 4, the sample size is increased to 200 therefore all estimators gain efficiency. For example, biases, MSEs and therefore Def values decrease as expected. Results show that MPS is more preferable to its rivals in most of the cases. ML is the second best and LS presents relatively good results in some cases.

Overall, we recommend to use MPS method used to estimate the parameters of FEP distribution since it has the smallest bias, MSE and Def values in most of the cases.

Table 1. Simulated biases, MSEs and Def values of the ML, LS, WLS, CVM and MPS estimators for $\mu = 2, \sigma = 1$ and $n = 100$.

Estimation Method	μ		σ		p		Def
	Bias	MSE	Bias	MSE	Bias	MSE	
<i>p = 0.8</i>							
ML	0.0035	0.0108	-0.0017	0.0225	0.0549	0.0578	0.0911
LS	-0.0084	0.0108	0.0281	0.0240	0.0713	0.0865	0.1214
WLS	-0.0061	0.0119	0.0219	0.0247	0.0670	0.8715	0.9081
CVM	-0.0030	0.0107	0.0072	0.0223	0.0865	0.0979	0.1310
MPS	-0.0041	0.0114	0.0147	0.0232	-0.0198	0.0394	0.0739
<i>p = 1</i>							
ML	-0.0049	0.0125	0.0009	0.0184	0.0774	0.1034	0.1342
LS	-0.0170	0.0125	0.0242	0.0192	0.0670	0.1301	0.1618
WLS	-0.0135	0.0131	0.0184	0.0193	0.0685	1.0731	1.1055
CVM	-0.0122	0.0123	0.0047	0.0179	0.0930	0.1484	0.1786
MPS	-0.0130	0.0129	0.0157	0.0190	-0.0303	0.0676	0.0995
<i>p = 1.5</i>							
ML	0.0003	0.0137	0.0138	0.0178	0.2179	0.6035	0.6351
LS	-0.0098	0.0139	0.0223	0.0173	0.1027	0.5064	0.5376
WLS	-0.0060	0.0134	0.0165	0.0173	0.1188	1.6329	1.6636
CVM	-0.0072	0.0136	0.0072	0.0166	0.1641	0.6004	0.6307
MPS	-0.0072	0.0142	0.0176	0.0173	-0.0047	0.3261	0.3576
<i>p = 2</i>							
ML	-0.0044	0.0121	0.0252	0.0172	0.3314	0.9214	0.9508
LS	-0.0095	0.0126	0.0232	0.0150	0.1414	0.9621	0.9897
WLS	-0.0071	0.0118	0.0186	0.0146	0.1702	2.1992	2.2256
CVM	-0.0082	0.0124	0.0117	0.0147	0.2408	1.1537	1.1808
MPS	-0.0095	0.0125	0.0174	0.0146	-0.0045	0.5180	0.5452
<i>p = 2.5</i>							
ML	-0.0059	0.0092	0.0178	0.0171	0.3494	1.1601	1.1865
LS	-0.0060	0.0105	0.0094	0.0138	0.2253	2.4578	2.4821
WLS	-0.0055	0.0094	0.0054	0.0124	0.1919	2.7287	2.7505
CVM	-0.0054	0.0103	0.0003	0.0139	0.3618	2.7212	2.7454
MPS	-0.0077	0.0095	0.0015	0.0123	-0.0586	0.9356	0.9575

Table 2. Simulated biases, MSEs and Def values of the ML, LS, WLS, CVM and MPS estimators for $\mu = 5, \sigma = 1$ and $n = 100$.

Estimation Method	μ		σ		p		Def
	Bias	MSE	Bias	MSE	Bias	MSE	
$p = 0.8$							
ML	-0.0762	0.0099	-0.0092	0.0204	0.0131	0.0338	0.0641
LS	0.0055	0.0125	0.0229	0.0205	0.0146	0.0485	0.0815
WLS	0.0064	0.0136	0.0147	0.0204	0.0217	0.8222	0.8562
CVM	0.0064	0.0126	0.0059	0.0196	0.0431	0.0547	0.0869
MPS	0.0060	0.0102	0.0018	0.0190	-0.0674	0.0286	0.0577
$p = 1$							
ML	-0.0810	0.0107	-0.0106	0.0183	0.0105	0.0502	0.0792
LS	0.0024	0.0122	0.0142	0.0177	0.0278	0.0802	0.1101
WLS	0.0029	0.0127	0.0072	0.0177	0.0334	1.0345	1.0650
CVM	0.0027	0.0122	-0.0003	0.0174	0.0652	0.0920	0.1216
MPS	0.0015	0.0117	-0.0049	0.0171	-0.0835	0.0451	0.0738
$p = 1.5$							
ML	-0.0901	0.0124	0.0059	0.0140	0.1186	0.2061	0.2325
LS	0.0035	0.0117	0.0075	0.0151	0.0568	0.2770	0.3037
WLS	0.0028	0.0116	0.0028	0.0145	0.0609	1.5646	1.5908
CVM	0.0035	0.0117	-0.0037	0.0151	0.1205	0.3272	0.3540
MPS	0.0043	0.0125	-0.0091	0.0136	-0.1139	0.1162	0.1424
$p = 2$							
ML	-0.0838	0.0111	0.0105	0.0148	0.2480	0.7436	0.7695
LS	-0.0005	0.0117	0.0040	0.0137	0.0943	0.8095	0.8349
WLS	0.0004	0.0112	0.0006	0.0128	0.0932	2.1019	2.1259
CVM	-0.0005	0.0117	-0.0053	0.0139	0.1936	0.9926	1.0183
MPS	0.0003	0.0114	-0.0083	0.0119	-0.1309	0.3034	0.3268
$p = 2.5$							
ML	-0.0733	0.0085	0.0139	0.0152	0.3781	1.8727	1.8965
LS	0.0008	0.0095	0.0028	0.0155	0.1895	2.4575	2.4825
WLS	0.0008	0.0085	-0.0006	0.0139	0.1587	2.6839	2.7063
CVM	0.0007	0.0095	-0.0051	0.0158	0.3405	3.2004	3.2257
MPS	0.0006	0.0083	-0.0111	0.0124	-0.1708	0.6494	0.6701

Table 3. Simulated biases, MSEs and Def values of the ML, LS, WLS, CVM and MPS estimators for $\mu = 2, \sigma = 1$ and $n = 200$.

Estimation Method	μ		σ		p		Def
	Bias	MSE	Bias	MSE	Bias	MSE	
$p = 0.8$							
ML	-0.0017	0.0052	-0.0009	0.0125	0.0120	0.0187	0.0364
LS	-0.0077	0.0056	0.0139	0.0129	0.0234	0.0306	0.0491
WLS	-0.0065	0.0061	0.0100	0.0129	0.0181	0.8185	0.8375
CVM	-0.0050	0.0056	0.0033	0.0125	0.0292	0.0322	0.0503
MPS	-0.0050	0.0053	0.0075	0.0127	-0.0292	0.0166	0.0345
$p = 1$							
ML	-0.0014	0.0065	-0.0010	0.0092	0.0429	0.0355	0.0512
LS	-0.0057	0.0064	0.0104	0.0094	0.0530	0.0639	0.0797
WLS	-0.0042	0.0068	0.0071	0.0096	0.0440	1.0459	1.0622
CVM	-0.0034	0.0064	0.0007	0.0091	0.0652	0.0687	0.0842
MPS	-0.0064	0.0066	0.0064	0.0094	-0.0187	0.0283	0.0443
$p = 1.5$							
ML	0.0031	0.0059	0.0012	0.0069	0.0802	0.1021	0.1150
LS	-0.0007	0.0059	0.0067	0.0070	0.0393	0.1513	0.1643
WLS	0.0010	0.0059	0.0040	0.0070	0.0522	1.5549	1.5677
CVM	0.0004	0.0059	-0.0007	0.0069	0.0675	0.1643	0.1771
MPS	-0.0006	0.0060	0.0019	0.0068	-0.0376	0.0774	0.0902
$p = 2$							
ML	0.0009	0.0052	0.0100	0.0065	0.1262	0.2496	0.2613
LS	-0.0009	0.0058	0.0081	0.0067	0.0219	0.3589	0.3715
WLS	0.0001	0.0054	0.0070	0.0064	0.0541	2.0570	2.0688
CVM	-0.0004	0.0058	0.0025	0.0067	0.0661	0.3900	0.4025
MPS	-0.0014	0.0053	0.0042	0.0060	-0.0591	0.1720	0.1833
$p = 2.5$							
ML	0.0008	0.0042	0.0052	0.0064	0.1546	0.4281	0.4387
LS	0.0020	0.0050	0.0028	0.0072	0.0674	0.6629	0.6751
WLS	0.0018	0.0045	0.0011	0.0066	0.0739	2.5793	2.5904
CVM	0.0022	0.0050	-0.0015	0.0072	0.1292	0.7253	0.7375
MPS	0.0000	0.0043	-0.0045	0.0059	-0.0949	0.2915	0.3017

Table 4. Simulated biases, MSEs and Def values of the ML, LS, WLS, CVM and MPS estimators for $\mu = 5, \sigma = 1$ and $n = 200$.

Estimation Method	μ		σ		p		Def
	Bias	MSE	Bias	MSE	Bias	MSE	
$p = 0.8$							
ML	-0.0538	0.0048	-0.0068	0.0090	0.0054	0.0146	0.0284
LS	-0.0017	0.0062	0.0090	0.0091	0.0019	0.0211	0.0364
WLS	-0.0013	0.0068	0.0055	0.0091	0.0096	0.8097	0.8255
CVM	-0.0013	0.0062	0.0006	0.0089	0.0152	0.0222	0.0373
MPS	0.0008	0.0047	-0.0022	0.0088	-0.0439	0.0144	0.0279
$p = 1$							
ML	-0.0631	0.0062	-0.0057	0.0086	0.0032	0.0253	0.0401
LS	0.0009	0.0068	0.0049	0.0084	0.0063	0.0385	0.0537
WLS	0.0008	0.0072	0.0019	0.0083	0.0140	1.0142	1.0296
CVM	0.0010	0.0068	-0.0024	0.0083	0.0236	0.0409	0.0560
MPS	0.0020	0.0063	-0.0058	0.0082	-0.0545	0.0245	0.0389
$p = 1.5$							
ML	-0.0643	0.0065	0.0010	0.0072	0.0509	0.0580	0.0718
LS	-0.0003	0.0062	0.0052	0.0078	0.0320	0.1262	0.1402
WLS	-0.0004	0.0061	0.0032	0.0074	0.0381	1.5396	1.5531
CVM	-0.0003	0.0062	-0.0005	0.0078	0.0613	0.1368	0.1507
MPS	-0.0004	0.0062	-0.0057	0.0069	-0.0740	0.0519	0.0651
$p = 2$							
ML	-0.0584	0.0055	0.0048	0.0059	0.0984	0.1578	0.1692
LS	-0.0007	0.0058	0.0039	0.0072	0.0435	0.3222	0.3352
WLS	-0.0013	0.0054	0.0028	0.0066	0.0498	2.0523	2.0643
CVM	-0.0007	0.0058	-0.0008	0.0073	0.0869	0.3524	0.3655
MPS	-0.0018	0.0053	-0.0068	0.0060	-0.1061	0.1184	0.1297
$p = 2.5$							
ML	-0.0510	0.0041	-0.0005	0.0059	0.1202	0.2914	0.3015
LS	0.0008	0.0051	-0.0054	0.0070	0.0309	0.5913	0.6034
WLS	0.0006	0.0046	-0.0063	0.0063	0.0368	2.5381	2.5490
CVM	0.0008	0.0051	-0.0095	0.0071	0.0900	0.6484	0.6606
MPS	0.0009	0.0043	-0.0156	0.0058	-0.1685	0.2196	0.2297

4. Applications

In this section two real life data set taken from the literature is analyzed. It should be mentioned that we use ML and MPS methods in this section since they are more preferable in terms of efficiency, see Table 1 - Table 4. In applications, we also conduct Kolmogorov-Simornov tests. Therefore, test statistic (KS) and corresponding p –values obtained from Kolmogrov-Smirnov test are also provided. The method having smaller (larger) KS (p –) value is more desirable.

4.1. Lead wires data

This application is taken from Leone et al. [5] in which FN distribution is used. The data consists of camber of 497 lead wires which are used in manufacture of miniature radio tubes. The full data set and more explanation are available in Leone et al. [5]. See also Cooray et al. [8] in which FL distribution is considered for modelling purposes. Different from these studies, we here use FEP distribution to model lead wires data. Parameter estimates based on FEP, FN and FL distributions and corresponding KS test results are given in Table 5. It should be noted that parameter estimates are obtained using the ML and MPS methods for FEP distribution. On the other hand, results for FN and FL distributions are based on the ML method. It is clear from Table 5 that the ML is more desirable according to KS and p – values for FEP distribution. It should be recognized that the fitting performances of the ML and MPS methods are similar, see Figure 2 in which histogram of data is given along with fitted densities based on ML and MPS estimates. Cooray et al. [8] show that FL distribution is more appropriate than FN distribution for modelling this data. Results given in Table 5 provide that the FEP distribution can also be considered as an alternative to FL distribution.

Table 5. Parameter estimates, KS and p –values for lead wires data.

Distribution	Estimation Method	$\hat{\mu}$	$\hat{\sigma}$	\hat{p}	KS	p –value
FEP	ML	13.1291	7.7628	1.4488	0.0612	0.0482
	MPS	13.0694	7.7759	1.3951	0.0632	0.0375
FN	ML	13.6120	8.4626	-	0.0709	0.0135
FL	ML	13.2860	4.8400	-	0.0611	0.0487

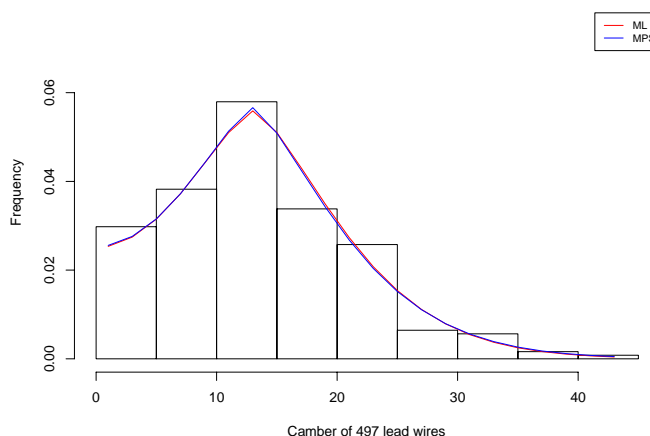


Figure 2. The histogram and fitted densities for lead wires data.

4.2. Log-Norwegian fire claim data

Brazauskas and Kleefeld [2] use this data in the context of insurance claims distributional modeling. The full data set is available at the following web site with name NORWEGIANFIRE.TXT: <http://lstat.kuleuven.be/Wiley/>. It should be noted that log-Norwegian fire claim data is obtained by taking logarithm of original data divided by 500 and is for the year 1988.

Nadarajah and Bakar [1] consider different folded distributions, including FEP, to model log-Norwegian fire claim data. They show that FEP is the best distribution modeling log-Norwegian fire claim by using ML estimates. In addition to ML, we use MPS method to estimate the parameters of FEP distribution in the current study. Results are given in Table 6. It is clear that MPS method is more preferable for this data set. See also Figure 3 in which histogram and fitted densities are provided.

Table 6. Parameter estimates, KS and p –values for log-Norwegian fire claim data.

Distribution	Estimation Method	$\hat{\mu}$	$\hat{\sigma}$	\hat{p}	KS	p –value
FEP	ML	0.6109	0.9322	1.1282	0.0213	0.8475
	MPS	0.6109	0.9259	1.0981	0.0211	0.8566

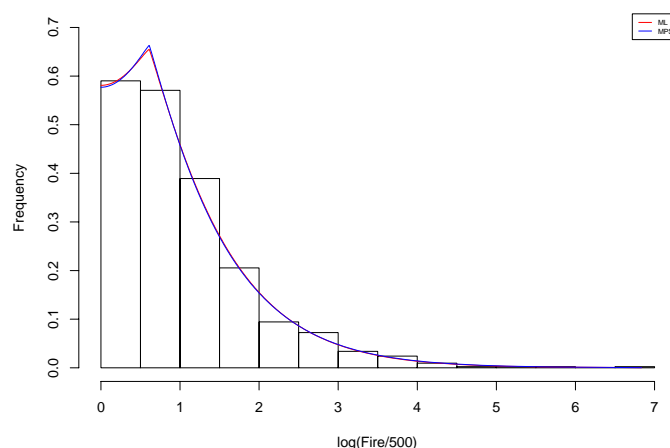


Figure 3. The histogram and fitted densities log-Norwegian fire claim data.

