Çeşitli Sıralı Küme Örneklemeleri

Hülya ÇINGI * Yeşim ÜNYAZICI **

ÖZET

Sıralı küme örneklemesi, basit rasgele örneklemeye alternetif bir yöntem olarak McIntyre (1952) tarafından geliştirilmiştir. Sıralama hatalarını azaltmak amacıyla, daha sonraki yıllarda, aslı sıralı küme örneklemesine dayanan yöntemler geliştirilmiştir. Yeni geliştirilen yöntemler ile, kitle ortalamasının yansız tahminleri yapılabilmektedir. Bulunan ortalama tahminleri için varyans değerleri elde edilmiştir.

***Anahtar Kelimeler** : Sıralı küme örneklemesi, ortancaya dayalı sıralı küme örneklemesi, rasgele sıralı küme örneklemesi, sıralama hataları*

1. GİRİŞ

Basit rasgele örnekleme, örnekleme yöntemleri arasında temel niteliğinde olan en eski yöntemlerden biri olarak kullanılmaktadır Ancak basit rasgele örnekleme ile örneklem seçmek, bazı durumlarda zaman alıcı ve maliyetli olabilir. Örnekleme seçilen birimlere ait ölçümlerin yapılmasının, maliyetli ve zaman alıcı olduğu durumlarda tercih edilir. Basit rasgele örneklemeye seçenek olarak sıralı küme örneklemesi düşünülebilir. Sıralı küme örneklemesi (Ranked Set Sampling), 1952 yılında McIntyre tarafından önerilmiş bir örnekleme yöntemidir. Sıralı küme örneklemesi ile seçilen tüm birimlerin ölçülmesine gerek kalmadan, parametreler tahmin edilebilir. Yapılan çalışmalarda, sıralı küme örneklemesi ile örneklem seçiminin zaman ve emek açılarından tasarruf sağladığı ve maliyeti en aza indirdiği görülmüştür. Sıralı küme örneklemesi ile yapılan tahminler tahmin edicilerin yansızlık, tutarlılık gibi bazı özelliklerini de göstermektedirler.

Daha sonraki yıllarda, sıralı küme örneklemesinde bir takım değişiklikler yapılarak, temeli sıralı küme örneklemesine dayanan ancak ölçülen birimler açısından farklılık gösteren çeşitli yöntemler geliştirilmiştir. Amaç, basit rasgele örneklemeye alternatif olarak geliştirilen sıralı küme örneklemesini ve çeşitlerini tanıtmak ve kitle ortalamasının tahminini yeni yöntemlerle elde etmektir. Duyarlılığı yüksek tahmin ediciler elde edilebilir.

* Prof.Dr., Hacettepe Üniversitesi, İstatistik Bölümü

** Hacettepe Üniversitesi, İstatistik Bölümü

2. SIRALI KÜME ÖRNEKLEMESİ (SKÖ) (RANKED SET SAMPLİNG RSS)

1952 yılında Mclntyre tarafından önerilmiş bir örnekleme yöntemidir. Mclntyre, mera hasılası ortalamasının tahmininde, basit rasgele örnekleme'ye (BRÖ) göre daha etkin olan sıralı küme örneklemesi (SKÖ) yöntemini kullanmıştır. Daha sonra SKÖ yöntemi 1966'da Halls ve Dell tarafından, bir ormandaki bitkilerin ağırlığını ölçmek amacıyla kullanılmıştır. Takahasi ve Wakimoto, 1968 yılında, SKÖ ile bulunan kitle ortalaması tahmininin yansız olduğunu ve BRÖ'ye göre daha küçük varyanslı bir tahmin edici olduğunu göstermişlerdir. Ancak Takahasi ve Wakimoto, sıralamada hata olmayacağını düşünmüşlerdir. 1972 yılında, Dell ve Clutter, sıralamada hata olsa da olmasa da SKÖ ile yapılan kitle ortalaması tahmininin yansız ve BRÖ'ye göre etkili bir yöntem olduğunu göstermişlerdir. Son yıllarda SKÖ üzerinde bir takım değişiklikler yapılarak, temeli SKÖ'ye dayanan ancak SKÖ'den daha etkili olan yöntemler Muttlak H.A tarafından 1995 yılında geliştirilmiştir. Çiftli Sıralı Küme Örneklemesi (Paired Rank Set Sampling PRSS) , Ortancaya Dayalı Sıralı Küme Örneklemesi (Median Rank Set Sampling MRSS), Eşit Aralıklı Sıralı Küme Örneklemesi (Equal Interval Ranked Set Sampling EIRSS) , Rasgele Sıralı Küme Örneklemesi (Random Rank Set Sampling RRSS) gibi. Bu yöntemler için göreli duyarlılıklar bulunarak, çeşitli dağılımlar için karşılaştırmalar yapılmıştır.