5. Conclusion

In this study, we consider the ML, LS, WLS, CVM and MPS methods to estimate the unknown parameters of FEP distribution. As far as we know, LS, WLS, CVM and MPS methods have not been used for FEP distribution previously. We aim to investigate the most efficient estimator in terms of some criteria such as bias, MSE and Def. Results of the simulation study show that MPS is slightly better than the other estimators. It should be noticed that the ML estimator has also a remarkable performance since it is the second best estimator. Two real life examples taken from literature are considered in the application, i.e. Lead wires and log Norwegian data. They are modelled using FEP distribution and the corresponding parameter estimates are obtained based on ML and MPS methods. KS test results show that these methods are desirable for modelling the data sets.

References

- [1] S. Nadarajah, S.A.A. Bakar, 2015, New folded models for the log-transformed Norwegian fire claim data, *Communications in Statistics-Theory and Methods*, 44(20), 4408-4440.
- [2] V. Brazauskas, A. Kleefeld, 2011, Folded and log-folded-t distributions as models for insurance loss data, *Scandinavian Actuarial Journal*, 2011(1), 59-74.
- [3] M. Chen, F. Kianifard, 2003, Estimation of treatment difference and standard deviation with blinded data in clinical trials, *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 45(2), 135-142.
- [4] S. Jung, M. Foskey, J. S. Marron, 2011, Principal arc analysis on direct product manifolds, *The Annals of Applied Statistics*, 5(1), 578-603.
- [5] F. C. Leone, L. S. Nelson, R. B. Nottingham, 1961, The folded normal distribution, *Technometrics*, 3, 543-550.
- [6] M. Y. Liao, 2010, Economic tolerance design for folded normal data, *International Journal of Production Research*, 48(14), 4123-4137.
- [7] S. Psarakis, J. Panaretos, 1990, The folded t distribution, *Communications in Statistics-Theory and Methods*, 19, 2717-2734.
- [8] K. Cooray, S. Gunasekera, M. M. A. Ananda, 2006, The folded logistic distribution, *Communications in Statistics-Theory and Methods*, 35, 385-393.
- [9] M. T. Subbotin, 1923, On the law of frequency of errors, *Matematicheskii Sbornik*, 31, 296-301.
- [10] K. Cooray, 2010, Generalized Gumbel distribution, *Journal of Applied Statistics*, 37, 171-179.
- [11] W. Gui, 2013, Statistical inferences and applications of the half exponential power distribution, *Journal of Quality and Reliability Engineering*, Article ID 219473, 9 pages.
- [12] F. G. Akgul, 2018, Comparison of the estimation methods for the parameters of exponentiated reduced kies distribution, *Süleyman Demirel University Journal of Natural and Applied Sciences*, 22(3), 1209-1216.
- [13] Y. M. Kantar, B. Senoglu, 2008, A comparative study for the location and scale parameters of the Weibull distribution with given shape parameter, *Computers & Geosciences*, 34(12), 1900-1909.
- [14] F. Louzada, P. L. Ramos, G. S. Perdoná, 2016, Different estimation procedures for the parameters of the extended exponential geometric distribution for medical data, *Computational and mathematical methods in medicine*, Article ID 8727951, 12 pages.
- [15] J. J. Swain, S. Venkatraman, J. R. Wilson, 1988, Least-squares estimation of distribution functions in Johnson's translation system, *Journal of Statistical Computation and Simulation*, 29(4), 271-297.
- [16] R. C. H. Cheng, N. A. K. Amin, 1983, Estimating parameters in continuous univariate distributions with a shifted origin, *Journal of the Royal Statistical Society: Series B (Methodological)*, 45(3), 394-403.
- [17] B. Ranney, 1984, The maximum spacing method. An estimation method related to the maximum likelihood method, *Scandinavian Journal of Statistics*, 11(2), 93-112.
- [18] Statisticat, LLC., 2018, LaplacesDemon: Complete Environment for Bayesian Inference. Bayesian-Inference.com. R package version 16.1.1.
<https://web.archive.org/web/20150206004624/http://www.bayesian-inference.com/software>.




A class of estimators in median ranked set sampling

Nursel Koyuncu

Hacettepe University, Department of Statistics

Beytepe, Ankara, TURKEY

nkoyuncu@hacettepe.edu.tr

 0000-0003-1065-3411

Abstract

In this paper we have proposed a general class of estimators under median ranked set sampling (MRSS) both one auxiliary and multi-auxiliary variables are known. The mean square error (MSE) and bias formulas are derived by theoretically. We have conducted a simulation study to see the performance of proposed estimators. Also it is shown that the proposed estimators are always more efficient than existing estimators by theoretically. From simulation study we can say that suggested class of estimator performs better than Al-Omari [1] estimator and conclude that regression type estimator gives always more efficient results than Al-Omari [1] estimator.

Keywords: Auxiliary information, Efficiency, Median ranked set sampling, Monte Carlo simulation, Ratio estimator.

Öz

Bu çalışmada, medyan sıralı küme örnekleme (MSKÖ) altında hem bir yardımcı değişken hem de birden çok yardımcı değişken bilinmesi durumunda genel bir tahmin edici sınıfı önerilmiştir. Hata kareler ortalaması (HKO) ve yanlılık formülleri teorik olarak elde edilmiştir. Önerilen tahmin edicilerin performansını görmek için bir benzetim çalışması yapılmıştır. Ayrıca önerilen tahmin edicilerin mevcut tahmin edicilerden her zaman daha etkin olduğu teorik olarak gösterilmiştir. Benzetim çalışmasından, önerilen tahmin edici sınıfının Al-Omari [1] tahmin edicisinden daha iyi performans gösterdiği söylenebilir ve regresyon tipi tahmin edicinin, Al-Omari[1] tahmin edicisinden her zaman daha etkin sonuçlar verdiği sonucuna varılabilir.

Anahtar sözcükler: Yardımcı değişken, Etkinlik, Medyan sıralı küme örnekleme, Monte Carlo benzetimi, Oransal Tahmin.

1. Introduction

Ranked set sampling (RSS) is more efficient sampling plan than simple random sampling (SRS) when the measurements are difficult or expensive to obtain but ranking of units is relatively easy and cheap. Many authors developed and modified this sampling plan to estimate population parameters such as Al-Saleh and Al-Omari [2], Jemain and Al-Omari [3], Jemain et al. [4], Ozturk and Jafari Jozani [5], Amiri et al. [6] etc. Recently Al-Omari and Al-Nasser [7], used robust extreme ranked set sampling to estimate median. Al-Omari and Raqab [8] studied truncation-based ranked set samples to estimate population mean and

median. Haq et al. [9-10] introduced partial ranked set sampling and mixed ranked set sampling designs respectively. Some authors preferred various ranked set sampling designs to estimate parameters of distributions. Bhoj and Kushary [11-12] proposed a ranked set sampling procedure with unequal samples and unequal replications to estimate the population mean and have extended their study for skewed distributions. Samuh and Qtait [13] used median ranked set sampling to estimate parameters of exponentiated exponential distribution. Omar and Ibrahim [14] used extreme ranked set sampling to estimate shape and scale parameters of the Pareto distribution.

In sampling literature another way to get more efficient estimates is that using ratio estimator when the information of auxiliary variable is available. Moving this point, many authors proposed ratio estimators under different sampling designs (see [15-17]). But in median ranked set sampling (MRSS) there are a few studies related with ratio estimators. Al-Omari [1] first defined ratio estimators in MRSS and Koyuncu [18-19] proposed new ratio estimators in MRSS and extended ratio estimators to extreme ranked set and double robust extreme ranked set sampling.

In this study we have proposed more general class of estimators in MRSS when one auxiliary is known. Also we have extended our results for the case: multi-auxiliary variables are known. SRS and MRSS are defined in Section2 and Section3 respectively. In Section4, we have suggested a class of estimators in MRSS. Simulation study is given in Section5 and we have summarized our results in this section.

2. Simple Random Sampling

We have reviewed some well known estimators in SRS. Let $U = (u_1, u_2, \dots, u_N)$ denote a finite population of size N . Let Y be the study; X and Z be two auxiliary variables associated with each unit $u_j (j = 1, \dots, N)$ of the population. A sample of size n is drawn without replacement from the population. Let $(X_1, Z_1, Y_1), (X_2, Z_2, Y_2), \dots, (X_n, Z_n, Y_n)$ denote the observed values of Y, X and Z . Moreover let

$\mu_y, \mu_x, \mu_z, \bar{y}_{SRS} = \frac{\sum_{i=1}^n Y_i}{n}, \bar{x}_{SRS} = \frac{\sum_{i=1}^n X_i}{n}, \bar{z}_{SRS} = \frac{\sum_{i=1}^n Z_i}{n}$ be the population and sample means of study and two auxiliary variables respectively.

The well known ratio estimator using SRS is defined by

$$\bar{y}_{R_SRS} = \bar{y}_{SRS} \frac{\mu_x}{\bar{x}_{SRS}} \tag{1}$$

where $\bar{x}_{SRS} \neq 0$. If the quantiles of auxiliary variable, q_1, q_3 are known Al-Omari [1] defined ratio type estimators in SRS as

$$\bar{y}_{R1_SRS} = \bar{y}_{SRS} \frac{\mu_x + q_1}{\bar{x}_{SRS} + q_1} \tag{2}$$

$$\bar{y}_{R2_SRS} = \bar{y}_{SRS} \frac{\mu_x + q_3}{\bar{x}_{SRS} + q_3} \tag{3}$$

The regression-type estimator of population mean using one auxiliary variable is defined by

$$\bar{y}_{Reg1_SRS} = \bar{y}_{SRS} + b_{SRS1} (\mu_x - \bar{x}_{SRS}) \tag{4}$$

where $b_{SRS1} = \frac{S_{SRS-yx}}{S_{SRS-x}^2}$ denotes the estimator of regression coefficient and $s_{SRS-x}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x}_{SRS})^2}{n-1}$,

$$s_{SRS-yx} = \frac{\sum_{i=1}^n (y_i - \bar{y}_{SRS})(x_i - \bar{x}_{SRS})}{n-1}, [20].$$

If two auxiliary variables are available the ratio type and regression-type estimators are, respectively, defined as

$$\bar{y}_{R3_SRS} = \bar{y}_{SRS} \frac{\mu_x}{\bar{x}_{SRS}} \frac{\mu_z}{\bar{z}_{SRS}} \tag{5}$$

$$\bar{y}_{Reg2_SRS} = \bar{y}_{SRS} + b_{SRS1}(\mu_x - \bar{x}_{SRS}) + b_{SRS2}(\mu_z - \bar{z}_{SRS}) \tag{6}$$

where $b_{SRS2} = \frac{S_{SRS-yz}}{S_{SRS-z}^2}$ denotes the estimator of regression coefficient for second auxiliary variable and

$$s_{SRS-z}^2 = \frac{\sum_{i=1}^n (z_i - \bar{z}_{SRS})^2}{n-1}, \quad s_{SRS-yz} = \frac{\sum_{i=1}^n (y_i - \bar{y}_{SRS})(z_i - \bar{z}_{SRS})}{n-1}, [20].$$

3. Median Ranked Set Sampling

For the sake of brevity we follow Al-Omari [1] sampling design and notations given by

1. Randomly select n samples each of size n multivariate units from the population.
2. The units within each sample are ranked with respect to a variable of interest.
3. If n is odd, select the $\left(\frac{n+1}{2}\right)$ th-smallest ranked unit auxiliary variables together with the associated study variable from each set, i.e., the median of each set. If n is even, from the first $\frac{n}{2}$ sets select the $\left(\frac{n}{2}\right)$ th ranked unit auxiliary variables together with the associated study variable and from the other $\frac{n}{2}$ sets select the $\left(\frac{n+2}{2}\right)$ th ranked unit auxiliary variables together with the associated study variables.
4. The above procedure can be repeated m times to obtain a sample of size nm units.

To extend our results to multivariate case we assume that Y is study variable and X and Z are auxiliary variable. In application to ranked the samples we prefer highly correlated auxiliary variable with study variable.

Let $(X_{i(1)}, Z_{i(1)}, Y_{i[1]}), (X_{i(2)}, Z_{i(2)}, Y_{i[2]}), \dots, (X_{i(n)}, Z_{i(n)}, Y_{i[n]})$ be the order statistics of $X_{i1}, X_{i2}, \dots, X_{in}$, $Z_{i1}, Z_{i2}, \dots, Z_{in}$ and the judgement order of $Y_{i1}, Y_{i2}, \dots, Y_{in}$ ($i = 1, 2, \dots, n$), where $()$ and $[]$ indicate that the ranking of X, Z is perfect and ranking of Y has errors. For odd and even sample sizes the units measured using MRSS are denoted by MRSSO and MRSSE, respectively.

For odd sample size let

$$\left(X_{1\binom{n+1}{2}}, Z_{1\binom{n+1}{2}}, Y_{1\lceil\frac{n+1}{2}\rceil} \right), \left(X_{2\binom{n+1}{2}}, Z_{2\binom{n+1}{2}}, Y_{2\lceil\frac{n+1}{2}\rceil} \right), \dots, \left(X_{n\binom{n+1}{2}}, Z_{n\binom{n+1}{2}}, Y_{n\lceil\frac{n+1}{2}\rceil} \right)$$

denote the observed units by MRSSO. $\bar{x}_{MRSSO} = \frac{1}{n} \sum_{i=1}^n X_{i\binom{n+1}{2}}$, $\bar{z}_{MRSSO} = \frac{1}{n} \sum_{i=1}^n Z_{i\binom{n+1}{2}}$ and

$\bar{y}_{MRSSO} = \frac{1}{n} \sum_{i=1}^n Y_{i\lceil\frac{n+1}{2}\rceil}$ be the sample mean of X, Z and Y respectively.

For even sample size let

$$\left(X_{1\binom{n}{2}}, Z_{1\binom{n}{2}}, Y_{1\lceil\frac{n}{2}\rceil} \right), \left(X_{2\binom{n}{2}}, Z_{2\binom{n}{2}}, Y_{2\lceil\frac{n}{2}\rceil} \right), \dots, \left(X_{\frac{n}{2}\binom{n}{2}}, Z_{\frac{n}{2}\binom{n}{2}}, Y_{\frac{n}{2}\lceil\frac{n}{2}\rceil} \right), \left(X_{\frac{n+2}{2}\binom{n+2}{2}}, Z_{\frac{n+2}{2}\binom{n+2}{2}}, Y_{\frac{n+2}{2}\lceil\frac{n+2}{2}\rceil} \right), \dots, \left(X_{n\binom{n+2}{2}}, Z_{n\binom{n+2}{2}}, Y_{n\lceil\frac{n+2}{2}\rceil} \right)$$

denote the observed units by MRSSE.

$$\bar{x}_{MRSSE} = \frac{1}{n} \left(\sum_{i=1}^{\frac{n}{2}} X_{i\binom{n}{2}} + \sum_{i=\frac{n+2}{2}}^n X_{i\binom{n+2}{2}} \right), \bar{z}_{MRSSE} = \frac{1}{n} \left(\sum_{i=1}^{\frac{n}{2}} Z_{i\binom{n}{2}} + \sum_{i=\frac{n+2}{2}}^n Z_{i\binom{n+2}{2}} \right)$$

and $\bar{y}_{MRSSE} = \frac{1}{n} \left(\sum_{i=1}^{\frac{n}{2}} Y_{i\lceil\frac{n}{2}\rceil} + \sum_{i=\frac{n+2}{2}}^n Y_{i\lceil\frac{n+2}{2}\rceil} \right)$ be the sample mean of X, Z and Y respectively. Let rewrite the

sample means as $\bar{x}_{MRSS(j)}, \bar{z}_{MRSS(j)}, \bar{y}_{MRSS(j)}$ where $j = (E, O)$ denote the sample size even or odd.

In MRSS we can re-write estimators in defined in section two as given by, respectively

$$\bar{y}_{R_MRSS} = \bar{y}_{MRSS(j)} \frac{\mu_x}{\bar{x}_{MRSS(j)}} \tag{7}$$

$$\bar{y}_{R1_MRSS} = \bar{y}_{MRSS(j)} \frac{\mu_x + q_1}{\bar{x}_{MRSS(j)} + q_1} \tag{8}$$

$$\bar{y}_{R2_MRSS} = \bar{y}_{MRSS(j)} \frac{\mu_x + q_3}{\bar{x}_{MRSS(j)} + q_3} \tag{9}$$

$$\bar{y}_{Reg1_MRSS} = \bar{y}_{MRSS(j)} + b_{MRSS1} (\mu_x - \bar{x}_{MRSS(j)}) \tag{10}$$

where b_{MRSS1} denotes the estimator of regression coefficient under MRSS

$$b_{MRSS1(j)} \cong \begin{cases} \frac{\rho_{xy\left[\frac{n+1}{2}\right]} s_{yx\left[\frac{n+1}{2}\right]}}{s_x^2\left(\frac{n+1}{2}\right)}, n \text{ is odd} \\ \frac{\left(\rho_{xy\left[\frac{n}{2}\right]} + \rho_{xy\left[\frac{n+2}{2}\right]}\right) \left(s_{yx\left[\frac{n}{2}\right]} + s_{yx\left[\frac{n+2}{2}\right]}\right)}{\left(s_x^2\left(\frac{n}{2}\right) + s_x^2\left(\frac{n+2}{2}\right)\right)}, n \text{ is even} \end{cases}$$

$$s_{xy\left[\frac{n+1}{2}\right]} = \rho_{xy\left[\frac{n+1}{2}\right]} s_x\left(\frac{n+1}{2}\right) s_{y\left[\frac{n+1}{2}\right]}, \left(s_{xy\left[\frac{n}{2}\right]} + s_{xy\left[\frac{n+2}{2}\right]}\right) = \left(\rho_{xy\left[\frac{n}{2}\right]} + \rho_{xy\left[\frac{n+2}{2}\right]}\right) \left(s_x\left[\frac{n}{2}\right] + s_x\left[\frac{n+2}{2}\right]\right) \left(s_y\left[\frac{n}{2}\right] + s_y\left[\frac{n+2}{2}\right]\right)$$

and $\rho_{xy\left[\frac{n+1}{2}\right]}, \rho_{xy\left[\frac{n}{2}\right]}$ represent correlation coefficient between study and auxiliary variable for odd and even sample sizes.

If the two auxiliary variables are known, following estimators can be used

$$\bar{y}_{R3_MRSS} = \bar{y}_{MRSS(j)} \frac{\mu_x}{\bar{x}_{MRSS(j)}} \frac{\mu_z}{\bar{z}_{MRSS(j)}} \tag{11}$$

$$\bar{y}_{Reg2_MRSS} = \bar{y}_{MRSS(j)} + b_{MRSS1} (\mu_x - \bar{x}_{MRSS(j)}) + b_{MRSS2} (\mu_z - \bar{z}_{MRSS(j)}) \tag{12}$$

where b_{MRSS2} denotes the estimator of regression coefficient of second auxiliary variable under MRSS

$$b_{MRSS2(j)} \cong \begin{cases} \frac{\rho_{zy\left[\frac{n+1}{2}\right]} s_{yz\left[\frac{n+1}{2}\right]}}{s_z^2\left(\frac{n+1}{2}\right)}, n \text{ is odd} \\ \frac{\left(\rho_{zy\left[\frac{n}{2}\right]} + \rho_{zy\left[\frac{n+2}{2}\right]}\right) \left(s_{yz\left[\frac{n}{2}\right]} + s_{yz\left[\frac{n+2}{2}\right]}\right)}{\left(s_z^2\left(\frac{n}{2}\right) + s_z^2\left(\frac{n+2}{2}\right)\right)}, n \text{ is even} \end{cases}$$

$$s_{zy\left[\frac{n+1}{2}\right]} = \rho_{zy\left[\frac{n+1}{2}\right]} s_z\left(\frac{n+1}{2}\right) s_{y\left[\frac{n+1}{2}\right]},$$

$$\left(s_{zy\left[\frac{n}{2}\right]} + s_{zy\left[\frac{n+2}{2}\right]}\right) = \left(\rho_{zy\left[\frac{n}{2}\right]} + \rho_{zy\left[\frac{n+2}{2}\right]}\right) \left(s_z\left[\frac{n}{2}\right] + s_z\left[\frac{n+2}{2}\right]\right) \left(s_y\left[\frac{n}{2}\right] + s_y\left[\frac{n+2}{2}\right]\right).$$

4. Suggested Classes of Estimators

Following Srivastava and *Jhajj* [21] and Koyuncu and Kadilar [22], we have defined a class of estimators of population mean using MRSS as

$$t_g = \bar{y}_{MRSS(j)} H(u) \tag{13}$$

where $u = \frac{\bar{x}_{MRSS(j)}}{\mu_x}$ and $H(\cdot)$ is a function of u satisfies the following regularity conditions:

a) $H(1) = 1$

b) The first and second order partial derivatives of H with respect to u exist and are known constants at a given point $u = 1$. Note that the estimators are defined in (7)-(9) are members of t_g . Let us define following expectations under MRSS as follows

$$\varepsilon_{0(j)} = \frac{\bar{y}_{MRSS(j)} - \mu_y}{\mu_y}, \quad \varepsilon_{1(j)} = \frac{\bar{x}_{MRSS(j)} - \mu_x}{\mu_x},$$

such that $E(\varepsilon_{0(j)}) = E(\varepsilon_{1(j)}) = 0$ and $j = (E, O)$ denote the sample size even or odd.

If sample size n is odd we can write

$$E(\varepsilon_{1(o)}^2) = \frac{1}{n\mu_x^2} \sigma_{x\left(\frac{n+1}{2}\right)}^2, \quad E(\varepsilon_{0(o)}^2) = \frac{1}{n\mu_y^2} \sigma_{y\left(\frac{n+1}{2}\right)}^2, \quad E(\varepsilon_{0(o)}\varepsilon_{1(o)}) = \frac{1}{n\mu_y\mu_x} \sigma_{xy\left(\frac{n+1}{2}\right)}$$

If sample size n is even we can write

$$E(\varepsilon_{1(E)}^2) = \frac{1}{2n\mu_x^2} \left(\sigma_{x\left(\frac{n}{2}\right)}^2 + \sigma_{x\left(\frac{n+2}{2}\right)}^2 \right), \quad E(\varepsilon_{0(E)}^2) = \frac{1}{2n\mu_y^2} \left(\sigma_{y\left[\frac{n}{2}\right]}^2 + \sigma_{y\left[\frac{n+2}{2}\right]}^2 \right)$$

$$E(\varepsilon_{0(E)}\varepsilon_{1(E)}) = \frac{1}{2n\mu_x\mu_y} \left(\sigma_{yx\left[\frac{n}{2}\right]} + \sigma_{yx\left[\frac{n+2}{2}\right]} \right)$$

Expanding $H(u)$ in a second order Taylor's series we have

$$H(u) = H(1 + (u-1)) = H(1) + (u-1) \frac{\partial H}{\partial u} \Big|_{u=1} + (u-1)^2 \frac{1}{2} \frac{\partial^2 H}{\partial u^2} \Big|_{u=1} + \dots \tag{14}$$

Note that $|u - 1| < 1$ thus the higher order terms can be neglected. Putting (14) in (13) we have

$$t_g = \bar{y}_{MRSS(j)} \left[1 + (u-1)H_1 + (u-1)^2 H_2 + \dots \right] \tag{15}$$

where $H_1 = \frac{\partial H}{\partial u} \Big|_{u=1}$, $H_2 = \frac{1}{2} \frac{\partial^2 H}{\partial u^2} \Big|_{u=1}$. Expressing (15) with ε terms and extracting μ_y from both sides we have

$$t_g - \mu_y = \mu_y \left[\varepsilon_{1(j)} H_1 + \varepsilon_{1(j)}^2 H_2 + \varepsilon_{0(j)} + \varepsilon_{0(j)} \varepsilon_{1(j)} H_1 + \dots \right] \tag{16}$$

Taking expectation on both sides, the bias of class is given by

$$B(t_g) \cong \begin{cases} \frac{1}{n\mu_x} \left[\frac{\mu_y}{\mu_x} \sigma_{x\left(\frac{n+1}{2}\right)}^2 H_2 + \sigma_{xy\left(\frac{n+1}{2}\right)} H_1 \right] & \text{if odd} \\ \frac{1}{2n\mu_x} \left[\frac{\mu_y}{\mu_x} \left(\sigma_{x\left(\frac{n}{2}\right)}^2 + \sigma_{x\left(\frac{n+2}{2}\right)}^2 \right) H_2 + \left(\sigma_{yx\left[\frac{n}{2}\right]} + \sigma_{yx\left[\frac{n+2}{2}\right]} \right) H_1 \right] & \text{if even} \end{cases} \tag{17}$$

Squaring both sides of (16) and taking expectation we obtain the MSE is given by

$$(t_g - \mu_y)^2 \cong \mu_y^2 \left[\varepsilon_{0(j)}^2 + \varepsilon_{1(j)}^2 H_1^2 + 2\varepsilon_{0(j)} \varepsilon_{1(j)} H_1 \right]$$

$$MSE(t_g) \cong \begin{cases} \frac{1}{n} \left[\sigma_{y\left(\frac{n+1}{2}\right)}^2 + \frac{\mu_y^2}{\mu_x^2} \sigma_{x\left(\frac{n+1}{2}\right)}^2 H_1^2 + 2 \frac{\mu_y}{\mu_x} \sigma_{xy\left(\frac{n+1}{2}\right)} H_1 \right] & \text{if odd} \\ \frac{1}{2n} \left[\left(\sigma_{y\left[\frac{n}{2}\right]}^2 + \sigma_{y\left[\frac{n+2}{2}\right]}^2 \right) + \frac{\mu_y^2}{\mu_x^2} \left(\sigma_{x\left(\frac{n}{2}\right)}^2 + \sigma_{x\left(\frac{n+2}{2}\right)}^2 \right) H_1^2 + 2 \frac{\mu_y}{\mu_x} \left(\sigma_{yx\left[\frac{n}{2}\right]} + \sigma_{yx\left[\frac{n+2}{2}\right]} \right) H_1 \right] & \text{if even} \end{cases} \tag{18}$$

Differentiating (18) with respect to H_1 and putting again we obtain minimum MSE as

$$MSE_{\min}(t_g) \cong \begin{cases} \frac{1}{n} \sigma_{y\left(\frac{n+1}{2}\right)}^2 \left(1 - \rho_{xy\left(\frac{n+1}{2}\right)}^2 \right), & n \text{ is odd} \\ \frac{1}{2n} \left(\sigma_{y\left[\frac{n}{2}\right]}^2 + \sigma_{y\left[\frac{n+2}{2}\right]}^2 \right) \left(1 - \left(\rho_{xy\left(\frac{n+1}{2}\right)}^2 + \rho_{xy\left[\frac{n+2}{2}\right]}^2 \right) \right), & n \text{ is even} \end{cases} \tag{19}$$

Note that this is also MSE of \bar{y}_{Reg1_MRSS} . (see Koyuncu [19] for details)

Srivastava [23] defined a wider class in simple random sampling. Secondly taking motivation from Srivastava [23] we have suggested wider class of estimators as

$$t_w = H(\bar{y}_{MRSS(j)}, u) \tag{20}$$

where $H(\bar{y}_{MRSS(j)}, u)$ is a function of $\bar{y}_{MRSS(j)}$ and u satisfy the following regularity conditions:

- a) The point $(\bar{y}_{MRSS(j)}, u)$ assumes the value in a bounded, closed convex subset R_2 of two-dimensional real space containing the point $(\bar{Y}, 1)$.

- b) The function $H(\bar{y}_{MRSS(j)}, u)$ is continuous and bounded in R_2 .
- c) $H(\bar{Y}, 1) = \bar{Y}$ and $H_0(\bar{Y}, 1) = 1$ which is first order partial derivative of H with respect to \bar{y}_{MRSS} .
- d) The first and second order partial derivatives of $H(\bar{y}_{MRSS}, u)$ exist and are continuous and bounded in R_2 .

Expanding $H(\bar{y}_{MRSS(j)}, u)$ about point of $(\bar{Y}, 1)$ in a second order Taylor series we have

$$t_w = H(\bar{y}_{MRSS(j)}, u) = H\left[\left(\bar{Y} + (\bar{y}_{MRSS(j)} - \mu_y)\right), (1 + (u - 1))\right] \tag{21}$$

$$t_w \cong \bar{y}_{MRSS(j)} + (u - 1)H_1 + (u - 1)^2 H_2 + (\bar{y}_{MRSS(j)} - \mu_y)(u - 1)H_3 + (\bar{y}_{MRSS(j)} - \mu_y)^2 H_4$$

where

$$H_1 = \frac{\partial H}{\partial u} \Big|_{\bar{y}_{MRSS} = \mu_y, u=1}, H_2 = \frac{1}{2} \frac{\partial^2 H}{\partial u^2} \Big|_{\bar{y}_{MRSS} = \mu_y, u=1}, H_3 = \frac{1}{2} \frac{\partial^2 H}{\partial \bar{y}_{MRSS} \partial u} \Big|_{\bar{y}_{MRSS} = \mu_y, u=1}, H_4 = \frac{1}{2} \frac{\partial^2 H}{\partial \bar{y}^2} \Big|_{\bar{y}_{MRSS} = \mu_y, u=1}$$

Extracting μ_y from both sides of (21) we obtain

$$t_w - \mu_y \cong \mu_y \varepsilon_{0(j)} + \varepsilon_{1(j)} H_1 + \varepsilon_{1(j)}^2 H_2 + \varepsilon_{0(j)} \varepsilon_{1(j)} H_3 + \varepsilon_{0(j)}^2 H_4 \tag{22}$$

The bias of t_w is, given by

$$B(t_w) \cong \begin{cases} \frac{1}{n} \left(\frac{1}{\mu_x^2} \sigma_{x\left(\frac{n+1}{2}\right)}^2 H_2 + \frac{1}{\mu_y \mu_x} \sigma_{xy\left(\frac{n+1}{2}\right)} H_3 + \frac{1}{\mu_y^2} \sigma_{y\left(\frac{n+1}{2}\right)}^2 H_4 \right) & \text{if odd} \\ \frac{1}{2n} \left(\frac{1}{\mu_x^2} \left(\sigma_{x\left(\frac{n}{2}\right)}^2 + \sigma_{x\left(\frac{n+2}{2}\right)}^2 \right) H_2 + \frac{1}{\mu_x \mu_y} \left(\sigma_{yx\left[\frac{n}{2}\right]} + \sigma_{yx\left[\frac{n+2}{2}\right]} \right) H_3 + \frac{1}{\mu_y^2} \left(\sigma_{y\left[\frac{n}{2}\right]}^2 + \sigma_{y\left[\frac{n+2}{2}\right]}^2 \right) H_4 \right) & \text{if even} \end{cases} \tag{23}$$

The MSE of t_w is, given by

$$MSE(t_w) \cong \begin{cases} \frac{1}{n} \left(\sigma_{y\left(\frac{n+1}{2}\right)}^2 + \frac{1}{\mu_x^2} \sigma_{x\left(\frac{n+1}{2}\right)}^2 H_1^2 + 2 \frac{1}{\mu_x} \sigma_{xy\left(\frac{n+1}{2}\right)} H_1 \right) & \text{if odd} \\ \frac{1}{2n} \left(\left(\sigma_{y\left[\frac{n}{2}\right]}^2 + \sigma_{y\left[\frac{n+2}{2}\right]}^2 \right) + \frac{1}{\mu_x^2} \left(\sigma_{x\left(\frac{n}{2}\right)}^2 + \sigma_{x\left(\frac{n+2}{2}\right)}^2 \right) H_1^2 + 2 \frac{1}{\mu_x} \left(\sigma_{yx\left[\frac{n}{2}\right]} + \sigma_{yx\left[\frac{n+2}{2}\right]} \right) H_1 \right) & \text{if even} \end{cases} \tag{24}$$

After differentiating with respect to H_1 and putting it (24) we obtain the $MSE_{\min}(t_w)$ is equal to $MSE_{\min}(t_g)$. This means that the usual linear regression estimator is special case of wider class of estimator.

Now we define a general class of estimators using multivariate auxiliary variable as

$$\bar{y}_s = \bar{y}_{MRSS(j)} H(u_1, u_2, \dots, u_p) = \bar{y}_{MRSS(j)} H(\underline{u}) \tag{25}$$

where $u_k = \frac{\bar{x}_{MRSS(j)k}}{\mu_{xk}}$ $k = 1, 2, \dots, p$ and $H(\underline{u})$ is a parametric function such that $H(\underline{\epsilon}) = 1$ for

$\underline{\epsilon} = (1, 1, \dots, 1)_{1 \times p}$, satisfying certain regularity condition such as the first and second order partial derivatives of H with respect to \underline{u} exist and are known. Expanding $H(\underline{u})$ around the $\underline{\epsilon}$ by using second order Taylor's series we have

$$H(\underline{u}) = H[\underline{\epsilon} + (\underline{u} - \underline{\epsilon})] = H(\underline{\epsilon}) + (\underline{u} - \underline{\epsilon}) \frac{\partial H}{\partial \underline{u}} \Big|_{\underline{u}=\underline{\epsilon}} + \frac{1}{2} (\underline{u} - \underline{\epsilon})' \frac{\partial^2 H}{\partial \underline{u}' \partial \underline{u}} \Big|_{\underline{u}=\underline{\epsilon}} (\underline{u} - \underline{\epsilon}) + \dots \tag{26}$$

Putting (26) in (25) we have

$$\bar{y}_s = \mu_y \left[1 + \varepsilon_{0(j)} + (\underline{u} - \underline{\epsilon}) \underline{H}_1 + \frac{1}{2} (\underline{u} - \underline{\epsilon})' \underline{H}_2 (\underline{u} - \underline{\epsilon}) + \varepsilon_{0(j)} (\underline{u} - \underline{\epsilon}) \underline{H}_1 + \dots \right] \tag{27}$$

where $\underline{H}_1 = \frac{\partial H}{\partial \underline{u}} \Big|_{\underline{u}=\underline{\epsilon}}$ and $\underline{H}_2 = \frac{\partial^2 H}{\partial \underline{u}' \partial \underline{u}} \Big|_{\underline{u}=\underline{\epsilon}}$ are the matrices consisting of first and second order partial derivatives of the function H with respect to \underline{u} and evaluated at $\underline{u} = \underline{\epsilon}$.

$$B(\bar{y}_s) = \mu_y E \left[\frac{1}{2} (\underline{u} - \underline{\epsilon})' \underline{H}_2 (\underline{u} - \underline{\epsilon}) + \varepsilon_{0(j)} (\underline{u} - \underline{\epsilon}) \underline{H}_1 + \dots \right] \tag{28}$$

$$B(\bar{y}_s) \cong \begin{cases} \frac{1}{n} \mu_y \left[\sum_{k=1}^p \frac{1}{\mu_{xk}^2} \sigma_{xk}^2 \left(\frac{n+1}{2} \right) H_{2k} + \sum_{k=1}^p \frac{1}{\mu_{xk} \mu_y} \sigma_{yxk} \left(\frac{n+1}{2} \right) H_{1k} \right] & \text{if odd} \\ \frac{1}{2n} \mu_y \left[\sum_{k=1}^p \frac{1}{\mu_{xk}^2} \left(\sigma_{xk}^2 \left(\frac{n}{2} \right) + \sigma_{xk}^2 \left(\frac{n+2}{2} \right) \right) H_{2k} + \sum_{k=1}^p \frac{1}{\mu_{xk} \mu_y} \left(\sigma_{yx} \left[\frac{n}{2} \right] + \sigma_{yx} \left[\frac{n+2}{2} \right] \right) H_{1k} \right] & \text{if even} \end{cases} \tag{29}$$

where H_{1j} and H_{2j} denote j^{th} diagonal element of H_2 and j^{th} component of H , respectively.

$$MSE(\bar{y}_s) \cong \mu_y^2 E \left[\varepsilon_0^2 + H_1' (\underline{u} - \underline{\epsilon}) (\underline{u} - \underline{\epsilon}) \underline{H}_1 + 2\varepsilon_0 (\underline{u} - \underline{\epsilon}) \underline{H}_1 \right] \tag{30}$$

$$MSE(\bar{y}_s) \cong \begin{cases} \frac{\mu_y^2}{n} \left[\frac{1}{\mu_y^2} \sigma_{y\left(\frac{n+1}{2}\right)}^2 + 2 \frac{1}{\mu_y \mu_{xt}} \sum_{t=1}^k \rho_{yxt} \sigma_y \sigma_{xt} H_{1t} + \frac{1}{\mu_{xt} \mu_{xs}} \sum_{t=1}^k \sum_{s=1}^k H_{1t} H_{1s} \rho_{xt} \rho_{xs} \sigma_{xt} \sigma_{xs} \right] & \text{if odd} \\ \frac{\mu_y^2}{2n} \left[\frac{1}{\mu_y^2} \left(\sigma_{y\left[\frac{n}{2}\right]}^2 + \sigma_{y\left[\frac{n+2}{2}\right]}^2 \right) + 2 \frac{1}{\mu_y \mu_{xt}} \sum_{t=1}^k \left(\sigma_{yxt\left[\frac{n}{2}\right]} + \sigma_{yxt\left[\frac{n+2}{2}\right]} \right) H_{1t} \right. \\ \left. + \frac{1}{\mu_{xt} \mu_{xs}} \sum_{t=1}^k \sum_{s=1}^k H_{1t} H_{1s} \rho_{xt} \rho_{xs} \left(\sigma_{xt\left[\frac{n}{2}\right]} + \sigma_{xt\left[\frac{n+2}{2}\right]} \right) \left(\sigma_{xs\left[\frac{n}{2}\right]} + \sigma_{xs\left[\frac{n+2}{2}\right]} \right) \right] & \text{if even} \end{cases} \quad (31)$$

On differentiating with respect to $\underline{H}_1 = (H_{11}, \dots, H_{1p})^t$ and equating to zero, we will obtain a set of p equations, as

$$\begin{bmatrix} \frac{1}{\mu_{x1}^2} \sigma_{x1\left(\frac{n+1}{2}\right)}^2 & \frac{\sigma_{x1x2\left(\frac{n+1}{2}\right)}}{\mu_{x1} \mu_{x2}} & \dots & \frac{\sigma_{x1xp\left(\frac{n+1}{2}\right)}}{\mu_{x1} \mu_{xp}} \\ \frac{\sigma_{x1x2\left(\frac{n+1}{2}\right)}}{\mu_{x1} \mu_{x2}} & \frac{1}{\mu_{x2}^2} \sigma_{x2\left(\frac{n+1}{2}\right)}^2 & \dots & \frac{\sigma_{x2xp\left(\frac{n+1}{2}\right)}}{\mu_{x2} \mu_{xp}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\sigma_{x1xp\left(\frac{n+1}{2}\right)}}{\mu_{x1} \mu_{xp}} & \frac{\sigma_{x2xp\left(\frac{n+1}{2}\right)}}{\mu_{x2} \mu_{xp}} & \dots & \frac{1}{\mu_{xp}^2} \sigma_{xp\left(\frac{n+1}{2}\right)}^2 \end{bmatrix} \begin{bmatrix} H_{11} \\ H_{12} \\ \vdots \\ H_{1p} \end{bmatrix} = \begin{bmatrix} -\mu_y \frac{\sigma_{yx1\left(\frac{n+1}{2}\right)}}{\mu_y \mu_{x1}} \\ -\mu_y \frac{\sigma_{yx2\left(\frac{n+1}{2}\right)}}{\mu_y \mu_{x2}} \\ \vdots \\ -\mu_y \frac{\sigma_{yxp\left(\frac{n+1}{2}\right)}}{\mu_y \mu_{xp}} \end{bmatrix}$$

or $\underline{A} \underline{H}_1 = \underline{C}$. This equation can easily be solved for unknown parameters as $\underline{H}_1 = \underline{A}^{-1} \underline{C}$

$$MSE(\bar{y}_s)_{\min} \cong \begin{cases} \frac{1}{n \mu_y^2} \sigma_{y\left(\frac{n+1}{2}\right)}^2 [1 - R_{y.x1x2\dots xp}^2] & \text{if odd} \\ \frac{1}{2n \mu_y^2} \left(\sigma_{y\left[\frac{n}{2}\right]}^2 + \sigma_{y\left[\frac{n+2}{2}\right]}^2 \right) [1 - R_{y.x1x2\dots xp}^2] & \text{if even} \end{cases} \quad (32)$$

Note that this MSE is same with MSE of multivariate regression estimator.

Corollary: A wider class of estimators for estimating population mean μ_y using p auxiliary variables $\mu_{x1}, \dots, \mu_{xp}$ can easily be defined as

$$\bar{y}_w = H(\bar{y}_{MRSS(j)}, \underline{u})$$

where $H(\cdot, \cdot)$ is a parametric function such that $H(\bar{Y}, \underline{\Xi}) = 1$, satisfying certain regularity conditions. It is easy to show that a wider class of estimators \bar{y}_w has the same asymptotic mean square error as that of the general class of ratio type estimators of population mean defined at $\bar{y}_{MRSS(j)} H(\underline{u})$.

5. Simulation Study

In this section, we conducted a simulation study to investigate the properties of proposed estimators. In the simulation study, we track following steps:

1. Generate correlated four finite populations of size $N = 10000$ using bivariate normal distribution using “mvrnorm” function in R programme. In the simulation, we considered $\mu_x = 2, \mu_z = 3, \mu_y = 4$. To see the efficiency of correlation we consider different correlations. To use the ratio type estimators we have taken into account only positive correlations. We have generated populations with different covariance matrices as given below:

Population I

$$\sigma^2 = \begin{bmatrix} 18 & 3 & 2.9 \\ 3 & 3 & 1.1 \\ 2.9 & 1.1 & 4 \end{bmatrix} \quad \rho_{yx} = 0.4082 \quad \rho_{yz} = 0.3418 \quad \rho_{xz} = 0.3175$$

Population II

$$\sigma^2 = \begin{bmatrix} 14 & 3 & 2.9 \\ 3 & 2 & 1.1 \\ 2.9 & 1.1 & 2 \end{bmatrix} \quad \rho_{yx} = 0.5669 \quad \rho_{yz} = 0.5480 \quad \rho_{xz} = 0.55$$

Population III

$$\sigma^2 = \begin{bmatrix} 10 & 3 & 2.9 \\ 3 & 2 & 1.1 \\ 2.9 & 1.1 & 2 \end{bmatrix} \quad \rho_{yx} = 0.6708 \quad \rho_{yz} = 0.6485 \quad \rho_{xz} = 0.55$$

Population IV

$$\sigma^2 = \begin{bmatrix} 6 & 3 & 2.9 \\ 3 & 2 & 1.1 \\ 2.9 & 1.1 & 2 \end{bmatrix} \quad \rho_{yx} = 0.8660 \quad \rho_{yz} = 0.8372 \quad \rho_{xz} = 0.55$$

2. Select samples size from bivariate normal distributions using SRS and MRSS for $n = 5, 6, 7, 8$ on the basis of 60.000 replications respectively.
3. Compute mean estimation using $\bar{y}_{R_SRS}, \bar{y}_{R1_SRS}, \bar{y}_{R2_SRS}, \bar{y}_{R3_SRS}, \bar{y}_{Reg1_SRS}, \bar{y}_{Reg2_SRS}, \bar{y}_{R_MRSS}, \bar{y}_{R1_MRSS}, \bar{y}_{R2_MRSS}, \bar{y}_{R3_MRSS}, \bar{y}_{Reg1_MRSS}$ and \bar{y}_{Reg2_MRSS} from samples.
4. Computed MSEs and percent relative efficiencies (PREs) of estimators with respect to \bar{y}_{R_SRS} for $n = 5, 6, 7, 8$ on the basis of 60.000 replications and displayed in Table 1-Table 4. We used following formula to calculate PRE values:

$$PRE_i = \frac{MSE(\bar{y}_{R_SRS})}{MSE(\bar{y}_i)} * 100$$

where i=R_SRS, R1_SRS, R2_SRS, R3_SRS, Reg1_SRS, Reg2_SRS, R_MRSS, R1_MRSS, R2_MRSS, R3_MRSS, Reg1_MRSS, Reg2_MRRS

Table 1: MSE and Efficiency of Estimators for Population I

Estimator	n = 5		n = 6		n = 7		n = 8	
	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
\bar{y}_{R_SRS}	13164.0042	100.00	948.5593	100.00	753.8487	100.00	52.6932	100.00
\bar{y}_{R1_SRS}	8.6828	151609.71	4.1258	22991.10	2.9104	25901.97	2.4331	2165.68
\bar{y}_{R2_SRS}	3.3443	393622.86	2.7754	34177.35	2.3691	31819.90	2.0342	2590.32
\bar{y}_{R3_SRS}	26038.5650	50.56	2531.9695	37.46	2295.0400	32.85	172.1833	30.60
\bar{y}_{Reg1_SRS}	3.0754	428046.87	2.5875	36658.93	2.2239	33897.70	1.9212	2742.74
\bar{y}_{Reg2_SRS}	2.9395	447824.38	2.4802	38245.61	2.1325	35350.10	1.8431	2858.95
\bar{y}_{R_MRSS}	3.8511	341828.74	1.5494	61222.79	2.4876	30304.07	1.0921	4824.86
\bar{y}_{R1_MRSS}	3.3535	392550.28	1.3704	69216.08	2.3136	32582.96	1.0117	5208.15
\bar{y}_{R2_MRSS}	3.1856	413229.24	1.3215	71779.46	2.2542	33442.67	0.9921	5311.12
\bar{y}_{R3_MRSS}	115.7138	11376.35	2.1826	43460.93	3.9730	18974.23	1.3606	3872.78
\bar{y}_{Reg1_MRSS}	3.1153	422555.38	1.2978	73089.57	2.2225	33918.36	0.9807	5372.81
\bar{y}_{Reg2_MRSS}	2.9281	449576.12	1.2190	77812.19	2.0765	36304.05	0.9137	5766.88

Table 2: MSE and Efficiency of Estimators for Population II

Estimator	n = 5		n = 6		n = 7		n = 8	
	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
\bar{y}_{R_SRS}	211.6173	100.00	6.7393	100.00	7.7910	100.00	216.1899	100.00
\bar{y}_{R1_SRS}	2.1825	9696.16	1.7635	382.16	1.4788	526.85	1.2896	16764.36
\bar{y}_{R2_SRS}	2.1703	9750.69	1.7939	375.68	1.5215	512.07	1.3375	16163.55
\bar{y}_{R3_SRS}	445.1171	47.54	16.1187	41.81	20.0734	38.81	519.0859	41.65
\bar{y}_{Reg1_SRS}	1.8579	11390.24	1.5479	435.40	1.3277	586.80	1.1668	18528.53
\bar{y}_{Reg2_SRS}	1.9815	10679.46	1.6448	409.72	1.4072	553.63	1.2367	17480.73
\bar{y}_{R_MRSS}	2.0655	10245.43	0.8371	805.03	1.4182	549.37	0.6235	34672.35
\bar{y}_{R1_MRSS}	1.9183	11031.28	0.7906	852.48	1.3612	572.35	0.5986	36118.91
\bar{y}_{R2_MRSS}	1.9222	11009.15	0.8097	832.31	1.3687	569.23	0.6098	35455.12
\bar{y}_{R3_MRSS}	2.4221	8736.95	0.9280	726.25	1.5159	513.94	0.6480	33360.22
\bar{y}_{Reg1_MRSS}	1.8496	11440.94	0.7730	871.86	1.3374	582.54	0.5914	36555.78
\bar{y}_{Reg2_MRSS}	1.7780	11901.87	0.7588	888.11	1.2653	615.75	0.5643	38307.82

Table 3: MSE and Efficiency of Estimators for Population III

Estimator	n = 5		n = 6		n = 7		n = 8	
	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
\bar{y}_{R_SRS}	27.3732	100.00	24.3081	100.00	1.7696	100.00	1.2123	100.00
\bar{y}_{R1_SRS}	1.2764	2144.49	1.0329	2353.35	0.8629	205.07	0.7527	161.07
\bar{y}_{R2_SRS}	1.3489	2029.24	1.1142	2181.74	0.9431	187.63	0.8285	146.33
\bar{y}_{R3_SRS}	145.0732	18.87	52.8522	45.99	3.1282	56.57	2.0634	58.75
\bar{y}_{Reg1_SRS}	1.0813	2531.59	0.9002	2700.33	0.7720	229.23	0.6776	178.91
\bar{y}_{Reg2_SRS}	1.1961	2288.51	0.9925	2449.24	0.8450	209.42	0.7432	163.12
\bar{y}_{R_MRSS}	1.2418	2204.27	0.4960	4900.34	0.8261	214.21	0.3642	332.85
\bar{y}_{R1_MRSS}	1.1382	2404.96	0.4618	5263.39	0.7906	223.84	0.3462	350.19
\bar{y}_{R2_MRSS}	1.1614	2356.98	0.4856	5005.79	0.8082	218.97	0.3604	336.36
\bar{y}_{R3_MRSS}	1.4280	1916.88	0.5558	4373.20	0.8517	207.78	0.3737	324.38
\bar{y}_{Reg1_MRSS}	1.1006	2487.15	0.4517	5381.40	0.7789	227.20	0.3426	353.82
\bar{y}_{Reg2_MRSS}	1.0248	2671.10	0.4398	5527.28	0.7112	248.83	0.3194	379.52

Table 4: MSE and Efficiency of Estimators for Population IV

Estimator	n = 5		n = 6		n = 7		n = 8	
	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
\bar{y}_{R_SRS}	30.9339	100.00	10.1850	100.00	0.6340	100.00	0.4334	100.00
\bar{y}_{R1_SRS}	0.3656	8461.66	0.2963	3437.76	0.2449	258.86	0.2137	202.82
\bar{y}_{R2_SRS}	0.5267	5872.75	0.4340	2346.62	0.3649	173.73	0.3202	135.35
\bar{y}_{R3_SRS}	73.7932	41.92	27.3506	37.24	1.3909	45.58	0.8715	49.73
\bar{y}_{Reg1_SRS}	0.2965	10433.07	0.2464	4133.89	0.2111	300.29	0.1854	233.77
\bar{y}_{Reg2_SRS}	0.4061	7617.18	0.3399	2996.27	0.2880	220.19	0.2556	169.59
\bar{y}_{R_MRSS}	0.3590	8617.02	0.1481	6879.10	0.2347	270.11	0.1047	413.83
\bar{y}_{R1_MRSS}	0.3132	9876.42	0.1272	8008.96	0.2161	293.41	0.0952	455.33
\bar{y}_{R2_MRSS}	0.3601	8590.04	0.1577	6459.30	0.2407	263.38	0.1137	381.20
\bar{y}_{R3_MRSS}	0.4688	6598.32	0.1785	5706.77	0.2098	302.23	0.1094	396.31
\bar{y}_{Reg1_MRSS}	0.2964	10435.62	0.1217	8370.56	0.2097	302.28	0.0922	469.95
\bar{y}_{Reg2_MRSS}	0.2325	13302.55	0.1105	9213.11	0.1534	413.41	0.0766	565.93

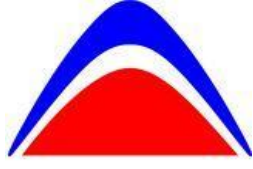
From the simulation results given in Tables 1-4, we can conclude that:

1. When we compare same type estimators in SRS and MRSS sampling designs we can say that MRSS estimators give more efficient results than SRS estimators.
2. Ratio type estimator using two auxiliary variables \bar{y}_{R3_SRS} is the worst estimator in SRS design for all populations. And \bar{y}_{R3_MRSS} is the worst estimator in MRSS except the sample size $n = 7$ for Population IV.

3. From the Table 1-Table 4, we can say that suggested class of estimator performs better than Al-Omari (2012) estimator. We can conclude that regression type estimator gives always more efficient results than Al-Omari (2012) estimator.
4. If we use any function of u in our estimator we can not reduced the MSE than MSE of regression estimator. The usual ratio, product and power estimators are special case of class of estimator.
5. To use ratio type estimators in SRS or MRSS, we assume that there are positive correlations between study and auxiliary variables. Otherwise to get efficiency we need to use product type estimators.

References:

- [1] A. I., Al-Omari, 2012, Ratio estimation of the population mean using auxiliary information in simple random sampling and median ranked set sampling, *Statistics and Probability Letters*, 82, 1883-1890.
- [2] M.F., Al-Saleh, A.I., Al-Omari, 2002, Multistage ranked set sampling, *Journal of Statistical Planning and Inference*, 102, 273-286.
- [3] A.A., Jemain, A.I., Al-Omari, 2006, Multistage median ranked set samples for estimating the population mean, *Pakistan Journal of Statistics*, 22, 195-207.
- [4] A.A., Jemain, A.I., Al-Omari, K., Ibrahim, 2007, Multistage extreme ranked set sampling for estimating the population mean, *Journal of Statistical Theory and Application*, 6, 4, 456-471.
- [5] O., Ozturk, M., Jafari Jozani, 2014, Inclusion probabilities in partially rank ordered set sampling, *Computational Statistics and Data Analysis*, 69, 122-132.
- [6] S., Amiri, R., Modarres, D.S., Bhoj, 2015, Ranked set sampling with random subsamples, *Journal of Statistical Computation and Simulation*, 85, 5, 935-946.
- [7] A.I., Al-Omari, A.D., Al-Nasser, 2012, On the population median estimation using robust extreme ranked set sampling, *Monte Carlo Methods and Applications*, 18, 2, 109-118.
- [8] A.I., Al-Omari, M.Z., Raqab, 2013, Estimation of the population mean and median using truncation-based ranked set samples, *Journal of Statistical Computation and Simulation*, 83, 8, 1453-1471.
- [9] A., Haq, J., Brown, E., Moltchanova, A.I., Al-Omari, 2013, Partial ranked set sampling design, *Environmetrics*, 24, 3, 201-207.
- [10] A., Haq, J., Brown, E., Moltchanova, A.I., Al-Omari, 2014, Mixed ranked set sampling design, *Journal of Applied Statistics*, 41, 10, 2141-2156.
- [11] D.S., Bhoj, D., Kushary, 2014, Ranked set sampling with unequal samples and unequal replications, *Pakistan Journal of Statistics*, 30, 3, 351-362.
- [12] D.S., Bhoj, D., Kushary, 2016, Ranked set sampling with unequal samples for skew distributions, *Journal of Statistical Computation and Simulation*, 86, 4, 676-681.
- [13] M.H., Samuh, A., Qtait, 2015, Estimation for the parameters of the exponentiated exponential distribution using a median ranked set sampling, *Journal of Modern Applied Statistical Methods*, 14, 1, 215-237.
- [14] A., Omar, K., Ibrahim, 2013, Estimation of the shape and scale parameters of the pareto distribution using extreme ranked set sampling, *Pakistan Journal of Statistics*, 29, 1, 33-47.
- [15] S., Gupta, J., Shabbir, 2015, Estimation of finite population mean in stratified random sampling with two auxiliary variables under double sampling design, *Communications in Statistics - Theory and Methods*, 44, 13, 2798-2808.
- [16] J., Shabbir, S., Gupta, 2015, A note on generalized exponential type estimator for population variance in survey sampling, *Revista Colombiana de Estadística*, 38, 2, 385-397.
- [17] J., Shabbir, A., Haq, S., Gupta, 2014, A new difference-cum-exponential type estimator of finite population mean in simple random sampling, *Revista Colombiana de Estadística*, 37, 1, 197-209.
- [18] N., Koyuncu, 2015, Ratio estimation of the population mean in extreme ranked set and double robust extreme ranked set sampling, *International Journal of Agricultural and Statistical Sciences*, 11, 1, 21-28.
- [19] N., Koyuncu, 2016, New difference-cum-ratio and exponential type estimators in median ranked set sampling, *Hacettepe Journal of Mathematics and Statistics*, 45, 1, 207-225.
- [20] S., Singh, 2003, *Advanced Sampling Theory With Applications: How Michael "Selected" Amy*, Springer Science & Business Media.
- [21] S.K., Srivastava, S.K., Jhaji, 1980, A class of estimators using auxiliary information for estimating finite population variance, *Sankhya*, C, 42, 87-96.
- [22] N., Koyuncu, C., Kadilar, 2010, On improvement in estimating population mean in stratified random sampling, *Journal of Applied Statistics*, 37, 6, 999-1013.
- [23] S.K., Srivastava, 1980, A class of estimator using auxiliary information in sample surveys, *Canadian Journal of Statistics*, 8, 253-254.



Aktüerya Derneği

İstatistikçiler Dergisi: İstatistik & Aktüerya

Journal of Statisticians: Statistics and Actuarial Sciences

IDIA 12, 2019, 2, 72-89

Geliş/Received:01.07.2019, Kabul/Accepted: 01.12.2019

www.istatistikciler.org

Araştırma Makalesi / Research Article

Süreç ortalamasının izlenmesi için sıra istatistiklerine dayalı kalite kontrol kartları

Nejla Turhan


Ankara Üniversitesi

Fen Fakültesi,

İstatistik Bölümü

06100 Beşevler, Ankara, Türkiye

nturhan@ankara.edu.tr

 0000-0003-1557-3314

Sevgi Yurt Öncel


Kırıkkale Üniversitesi

Fen-Edebiyat Fakültesi

İstatistik Bölümü,

71450, Yenişehir, Yahşihan, KIRIKKALE

syoncel@kku.edu.tr

 0000-0002-0990-292X

Öz

Bu çalışmada, herhangi bir dağılım varsayımı olmadan sıra istatistiklerine dayalı ortalamaya ilişkin kalite kontrol kartı limitlerinin belirlenmesi konusu ele alınmıştır. Dağılımdan bağımsız kontrol kartları, çarpık ya da ağır kuyruklu dağılımlar için dayanıklı bir yöntemdir. Ayrıca, çizilecek kontrol kartının dağılımının varyansını bilmeye ya da tahmin etmeye gerek yoktur. Farklı örnek hacimleri için dağılımdan bağımsız kontrol limitleri, yanlış alarm oranı ve ortalama çalışma uzunluğu hesaplamaları tablo halinde sunulmuştur. Shewhart kontrol kartı ve medyan kontrol kartının çizimi hakkında bir uygulama yapılmıştır.

Anahtar Sözcükler: Dağılımdan bağımsız kalite kontrol, kontrol kartı, medyan kontrol kartı, sıra istatistikleri, ön istatistikler

Abstract

Quality control cards based on order statistics to monitoring process average

In this study, determination of quality control card limits based on order statistics without any distribution assumption is discussed. Distribution-free control cards are a robust method for skewed or heavily tailed distributions. In addition, there is no need to estimate the variance of the distribution of the control card to be drawn. Distribution independent control limits for different sample volumes, false alarm rate, and average run length calculations are presented in a table. An application was made about drawing of Shewhart control card and median control card graph.

Keywords: Distribution-free quality control, control charts, median control charts, order statistics, precedence statistics.

1. Giriş

Kontrol kartlarını kullanmanın temel amacı, süreç dağılımında meydana gelebilecek değişiklikleri hızlıca belirleyebilmektir. Doğal nedenlerden kaynaklanan değişimler göz ardı edilerek özel sebeplerin yol açabileceği değişim yakalanmak istenilmektedir. Bir imalat sürecinde

ham maddeden, üretimde kullanılan cihazlardan veya görevli kişilerden kaynaklanan değişimler özel sebepler olarak adlandırılır.

Kontrol kartlarının çoğu, dağılım varsayımına dayanan prosedürlerdir. Süreç çıktısının belirli bir olasılık dağılımına (genellikle normal dağılıma) sahip olduğu varsayılır.

Kalite kontrol kartları, hipotez testleri ile yakından ilişkilidir. Kitle ortalamasına ilişkin $H_0 : \mu = \mu_0$ iddiasına karşılık $H_1 : \mu = \mu_1$ ($\mu_1 = \mu_0 + \delta\sigma$) iddiasının doğru olup olmadığının α anlam düzeyinde araştırıldığı hipotez testi sonucunda H_0 hipotezinin reddedilememesi sürecin kontrol altında olmasına denktir. \bar{X} -kalite kontrol kartlarının amacı, süreç ortalamasındaki kaymaların hızlı bir şekilde tespit edilmesidir. Kontrol kartının limitleri $\mu_{\bar{x}} \mp k\sigma_{\bar{x}}$ dir. Süreç kontrol dışı iken örneklem ortalaması \bar{x} 'nin kontrol limitlerinin dışına çıkması olasılığı

$$P = 1 - \beta = \int_{-\infty}^{\mu_0 - k\sigma/\sqrt{n}} f(\bar{x}|\mu_1) d\bar{x} + \int_{\mu_0 + k\sigma/\sqrt{n}}^{\infty} f(\bar{x}|\mu_1) d\bar{x}$$

dir. Burada P , doğru alarm verme olasılığı, β ise II. Tip hata yapma olasılığıdır (tüketici riskidir) ve $\beta = P(\text{Süreç kontrol altında} | \text{Süreç kontrol dışında iken})$ biçiminde ifade edilir. Burada yer alan μ_0 ve μ_1 parametreleri sırasıyla süreç kontrol altında ve süreç kontrol dışında iken X rasgele değişkeninin kitle ortalamasını göstermektedir. X rasgele değişkeninin dağılımı normal dağılım olduğunda

$$P = 1 - \Phi(k - \sqrt{n}\delta) + \Phi(-k + \sqrt{n}\delta)$$

dir. Burada Φ , standart normal dağılım fonksiyonu ve $\delta = (\mu_1 - \mu_0) / \sigma$ dir. Yanlış alarm verme olasılığı yani I. tip hata olasılığı (üretici riski) ise $\alpha = P(\text{Süreç kontrol dışı} | \text{Süreç kontrol altında iken}) = 2\Phi(-k)$ dir [1]. Kontrol kartında işaretlenen örnek ortalamasının kontrol limitlerinin dışına çıkmadan önce çizilmesi beklenen noktaların sayısı, ortalama çalışma uzunluğu (average run length-ARL) olarak adlandırılır ve kontrol kartının performansını değerlendirmek için kullanılır. Süreç kontrol altındayken $ARL = \frac{1}{\alpha}$, süreç

kontrol dışında iken $ARL = \frac{1}{1 - \beta}$ dir. Örneklemin dağılımı Normal dağılım iken ortalama için

Shewhart Kontrol kartının ortalama çalışma uzunluğu $\alpha = 0.0027$ için 370'dir. Yani süreç kontrol altındayken ortalama 370 noktada bir kez kontrol karıt, süreç kontrol dışında sinyali verecektir [2].

Birçok durumda ilgilenilen kalite değişkeninin gösterdiği dağılım normal dağılım olmayabilir, çarpık ya da ağır kuyruklu dağılımlar olabilir. Bu durumda kontrol kartlarının çiziminde parametrik olmayan (dağılımdan bağımsız) yöntemler geliştirilmiştir. Özellikle küçük örneklem hacmi ile çalışıldığında parametrik olmayan yöntemler tercih edilmektedir. Ancak parametrik olmayan kartların algılanan dezavantajı, hesaplamaların zorluğundan dolayı özel tablolara ihtiyaç duymasıdır. Literatürde, örneğin yerini kontrol etmek için farklı parametrik olmayan kontrol kartı tabloları bulunmaktadır. İşaret testine dayanan kontrol kartları, Hodges-Lehmann tahmincisine dayanan kontrol kartları ve Mann-Whitney istatistiklerine dayanan kontrol kartları konumdaki kaymayı tespit etmek için kullanılan yöntemlerdir [3].

Parametrik olmayan kalite kontrol hakkında ilk çalışma Bakir ve Reynolds [4] tarafından yapılmıştır ve Wilcoxon işaretli sıra istatistiklerine dayalı olarak bir prosedür geliştirmişlerdir. Daha sonraki yıllarda Janacek ve Meikle [5], Chakraborti ve ark.[6,7], Balakrishnan ve ark. [8,9], Triantafyllou [10,11]'nin parametrik olmayan kalite kontrol kartları hakkında önemli çalışmaları bulunmaktadır.

Bakir [12], işaret istatistiklerine dayalı bir Shewhart tipi kontrol kartları hazırlamıştır. Chakraborti ve Eryılmaz [13], daha küçük yanlış alarm oranları ve daha büyük kontrol ortalama çalışma uzunluklarına sahip Wilcoxon istatistiğine dayanan alternatif bir yöntem geliştirmişlerdir. Chakraborti, van der Laan ve van de Wiel [7], kontrol içi (referans) örneklemeden hesaplanan kontrol limitleriyle sıra istatistiklerine dayanan, parametrik olmayan kontrol kartları tanıttı.

Balakrishnan, Triantafyllou ve Koutras [14], test örnekleminin bir sıra istatistiği (ortanca gibi) olmasının yanı sıra test örnekleminin kontrol limitleri arasına düşen gözlem sayısını da modele dahil etmiştir. Ek olarak, Balakrishnan, Triantafyllou ve Koutras [15] izlenen bir sürecin olası kaymalarını saptamak için koşu (run) ve Wilcoxon sıra (rank) toplamı istatistiklerinden yararlanırken, Mukherjee ve Chakraborti [16], Shewhart-Lepage grafiği olarak adlandırılan Lepage istatistiğine dayalı olarak konum ve ölçeğin ortak izlenmesi için bir grafik önermiştir. Benzer bir çerçevede, Chowdhury ve ark. [17], Shewhart-Cucconi şeması olarak adlandırılan Cucconi istatistiklerine dayanan, dağılım varsayımı gerektirmeyen Shewhart tipi kart hazırladı.

Dağılımdan bağımsız kontrol kartları ile ilgili son gelişmeler, Chakraborti ve ark. [6,7], Chakraborti ve Graham [18], Chakraborti [19], Malela-Majika ve ark. [20] tarafından yapılmıştır. Sürekli değişkenler için dağılımdan bağımsız kontrol kartlarına genel bakış için, ilgilenen okuyucular bu çalışmalarını okuması önerilir.

Triantafyllou [21] , test örnekleminin iki sıra istatistiğinin aynı anda kontrol limitleri arasında olmasını dikkate alan yaklaşımı ile dağılımdan bağımsız Shewhart tipi kontrol kartlarının kurulumunu ayrıntılı olarak açıklamış, yanlış alarm oranı ve çalışma karakteristiği işlevi için formüller elde etmiştir.

Ayrıca çarpık dağılımlar için \bar{X} ve R kontrol kartı için kontrol limitlerinin oluşturulması için Bai ve Choi [22] ağırlıklı varyans yöntemini, Chan ve Cui [23] çarpıklık düzeltmesi yöntemini, Tadikamalla ve Popescu [24] basıklık düzeltmesi yöntemin önerilmiştir. Chan ve Cui [23]'nin yöntemine dayalı olarak Kan ve Yazıcı [25], Normal dağılımına sahip olmayan veriler asimetrik kontrol limitlerini oluşturmak için yeni bir teknik sunmuşlardır.

Karagöz ve Hamurkaroğlu [26] çarpık dağılımlar için \bar{X} ve R kontrol kartlarının limitlerini, klasik, ağırlıklı varyans, ağırlıklı standart sapma ve çarpıklık düzeltmesi yöntemlerini kullanarak elde etmişlerdir. Bu yöntemlerin Monte Carlo simülasyonu kullanılarak karşılaştırılması sonucunda Weibull, Gamma ve Lognormal dağılımları için çarpıklık düzeltmesi yönteminin I. tip hata olasılığının daha düşük olduğu ifade etmişlerdir. Yıldız ve Vahaplar [27], Balakrishnan, Triantafyllou ve Koutras [14] çalışmasında verilen teoriyi kullanarak bayan gömlekleri için dokuma yapan bir fabrikaya ait verilerin analizi ile sürecin istatistiksel olarak kontrol altında olup olmadığını incelemişlerdir.

Kontrol dışı bir sinyal alana kadar kartta çizilen örneklerin sayısı bir rasgele değişkendir ve çalışma uzunluğu (run length) olarak adlandırılır. Bir rastgele değişken olan bu çalışma uzunluğunun beklenen değeri, kontrol kartının performansını ölçmek için çok sık olarak

kullanılan Ortalama Çalışma Uzunluğu (average run length-ARL) 'dur. Süreç kontrol altındayken ARL'nin büyük (uzun), süreç kontrol dışındayken ise ARL'nin küçük (kısa) olması ve varsa değişimin hızlı bir şekilde tespit edilmesi arzulanır. Süreçte incelenen değişkenlerin dağılımı hangi dağılım olursa olsun, süreç kontrol altındayken çalışma uzunluğunun dağılımı aynı kalıyorsa o zaman kontrol kartı dağılımdan bağımsızdır denir.

Bu çalışmada süreç ortalamasındaki değişimi tespit edebilmek için herhangi bir dağılım varsayımı yapılmadan sıra istatistikleri teorisine dayalı ön istatistikleri kullanılarak medyan kontrol kartları incelenmiştir. Shewhart kontrol kartlarındaki aritmetik ortalama yerine medyan istatistiği kullanılmıştır. Dağılımdan bağımsız kontrol kartları, çarpık ya da ağır kuyruklu dağılımlar için dayanıklı (robust) bir yöntemdir. Test örnekleminin medyanının, referans örnekleminden belirlenen kontrol limitleri arasında olup olmadığını araştırmak için kullanılacak formüller Chakraborti ve ark [7]'nin çalışmasından verilmiştir. Uygulama bölümü iki kısımdan oluşmaktadır. Önce farklı örneklem hacimleri için dağılımdan bağımsız kontrol limitleri bulunmuştur.

Ayrıca, bu kontrol limitlerinin iyilik ölçütü olarak yanlış alarm oranı (False Alarm Rate-FAR) ve ortalama çalışma uzunluğu (ARL) hesaplamaları yapılarak 3 yeni tablo hazırlanmıştır. Daha sonra da tabloların nasıl kullanılacağını göstermek için literatürden alınan veriler kullanılarak parametrik olmayan kontrol kartların çizimi ayrıntılı bir şekilde açıklanmıştır. Shewhart kontrol kartı ile medyan kontrol kartının karşılaştırması bir uygulama üzerinde karşılaştırmalı olarak verilmiştir.

2. Ön (Precedence) İstatistikler

Bir imalat sürecinde ortaya konulan ürünün kalitesi hakkında incelenebilecek tüm özellikleri bir rasgele değişkendir. Bir üründen ölçme yoluyla elde edilen uzunluk, ağırlık, dayanıklılık, en, boy, çap, vb. karakteristiklerini kaydedip, gerekli kalite kontrollerini yapmak işletmelerin verimliliğini artırmasını sağlar. Ürünün belli bir karakteristiğinin ölçülen değerleri örneklem kümesini oluşturur.

Bir örnekleme de ki ön istatistiği (precedence statistics) $S_n = \#\{j \leq n : Y_j \leq X\}$, $j = 1, 2, \dots, n$ biçiminde tanımlanır. Burada X rasgele değişkeni, ilk örneklemden alınan bir rasgele bariyer noktasını, Y_j ise sonradan elde edilen örnekleme göstermektedir. S_n ön istatistiği, X 'den küçük olan Y 'lerin sayısını göstermektedir. Ön istatistikler üzerine kurulan testlere de ön testler denilmektedir. Ön testler iki bağımsız örneklemden belirlenmiş iki sıra istatistiğine dayanan oldukça kullanışlı testlerdir. H_0 hipotezinin doğru olması, yani $H_0 : F(x) = G(x)$ durumunda ön olasılıklar rahatlıkla hesaplanabilmektedir. Bu istatistik birçok alanda kullanıldığı gibi istatistiksel kalite kontrolünde de kullanılabilir.

Farz edelim ki X_1, X_2, \dots, X_m referans örnekleme $F(x, \mu_0)$ dağılım fonksiyonuna sahip olsun. Sürecin kontrol altında olup olmadığını test etmek için alınan Y_1, Y_2, \dots, Y_n test örnekleme $G(x, \mu_1)$ dağılımına sahip olsun. Referans örneklemin sıra istatistikleri $X_{1:m}, X_{2:m}, \dots, X_{n:m}$, test örnekleminin sıra istatistikleri ise $Y_{1:m}, Y_{2:m}, \dots, Y_{n:m}$ olsun. \mathfrak{S}_c ile de tüm sürekli dağılım fonksiyonları sınıfını gösterelim. İstatistiksel kalite kontrolün temel problemi F ve G dağılımı arasında bir farklılık olup olmadığını yani, $F, G \in \mathfrak{S}_c$ olmak üzere $H_0 : F(x) = G(x)$ hipotezine

karşı $H_1 : F(x) \neq G(x)$ hipotezini test etmektir. G dağılımının konum parametresinin $\mu_1 = \mu_0 + \delta\sigma$ biçiminde bir kayma olması durumunda sürecin kontrol dışına çıkıp çıkmadığı araştırılmak istenebilir. Bu çalışmada normal dağılım varsayımının sağlanamadığı durumlarda Shewhart ortalama kontrol kartları yerine parametrik olmayan kontrol kartları üzerinde durulmuştur. Ön istatistikleri kullanılarak kontrol limitleri oluşturulmuştur.

Referans örneklemeden kontrol kartının alt ve üst kontrol limitleri $AKL = X_{am}$, $ÜKL = X_{bm}$, $1 \leq a < b \leq m$ olarak belirlenir. $Y_{j:n}$ test örneklemeden alınmış j . sıra istatistiğini gösterir. $Y_{j:n}$, AKL ve $ÜKL$ arasına düşerse süreç kontrol altındadır denir. Eğer $Y_{j:n}$, bu aralığın dışında kalırsa süreç kontrol dışındadır denir. $Y_{j:n}$ 'den küçük olan X gözlemlerinin sayısını W_j istatistiği ile gösterilsin. W_j 'ye ön istatistik (precedence statistics) denir. Süreç kontrol altındayken (K indisi ile gösterilsin) W_j istatistiğinin dağılımı

$$P_K(W_j = w) = \frac{\binom{j+w-1}{w} \binom{m+n-j-w}{m-w}}{\binom{m+n}{m}}, \quad w = 0, 1, 2, \dots, m \quad (1)$$

biçimindedir [7]. W_j istatistiğinin dağılımı, sadece m, n ve j 'ye bağlıdır, F ve G dağılımlarına bağlı değildir. Dolayısıyla W_j , dağılımdan bağımsız istatistiktir. Önerilen kartın "sinyal vermemesi" olasılığı, ön istatistiği W_j cinsinden ifade edilebilir. Test örnekleminin j . sıra istatistiği olan $Y_{j:n}$ 'nin X_{am} ve X_{bm} arasına düşmesi olayı ile W_j istatistiğinin a ve $b-1$ arasında olması aynı olaylardır. Böylece

$$p = p(m, n, j, a, b, F, G) = P(X_{am} \leq Y_{j:n} \leq X_{bm}) = P(a \leq W_j \leq b-1) \quad (2)$$

dır. Böylece süreç kontrol altındayken

$$p_0 = P_K(a \leq W_j \leq b-1) = p(m, n, j, a, b, F, F)$$

olacaktır. (1) ve (2) eşitlikleri kullanılarak a ve b sabitleri ($1 \leq a < b \leq m$)

$$\sum_{w=a}^{b-1} \frac{\binom{j+w-1}{w} \binom{m+n-j-w}{m-w}}{\binom{m+n}{m}} \geq p_0 \quad (3)$$

eşitliğinden bulunabilir [7]. Kontrol kartının $AKL = X_{am}$ ve $ÜKL = X_{bm}$, $1 \leq a < b \leq m$ limitlerinde yer alan a ve b değerleri belirlenmiş p_0 (veya $1 - p_0$) değerine bağlı olarak bulunur. Süreç kontrol altındayken yani $H_0 : F = G$ olması durumunda yanlış alarm oranı (FAR)

$$FAR = 1 - p(m, n, j, a, b, F, F)$$

şeklinde hesaplanır. Örneklemin dağılımının Normal dağılım olduğu varsayımına dayalı olarak çizilen kontrol kartlarında yanlış alarm oranı genellikle 0.0027 kabul edilir. Kesikli bir rasgele değişken olan ön istatistiğinin yanlış alarm olasılığı $p_0 \geq P_0$ olacak şekilde bir değer seçilir. Belirlenen bu değere bağlı olarak FAR, $1 - P_0$ değerinden daha büyük bir değer olamaz.

3. Medyan Kontrol Kartı

Medyan, dayanıklı bir tahmin edicidir ve büyük ölçüm hatalarının olması durumunda tercih edilir. Medyan, örneklem ortalamasının aksine daha esnek bir konum tahmin edicidir ve uç değerlerden etkilenmez. Örneklem hacmi $n = 2s + 1$ (tek) iken Y örnekleminin medyanı $j = s + 1$ olmak üzere $Y_{j:n}$ 'dir. Medyan istatistiği kullanıldığından dolayı (3) eşitliğinde $b = m - a + 1$ olmalıdır. (3) eşitliğinde $b = m - a + 1$ yerine yazıldığında a ($a \geq 1$) değeri

$$\sum_{w=a}^{m-a} \frac{\binom{j+w-1}{w} \binom{m+n-j-w}{m-w}}{\binom{m+n}{m}} \geq p_0 \quad (4)$$

ile verilen (4) eşitliğini sağlayacak biçimde bulunur.

4. Çalışma Uzunluğu Dağılımı

Bir kontrol kartının performansı genellikle, çalışma uzunluğu dağılımına göre incelenir. Faz I aşamasında test örneklemini ile çizilen kartta kontrol limitleri arasında kalan (işaretlenen) nokta sayısı olan çalışma uzunluğu N rastgele değişkenin dağılımı $P(N = k) = E_{X_{(am)}, X_{(bm)}}(p^{k-1} - p^k)$, $k = 1, 2, \dots$ dir. Burada $E_{X_{(am)}, X_{(bm)}}(p^k)$ ifadesinde yer alan p niceliği

$$p = P(X_{am} \leq Y_{jn} \leq X_{bm}) = P(GF^{-1}\{U_{am}\} \leq U_{jn} \leq GF^{-1}\{U_{bm}\})$$

eşitliği ile hesaplanır. Burada U_{am} ve U_{bm} sürekli düzgün dağılımının a ve b . sıra istatistikleridir. Çalışma uzunluğu dağılımı iki yanlı kartlar için

$$\begin{aligned} E_{X_{(am)}, X_{(bm)}}(p^k) &= \int_0^1 \int_0^t \left[\frac{1}{B(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(GF^{-1}(t)^{(j+h)} - GF^{-1}(s)^{(j+h)} \right) \right]^k \\ &\quad \times \frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1} (t-s)^{b-a-1} (1-t)^{m-b} ds dt \\ &= D^*(k) \end{aligned} \quad (5)$$

biçiminde elde edilir [7]. Bu durumda iki yanlı kontrol kartlarında çalışma uzunluğunun dağılımı

$$P(N = k) = D^*(k-1) - D^*(k), \quad k = 1, 2, 3, \dots \text{ ve } D^*(0) = 1 \quad (6)$$

dır. (5) formülünde $F = G$ olduğunda, süreç kontrol altındadır denir.

4.1 Sistem kontrol altında iken çalışma uzunluğunun dağılımı

İki yanlı kontrol kartı için, (6) eşitliğinde verilen süreç kontrol altındayken çalışma uzunluğunun dağılımı

$$P_k(N = k) = D(k-1) - D(k), \quad k = 1, 2, 3, \dots, \quad D(0) = 1 \quad (7)$$

dır. Burada (6) eşitliğinden

$$D(k) = \int_0^1 \int_0^t \left[\frac{1}{B(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} (t^{(j+h)} - s^{(j+h)}) \right]^k \times \frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1} (t-s)^{b-a-1} (1-t)^{m-b} ds dt \quad (8)$$

elde edilir [6]. (8) eşitliği kullanılarak süreç kontrol altındayken ön istatistiklere dayalı kalite kontrol kartının çalışma uzunluğu dağılımının olasılıkları hesaplanır.

4.2 Ortalama çalışma uzunluğu

Shewhart kartlarında, çalışma uzunluğunun dağılımı $P(X = x) = (1 - p(\delta))^{x-1} p(\delta)$, $x = 1, 2, \dots$ olasılık fonksiyonu ile Geometrik dağılımdır. Burada δ , incelenen işlem parametresindeki kayma miktarı ve p parametresi kartta işaretlenen istatistiklerin kontrol limitlerinin dışına çıkma olasılığıdır. Bu nedenle, Shewhart türü kartta işlem parametresi delta kayması içerdiğinde,

$ARL = \frac{1}{p(\delta)}$ olur. Süreç kontrol altında iken ARL_0 olarak adlandırılır. Süreçte işaretlenen

istatistikler çok uzun bir süre kontrol limitleri içinde olduğunda, sürecin kararlı durumda olduğu bilinmektedir. Bu kararlı durumda sürecin ARL 'si, kararlı durum ARL 'si olarak bilinir. Burada yer alan p , (2) eşitliğinden hesaplanır. Böylece X_{am} ve X_{bm} ile koşullu ARL denklemi

$$ARL = \sum_{k=0}^{\infty} P(N > k) = \sum_{k=0}^{\infty} p^k$$

ve koşulsuz ARL denklemi

$$ARL = \sum_{k=0}^{\infty} E_{X_{(am)}, X_{(bm)}}(p^k) = \sum_{k=0}^{\infty} D^*(k)$$

dır. (5) eşitliğini kullanarak

$$ARL = \int_0^t \int_0^t [1 - C(s, t, j, n, F, G)]^{-1} \frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1} (t-s)^{b-a-1} (1-t)^{m-b} ds dt \quad (9)$$

elde edilir [7]. Burada

$$C(s, t, j, n, F, G) = \frac{1}{B(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(\{GF^{-1}(t)\}^{(j+h)} - \{GF^{-1}(s)\}^{(j+h)} \right)$$

dır. Süreç kontrol altındayken $G=F$ olacağından (9) eşitliği

$$ARL_0 = \sum_{k=0}^{\infty} D(k) = \int_0^t \int_0^t [1 - C(s, t, j, n)]^{-1} \frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1} (t-s)^{b-a-1} (1-t)^{m-b} ds dt \quad (10)$$

biçimine gelir [7].

Burada,

$$C(s, t, j, n) = \frac{1}{B(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(t^{(j+h)} - s^{(j+h)} \right)$$

dır. Bu kesimde verilen formüller iki yanlı kontrol kartı içindir.

ARL_0 'ın sonlu olması için a, b, j, m ve n sabitleri için belirli koşullar sağlanmalıdır. Chakraborti ve ark.(2004)'ün çalışmasında ARL_0 'ın sonlu olması için gerek ve yeter koşulun $(a-j)(n-j+1) + j(m-b+1) > 0$ olduğu ispatlanmıştır.

Verilerin geldiği dağılımın çarpık ya da ağır kuyruklu dağılım olması durumunda hesaplanan ARL_0 değerleri aşağıdaki Tablo 1'de verilmiştir.

Tablo 1- Bazı dağılımlar için ARL_0 değerleri

Dağılım	Medyan Kontrol Kartı için ARL_0 *	\bar{X} Shewhart Kontrol Kartı için ARL_0 **
Normal (0,1)	501.89	500
Laplace (0,1)	501.89	200.01
Gamma (1,1)	501.89	118.13
t(4)	501.89	141.51

* $m = 1000, n = 5, P_0 = 0.9978, a = 48, b = 953$

** $n = 5, P_0 = 0.998$

Ön istatistikleri kullanarak oluşturulan medyan kontrol kartları dağılımdan bağımsızdır. Medyan kontrol kartları dayanıklı kartlardır ve Shewhart ortalama kontrol kartlarına göre daha büyük ARL_0 değerine sahiptir. Bu durumda Shewart \bar{X} kontrol kartları yerine ön istatistiklere dayalı kontrol kartları tercih edilmelidir [7].

5. Uygulama

Bilgisayar desteği, istatistiksel süreç kontrolünün uygulanmasında önemli bir rol oynamaktadır. Bugün piyasada çok çeşitli istatistiksel programlar, yazılımlar bulunmaktadır. Bunlardan bazıları: JPM [28], MİNİTAB [29], SAS [30], R [31], MATLAB [32], NCSS [33] PALSTAT [34], QC EXPERT [35], SIGMAXL [36], SPSS [37], STATGRAPHICS CENTURION [38], STATISTICA[39] dir. Ancak bu programlar parametrik Shewhart kontrol kartları için gerekli hesaplamaları yapıp, grafikleri çizdiren programlardır [41]. Tereza ve Darja [41], bazı parametrik olmayan kartlar için EXCEL[40] uygulaması hazırlamışlardır. Parametrik olmayan kontrol kartı hakkında literatürde yeterince teorik bilgi olmasına rağmen sanayi uygulamaları hakkında pratik bilgiler yetersizdir. Parametrik olmayan kalite kontrol kartları için kontrol limitlerinin hesaplanması, *ARL* ve *FAR* değerlerinin hesaplanması oldukça zahmetlidir. İyi bilgisayar programlama bilmek gerektirmektedir. Bu çalışmada üretim yapan sanayi kuruluşlarının ve kişilerin parametrik olmayan kontrol kartlarını çizebilmeleri için kontrol limitlerini, *ARL* ve *FAR* değerlerini kolayca bulabilmeleri için tablolar oluşturulmuştur. Bilinmektedir ki Shewhart kontrol kartlarının limitlerinin hesaplanmasında yer alan katsayıların tablo olarak sunulmuş olması kullanıcılara büyük kolaylık sağlamaktadır [2].

Bu çalışmanın uygulama kısmında *FAR* hesabı için (4), *ARL* hesabı için (10) eşitliği kullanılmıştır. Bu tabloları hazırlamak için R programı kullanılmıştır.

5.1. Medyan Kontrol Kartı için hesaplamalar

$m = 50, 100, 250, 500, 750, 1000$, $n = 5, 11, 15, 25, 31$ için a , $b = m - a + 1$, *ARL*, *FAR* değerleri Tablo 2’de $P_0 = 0.9973$ için, Tablo 3’de $P_0 = 0.99$ için ve Tablo 4’de $P_0 = 0.995$ için verilmiştir. Tablo 2-4, bu çalışmanın özgün yanını oluşturmaktadır. Bu tablolarda literatürde yer almayan ve sanayi uygulamalarında ihtiyaç duyulabilecek olan $n=15, 21, 31$ ve $m=200, 250, 750$ değerleri için gerekli hesaplamalar yapılmıştır. Çalışmanın özgün yanını bu tablolar oluşturmaktadır.

Tablo 2 -Süreç kontrol altındayken medyan kontrol kartı için seçilen keyfi m , n ve $P_0 = 0.9973$ değerlerine göre belirlenen $AKL=a$, $ÜKL= b$ değerleri için hesaplanan FAR ve ARL değerlerinin tablosu

			<i>Referans örneklem hacmi - m</i>							
P_0	n	j		50	100	200	250	500	750	1000
0.9973	5	3	(a,b)	(1,50)	(4,97)	(9,192)	(12,238)	(25,476)	(38,712)	(51,950)
			FAR	0.0004	0.00102	0.00109	0.00125	0.00127	0.00128	0.00129
			ARL	∞	1550	728.6	565.5	460.2	432.3	419.5
0.9973	11	6	(a,b)	(5,46)	(11,90)	(24,177)	(31,220)	(64,437)	(97,654)	(130,871)
			FAR	0.00125	0.00106	0.00115	0.00128	0.00131	0.00132	0.00131
			ARL	9503	1630	726.18	579.2	456.1	424.2	409.8
0.9973	15	8	(a,b)	(6,45)	(14,87)	(31,170)	(40,211)	(82,418)	(124,626)	(166,835)
			FAR	0.00090	0.00098	0.0012	0.00135	0.00134	0.00134	0.00133
			ARL	44789.1	2320	763.2	581.8	439.5	427.2	413.1
0.9973	21	11	(a,b)	(8,42)	(18,83)	(38,163)	(49,201)	(101,400)	(153,598)	(205,795)
			FAR	0.00130	0.00488	0.0011	0.00129	0.00133	0.00488	0.00135
			ARL	51690.4	2477.1	966.4	623.4	484.7	437.6	417
0.9973	25	13	(a,b)	(8,43)	(19,82)	(42,159)	(53,198)	(110,391)	(167,584)	(224,777)
			FAR	0.00075	0.00092	0.00119	0.00116	0.00127	0.00131	0.00133
			ARL	173700	5183	986.2	839.6	526.2	458.9	430.2
0.9973	31	16	(a,b)	(9,41)	(21,80)	(46,155)	(59,192)	(122,379)	(184,567)	(247,754)
			FAR	0.00083	0.00092	0.0011	0.00127	0.00135	0.00130	0.00133
			ARL	18143.5	5680.6	1262.1	848.8	516.2	476.5	438.4

Tablo 3 - Süreç kontrol altındayken medyan kontrol kartı için seçilen keyfi m , n ve $P_0=0.99$ değerlerine göre belirlenen $AKL=a$, $ÜKL=b$ değerleri için hesaplanan FAR ve ARL değerlerinin tablosu

				<i>Referans örneklem hacmi - m</i>						
P_0	n	j		50	100	200	250	500	750	1000
0.99	5	3	(a,b)	(3,48)	(7,94)	(15,186)	(19,232)	(40,461)	(60,691)	(82,919)
			FAR	0.0036	0.00432	0.0043	0,00436	0,00477	0,00469	0,00499
			ARL	635.7	214.9	148.4	139.1	114.5	113.1	104.6
0.99	11	6	(a,b)	(7,44)	(15,86)	(32,169)	(40,211)	(83,418)	(125,626)	(167,834)
			FAR	0.00465	0.00428	0.00469	0.00450	0.00494	0.0047	0.00487
			ARL	642.2	245	146.1	142.8	113.3	110	108.4
0.99	15	8	(a,b)	(8,43)	(18,83)	(39,162)	(49,202)	(100,401)	(152,599)	(203,798)
			FAR	0.00351	0.00397	0.00482	0.00472	0.00478	0,005	0.00495
			ARL	1815.9	321.4	151.2	142.5	120	109.1	107.6
0.99	21	11	(a,b)	(10,41)	(22,79)	(46,155)	(58,193)	(118,383)	(153,598)	(240,761)
			FAR	0.00472	0.00488	0.0048	0.00478	0.00470	0.00134	0.00498
			ARL	2330.5	317.4	168.7	152	126.2	114	108.7
0.99	25	13	(a,b)	(10,41)	(23,78)	(49,152)	(62,189)	(127,374)	(192,559)	(258,743)
			FAR	0.00305	0.00402	0.0048	0.00457	0.00473	0.00478	0.00497
			ARL	10990	510.8	197.1	168.9	128.3	118.3	109.8
0.99	31	16	(a,b)	(11,40)	(25,76)	(53,148)	(67,184)	(138,363)	(208,543)	(279,722)
			FAR	0.00343	0.00422	0.0045	0.00454	0.00495	0.00484	0.00497
			ARL	4401.6	642	219.4	184.9	126.3	119.1	111.5

Tablo 4 -Süreç kontrol altındayken medyan kontrol kartı için seçilen keyfi m , n ve $P_0 = 0.995$ değerlerine göre belirlenen $AKL=a$, $ÜKL= b$ değerleri için hesaplanan FAR ve ARL değerlerinin çizelgesi

				<i>Referans örneklem hacmi m</i>						
P_0	n	j		50	100	200	250	500	750	1000
			(a,b)	(2,49)	(5,96)	(12,189)	(15,236)	(31,470)	(48,702)	(64,937)
0.995	5	3	FAR	0.0015	0.00176	0.00236	0.00229	0.00233	0.00249	0.00246
			ARL	5671	678.4	293.9	284.1	242.3	216.9	215.1
			(a,b)	(5,46)	(13,88)	(28,173)	(35,216)	(72,429)	(109,642)	(146,855)
0.995	11	6	FAR	0.00125	0.00225	0.00246	0.00233	0.0024	0.00243	0.00244
			ARL	9503	574.5	304	294.9	240.9	226.4	219.8
			(a,b)	(7,44)	(16,85)	(34,167)	(44,207)	(90,411)	(136,615)	(182,819)
0.995	15	8	FAR	0.00186	0.00206	0.00211	0.00244	0.00245	0.00245	0.00245
			ARL	7289.1	792.4	392.8	297.6	242.5	228.1	221.5
			(a,b)	(9,42)	(19,82)	(42,159)	(53,198)	(109,392)	(164,587)	(220,780)
0.995	21	11	FAR	0.00256	0.00178	0.00242	0.00238	0.00249	0.00240	0.00244
			ARL	9887.4	1393.4	380.6	335.8	248.2	239.1	226.1
			(a,b)	(9,42)	(21,80)	(45,156)	(57,194)	(118,383)	(178,573)	(239,762)
0.995	25	13	FAR	0.00155	0.002	0.00217	0.00221	0.00243	0.00239	0.00245
			ARL	4475	148.8	472.4	393	261	244.3	227.5
			(a,b)	(10,41)	(23,74)	(49,152)	(63,188)	(129,372)	(195,556)	(261,740)
0.995	31	16	FAR	0.00175	0.00205	0.0021	0.00247	0.00246	0.00244	0.00244
			ARL	9906	585.3	567.38	381	269.5	244.1	233.2

5.2. Veri seti üzerinde bir uygulama

Bu kesimde yukarıda verilen tabloların nasıl kullanılacağını açıklamak için bir veri seti üzerinde uygulama yapılmıştır. Keyfi m , n ve P_0 değerlerinin kombinasyonlarına göre bulunan a , b değerlerini Tablo 2'den okuyarak medyan kontrol kartının çizimine örnek vermek için bir veri seti üzerinde uygulama yapılmıştır.

Bunun için Montgomery [2]' den alınan sert fırında işlem için akış genişliği (mikron) verisi kullanılmıştır. Bu veri seti için Minitab 17'de yapılan analizler sonucunda $\bar{x}=1.51342$, $st.sapma=0.132989$, $medyan=1.51025$, $min=1.0931$, $maks=1,8698$, $\text{çarpıklık katsayısı}=0,0164794$, $\text{basıklık katsayısı}=0,278253$ olarak bulunmuştur. Anderson Darling uyum iyiliği sonuçlarına göre veri setinin Normal dağılıma, Log-Normal dağılıma, Logistic dağılıma ve Loglogistic dağılıma uyduğu ($p>0.05$) görülmüştür.

Referans örneklemeden

$$\bar{\bar{x}} = \frac{\sum_{i=1}^m \bar{x}_i}{m} = 1.5134, \quad S_p^2 = \frac{\sum_{i=1}^m (n_i - 1) S_i^2}{\sum_{i=1}^m (n_i - 1)} = 0.01802 \quad \text{ve} \quad \hat{\sigma} = \sqrt{S_p^2} = 0.134237 \quad \text{olarak hesaplanmıştır ve}$$

Shewhart ortalama kontrol kartının limitleri

$$AKL = \bar{\bar{x}} - 3 \frac{\sigma}{\sqrt{n}} = 1.333,$$

$$\dot{ÜKL} = \bar{\bar{x}} + 3 \frac{\sigma}{\sqrt{n}} = 1.6938$$

olarak elde edilir.

$m = 40 \times 5 = 200$, $n = 5$, $j = 3$ için simetrik parametrik olmayan kontrol kartları için $P_0 = 0.9973$ için Tablo 2'den $a = 9$, $b = 192$ olarak okunur. Buna göre referans örneklemeden $AKL = X_{9;m} = 1.2831$ ve $\dot{ÜKL} = X_{192;m} = 1,7473$ olarak bulunur.

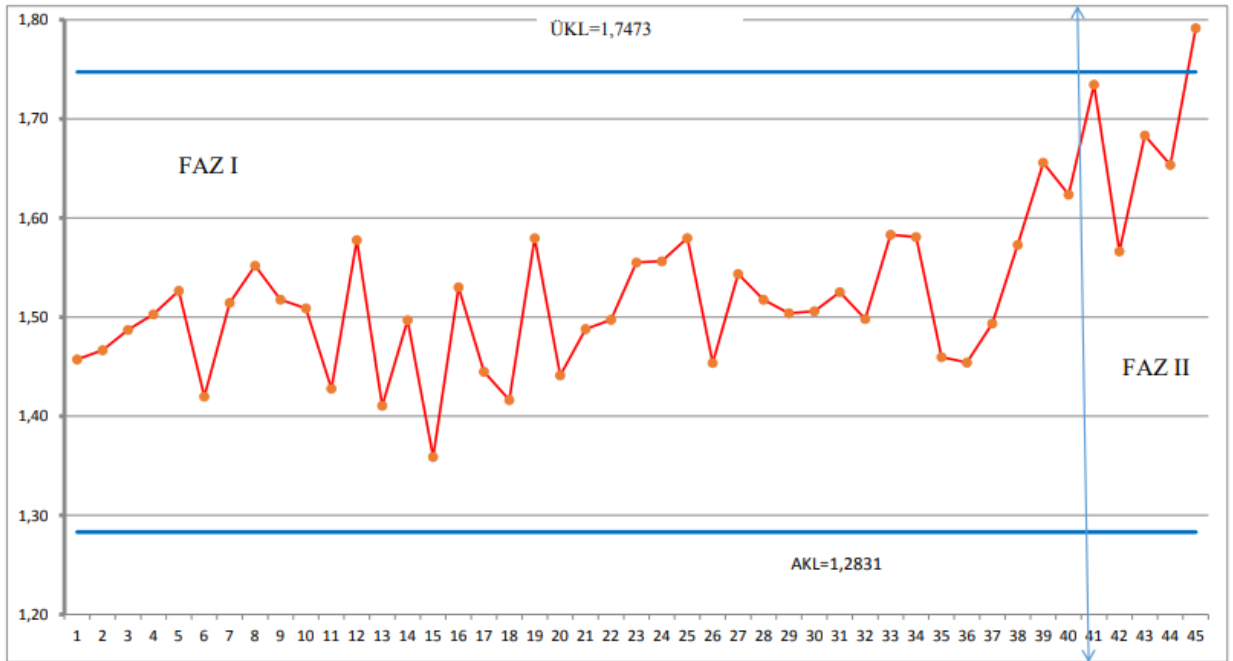
Veri Seti- Sert fırında işlem için akış genişliği (mikron) verisi

Sıra No	X_1	X_2	X_3	X_4	X_5	ORTALAMA	ORTANCA - X_{35}
1	1,3235	1,4128	1,6744	1,4573	1,6914	1,51188	1,4573
2	1,4314	1,3592	1,6075	1,4666	1,6109	1,49512	1,4666
3	1,4284	1,4871	1,4932	1,4324	1,5674	1,4817	1,4871
4	1,5028	1,6352	1,3841	1,2831	1,5507	1,47118	1,5028
5	1,5604	1,2735	1,5265	1,4363	1,6441	1,48816	1,5265
6	1,5955	1,5451	1,3574	1,3281	1,4198	1,44918	1,4198
7	1,6274	1,5064	1,8366	1,4177	1,5144	1,5805	1,5144
8	1,419	1,4303	1,6637	1,6067	1,5519	1,53432	1,5519
9	1,3884	1,7277	1,5355	1,5176	1,3688	1,5076	1,5176
10	1,4039	1,6697	1,5089	1,4627	1,522	1,51344	1,5089
11	1,4158	1,7667	1,4278	1,5928	1,4181	1,52424	1,4278
12	1,5821	1,3355	1,5777	1,3908	1,7559	1,5284	1,5777
13	1,2856	1,4106	1,4447	1,6398	1,1928	1,3947	1,4106
14	1,4951	1,4036	1,5893	1,6458	1,4969	1,52614	1,4969
15	1,3589	1,2863	1,5996	1,2497	1,5471	1,40832	1,3589
16	1,5747	1,5301	1,5171	1,1839	1,8662	1,5344	1,5301
17	1,368	1,7269	1,3957	1,5014	1,4449	1,48738	1,4449
18	1,4163	1,3864	1,3057	1,621	1,5573	1,45734	1,4163
19	1,5796	1,4185	1,6541	1,5116	1,7247	1,5777	1,5796
20	1,7106	1,4412	1,2361	1,382	1,7601	1,506	1,4412
21	1,4371	1,5051	1,3485	1,567	1,488	1,46914	1,488
22	1,4738	1,5936	1,6583	1,4973	1,472	1,539	1,4973
23	1,5917	1,4333	1,5551	1,5295	1,6866	1,55924	1,5551
24	1,6399	1,5243	1,5705	1,5563	1,553	1,5688	1,5563
25	1,5797	1,3663	1,624	1,3732	1,6887	1,52638	1,5797
26	1,4483	1,5458	1,4538	1,4303	1,6206	1,49976	1,4538
27	1,5435	1,6899	1,583	1,3358	1,4187	1,51418	1,5435
28	1,5175	1,3446	1,4723	1,6657	1,6661	1,53324	1,5175
29	1,5454	1,0931	1,4072	1,5039	1,5264	1,4152	1,5039
30	1,4418	1,5059	1,5124	1,462	1,6263	1,50968	1,5059
31	1,4301	1,2725	1,5945	1,5397	1,5252	1,4724	1,5252
32	1,4981	1,4506	1,6174	1,5837	1,4962	1,5292	1,4981
33	1,3009	1,506	1,6231	1,5831	1,6454	1,5317	1,5831
34	1,4132	1,4603	1,5808	1,7111	1,7313	1,57934	1,5808
35	1,3817	1,3135	1,4953	1,4894	1,4596	1,4279	1,4596
36	1,5765	1,7014	1,4026	1,2773	1,4541	1,48238	1,4541
37	1,4936	1,4373	1,5139	1,4808	1,5293	1,49098	1,4936
38	1,5729	1,6738	1,5048	1,5651	1,7473	1,61278	1,5729
39	1,8089	1,5513	1,825	1,4389	1,6558	1,65598	1,6558

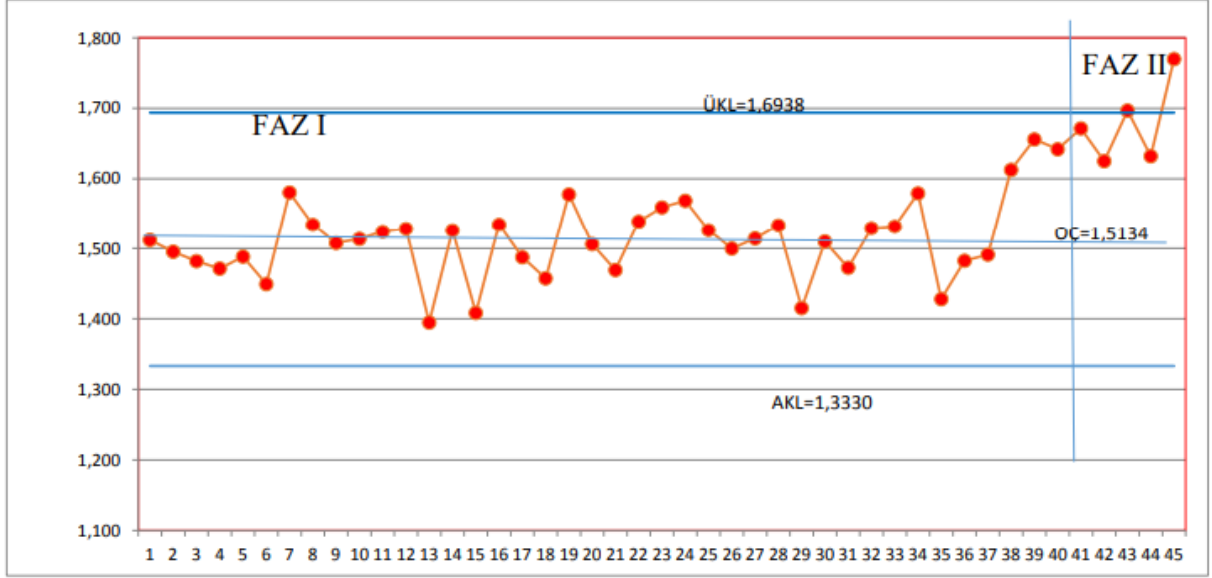
Veri Setinin devamı

Sıra No	X_1	X_2	X_3	X_4	X_5	ORTALAMA	ORTANCA – $X_{3,5}$
40	1,6236	1,5393	1,6738	1,8698	1,5036	1,64202	1,6236
41	1,412	1,7931	1,7345	1,6391	1,7791	1,67156	1,7345
42	1,7372	1,5663	1,491	1,7809	1,5504	1,62516	1,5663
43	1,5971	1,7394	1,6832	1,6677	1,7974	1,69696	1,6832
44	1,4295	1,6536	1,9134	1,7272	1,437	1,63214	1,6536
45	1,6217	1,822	1,7915	1,6744	1,9404	1,77	1,7915

Yukarıda verilen veri seti kullanılarak medyan kontrol kartı ve Shewhart kontrol kartı çizilerek sırasıyla Şekil-1 ve Şekil-2’de verilmiştir. Faz I de 40 örneklem kullanılmıştır. Son 5 örneklem Faz II’de kullanılmıştır. Medyan kontrol kartı için bulunan kontrol limitleri (1.2831, 1.7473) olarak Shewhart kontrol kartı için bulunan kontrol limitleri ise (1.3330, 1.6938) olarak bulunmuştur. Buna göre Shewhart kontrol kartı için bulunan kontrol limitlerinin daha dar olduğu görülmektedir. Sert fırında işlem için akış genişliği ölçümleri (mikron) verisi için Shewhart kontrol kartı çiziminde varyansların tahmin edicisi olarak birleştirilmiş örneklem varyansı kullanılmıştır. Medyan kontrol kartı için Tablo 2’den $FAR=0.00109$ ve $ARL=728.6$ olduğu görülür. Shewhart kontrol kartı için $FAR=0.002$ ve $ARL=500$ olarak hesaplanmıştır.



Şekil 1- Sert fırında işlem için akış genişliği ölçümleri (mikron) verisi için medyan kontrol kartı



Şekil 2- Sert fırında işlem için akış genişliği ölçümleri (mikron) verisi için Shewhart kontrol kartı

Şekil 1 ve Şekil-2’de görülen medyan ve Shewhart ortalama kontrol kartlarında Faz II’de süreç kontrol dışına çıkmıştır. Ancak Shewhart kontrol kartına göre medyan kontrol kartının ARL_0 değeri daha büyük ve FAR değeri daha küçük çıkmıştır. Medyan kontrol kartına göre 729 üründe 1 kez sistemin doğal nedenlerle kontrol dışına çıkacağı beklenmektedir.

6. Sonuç ve öneriler

Sanayide üretilen ürünün kalite kontrolü için Shewhart ortalama kontrol kartları yaygın olarak kullanılmaktadır. Ancak küçük örneklem durumunda, normal dağılım varsayımının şüpheli olduğu durumda, çarpık ya da ağır kuyruklu dağılım söz konusu olduğunda shewhart kontrol kartı ile sistemi takip etmek riskli olacaktır. Bu nedenle üretim yapan fabrikalara ürün kalitesini takip etmede ön istatistiklere dayalı ve dağılımdan bağımsız olan medyan kontrol kartlarının kullanması önerilir. Bir imalat hattında sürecin istatistiksel kontrolü için yapılan denetlemeler ile verimliliği artırmak, yüksek kar elde etmek ve zaman tasarrufu elde etmek açısından doğru kontrol kartının seçilmesi çok önemlidir. Bu çalışmada dağılımdan bağımsız kontrol kartlarının kullanımı için farklı P_0 , m , n değerleri için sıra istatistiklerine dayalı kontrol limitlerinin ARL ve FAR değerleri hesaplanmış ve tablo olarak verilmiştir. Ülkemizde üretim yapan ve istatistiksel kalite kontrol sürecine önem veren sanayi kuruluşları için hazırlanan bu tablolar büyük kullanım kolaylığı sağlayacaktır.

Kaynaklar

- [1] N.Balakrishnan, S. Kocherlakota, 1986, Effects of nonnormality on X charts: single assignable cause model. Sankhya: The Indian Journal of Statistics, Series B, 439-444.
- [2] D. C. Montgomery, 2005, Introduction to Statistical Quality Control. John Wiley & Sons.
- [3] N. Das, 2009, A comparison study of three non-parametric control charts to detect shift in location parameters. The International Journal of Advanced Manufacturing Technology, 41, 7-8, 799-807.
- [4] S. T. Bakir, M. R Reynold, 1979, A Nonparametric procedure for process control based on within-group rankin, Technometrics, 21, 75- 183.

- [5] G. J. Janacek, S. E. Meikle, 1997, Control charts based on medians, *The Statistician*, 46, 1, 19-31.
- [6] S Chakraborti, P. Van der Laan, S. T. Bakir, 2001, Nonparametric control charts: an overview and some results, *Journal of Quality Technology*, 33, 3, 304- 315.
- [7] S. Chakraborti, S., P. Van der Laan, M. A. Van de Wiel, 2004, A class of distribution-free control charts, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53, 3, 443-462.
- [8] N. Balakrishnan, I. S. Triantafyllou, M. V. Koutras, 2009, Nonparametric control charts based on runs and Wilcoxon-type rank-sum statistics, *Journal of Statistical Planning and Inference*, 139, 9, 3177-3192.
- [9] Balakrishnan, I. S. Triantafyllou, M. V. Koutras, 2010, A distribution-free control chart based on order statistics. *Communications in Statistics-Theory and Methods*, 39(20), 3652-3677.
- [10] I. S., Triantafyllou, 2018, Nonparametric control charts based on order statistics: Some advances. *Communications in Statistics-Simulation and Computation*, 47, 9, 2684-2702.
- [11] I. S. Triantafyllou, 2019, A new distribution-free control scheme based on order statistics. *Journal of Nonparametric Statistics*, 30, 1, 1-30.
- [12] S. T. Bakir, 2006, Distribution-free quality control charts based on signed-rank-like statistics. *Communications in Statistics-Theory and Methods*, 35(4), 743-757.
- [13] S. Chakraborti, S. Eryilmaz, 2007, A nonparametric Shewhart-type signed-rank control chart based on runs. *Communications in Statistics—Simulation and Computation*, 36, 2, 335-356.
- [14] N. Balakrishnan, I. S. Triantafyllou, M. V. Koutras, 2010, A distribution-free control chart based on order statistics. *Communications in Statistics-Theory and Methods*, 39, 20, 3652-3677.
- [15] N. Balakrishnan, I. S. Triantafyllou, M. V. Koutras, 2009, Nonparametric control charts based on runs and Wilcoxon-type rank-sum statistics, *Journal of Statistical Planning and Inference*, 139, 9, 3177-3192.
- [16] A. Mukherjee, S. Chakraborti, 2012, A distribution-free control chart for the joint monitoring of location and scale, *Quality and Reliability Engineering International*, 28, 3, 335-352.
- [17] S. Chowdhury, A. Mukherjee, S. Chakraborti, 2014, A new distribution-free control chart for joint monitoring of unknown location and scale parameters of continuous distributions, *Quality and Reliability Engineering International*, 30, 2, 191-204.
- [18] S. Chakraborti, M. A. Graham, 2007, Nonparametric control charts, *Encyclopedia of statistics in quality and reliability*, 1, 415-429.
- [19] S. Chakraborti, 2014, Nonparametric (distribution-free) quality control charts, *Wiley StatsRef: Statistics Reference Online*.
- [20] J. C. Malela-Majika, S. Chakraborti, M. A. Graham, 2016, Distribution-free precedence control charts with improved runs-rules, *Applied Stochastic Models in Business and Industry*, 32, 4, 423-439.
- [21] I. S. Triantafyllou, 2018, Nonparametric control charts based on order statistics: Some advances, *Communications in Statistics-Simulation and Computation*, 47, 9, 2684-2702.
- [22] D. S. Bai, I. S. Choi, 1995, X and R Control Charts For Skewed Population, *Journal of Quality Tecnology*, 27, 2, 305-701.
- [23] L. K. Chan, H. J. Cui, 2003, Skewness Correction ve Charts for Skewed Distributions, *Wilvy Periodicals, Inc., Naval Research Logistics*, 50, 1-19.
- [24] P. R. Tadikamalla, D. G. Popescu, 2007, Kurtosis Correction Metod for X and R Control Charts for Long-Tailed Symmetrical Distributions, *Naval Research Logistics, Wiley Periodicals*, 54, 4, 371-383.
- [25] B. Yazıcı, B. Kan, 2009, Asymmetric control limits for small samples, *Quality & Quantity*, 43, 865-874.
- [26] D. Karagöz, C. Hamurkaroglu C., 2012, Control charts for skewed distributions: Weibull, gamma, and lognormal, *Metodoloski Zvezki*, 9, 2, 95.
- [27] T. Ö. Yıldız, S. Ş. Vahaplar, 2015, An Application on fancy shirting fabric production through distribution-free quality control charts. *Tekstil ve Konfeksiyon*, 25, 2, 97-103.
- [28] Core Capabilities of JMP. JPM statistical discovery erişim adresi: https://www.jmp.com/en_us/home.html.
- [29] Minitab Statistical Software, Data Analysis, Statistical & Process Improvement Tools, State College, PA, erişim adresi: <https://www.minitab.com/en-us/products/minitab/free-trial>, 2019.
- [30] SAS/QC Software. SAS/QC, erişim adresi: https://www.sas.com/en_us/software/qc.html.
- [31] R Core Team.: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. 2019. Erişim adresi: <https://www.R-project.org/>.
- [32] Matlab, Mathworks, erişim adresi: <https://www.mathworks.com/>.

- [33] Quality Control. NCSS Statistical Software, erişim adrsi: <https://www.ncss.com/industries/quality-control/>.
- [34] Palstat CAQ, Quality Management System, erişim adresi: <https://www.palstat.cz/en/>.
- [35] QC Expert. TriloByte Statistical erişim adresi: <http://www.trilobyte.cz/en/>.
- [36] SigmaXL - Powerful Statistical and Graphical Analysis. SigmaXL Simply Smart Software [online]. [cit. 2018-01-18]. Available from: <https://www.sigmaxl.com/SigmaXL.shtml>.
- [37] Software IBM SPS. IBM Analytics, erişim adresi: <https://www.ibm.com/tr-tr/analytics/spss-statistics-software>.
- [38] Statgraphics Centurion XVII. Statgraphic erişim adresi: <http://www.statgraphics.com/statistical-process-control-charts>.
- [39] Statistica Product Features. TIBC Statistica, erişim adresi: <http://www.statsoft.com/Products/STATISTICA-Features>.
- [40] Excel, Microsoft, erişim adresi: <https://products.office.com/tr-tr/excel>.
- [41] S.Tereza, N.Darja, 2018, Software support of nonparametric control charts. 19th International Carpathian Control Conference (ICCC). IEEE, 2018. p. 456-461.