Sıralı Küme Örneklemesi (SKÖ), çevresel araştırmalarda, tarımda, ekolojide sıkça kullanılan bir örnekleme yöntemidir. Bu yönteme sıralı küme örneklemesi denilmesinin nedeni ise, kitleden rasgele seçilen birimlerin sıralanmasıyla oluşturulmasıdır. Sonsuz büyüklükteki kitlelerde kullanılabilir. Sonsuz büyüklüklü bir kitleden, m sayıda birim seçilir Bu birimler gözlemsel olarak, herhangi bir ölçüm yapılmadan sıralanır. En küçük birim ölçülür, diğer birimler kitleye iade edilir. İkinci kez m birimlik bir örneklem seçilir ve gözle sıralama yapılır. 2. en küçük birim ölçülür, diğer birimler kitleye iade edilir. Bu şekilde işlemler m küme için yapılır. m. kümede m birim sıralanır ve en büyük birim ölçülür. Böylelikle sadece m birimin ölçümü yapılmış olur. m küme büyüklüğü genellikle 5 ya da 6 olarak alınır. Daha çok alındığında sıralmada hata olacağı düşünülmüştür. Yapılan işlemler, r kez tekrarlanılabilir. m ve r araştırmacı tarafından belirlenen değerlerdir. m küme büyüklüğü, r ise tekrar sayısıdır. Böylece m^2r birim ölçmek yerine sadece mr birim ölçülerek, zaman, emek ve maliyet açısından kazanç sağlanır. SKÖ ile kitle ortalaması ve varyansı tahmin edilebilir.
SKÖ'nin BRÖ'ye göre göreli duyarlılığı, kitle dağılımına ve küme büyüklüğüne (m) de bağlıdır. Kitle dağılımının çarpıklığı arttıkça, göreli duyarlılık azalır. Göreli duyarlılık, m küme büyüklüğü ile doğru, r tekrar sayısı ile ters orantılıdır.

Kitle ortalamasının yansız tahmin edicisi aşağıdadır :

$$\bar{y}_{SKÖ} = \frac{1}{mr} \sum_{j=1}^{r} \sum_{i=1}^{m} Y_{(i:m)j} \tag{1}$$

Burada m küme büyüklüğünü ve r tekrar sayısını göstermektedir.

σ^2 varyansının sıralı küme örneklemesinde tahmin edicisi eşitlik (2) ile verilmiştir.

$$\hat{\sigma}^2 = \frac{1}{mr-1}\sum_{j=1}^{r}\sum_{i=1}^{m}(Y_{(i:m)j} - \bar{y}_{SKÖ})^2 \qquad (2)$$

SKÖ'nde kitle ortalamasının tahmininin varyansı, aşağıdaki eşitlik ile verilebilir.

$$\mathrm{Var}(\bar{y}_{SKÖ}) = \frac{1}{mr}\{\sigma^2 - \frac{1}{m}\sum\sum(\mu_{(i:m)j} - \mu)^2\} \qquad (3)$$

Eşitlik (3) ile verilen varyansın kare kökü de standart hatadır.

Burada $Y_{(i:m)j}$, j. tekrardaki m büyüklüklü i. kümedeki ölçülmüş birimdir.

$\mu_{(i:m)j}$ ise j. tekrardaki m büyüklüklü i. kümenin ortalamasıdır.

Ayrıca (3). eşitlikte verilen varyansın tahmin edicisi de bulunabilir.

$$\hat{Var}(\bar{y}_{SKÖ}) = \frac{1}{mr}\left\{\hat{\sigma}^2 - \frac{1}{m}\sum_{j=1}^{r}\sum_{i=1}^{m}(\mu_{(i:m)j} - \bar{y}_{SKÖ})^2\right\} \qquad (4)$$

3. ÇİFTLİ SIRALI KÜME ÖRNEKLEMESİ (ÇSKÖ) (PAİRED RANK SET SAMPLİNG PRSS)

1996 yılında Muttlak tarafından geliştirilen bir yöntemdir. SKÖ'de birtakım değişiklikler yapılmıştır. Sonsuz büyüklükteki kitleden, m birim rasgele olarak seçilir. Bu birimler gözlemsel olarak, herhangi bir ölçüm yapılmadan sıralanır. En küçük birim ve en büyük birim ölçülür, diğer birimler kitleye iade edilir. İkinci kez m birim rasgele olarak seçilir. İkinci en küçük ve ikinci en büyük birim ölçülür, diğer birimler kitleye iade edilir. Bu şekilde, k=m/2 (m çift ise) ya da L=(m+1)/2 (m tek ise) küme için aynı şekilde, ölçümler yapılır. . k ve L küme sayısı, m ise küme büyüklüğüdür. Küme içinde ilgilenilen değişkene göre yapılan sıralamadan sonra (ölçüm yapılmadan sadece gözle yapılan sıralama) 1. kümeden en büyük ve en küçük sıralı birimler, 2. kümeden 2. en büyük ve 2. en küçük sıralı birimler seçilir. m çiftse, k. kümeden k. en küçük ve (k+1). en büyük birim seçilmiş olur. Eğer m tekse, L. kümeden L. en küçük ve L. en büyük birim seçilmiş olur. $m^2/2$ birim seçilmiş ancak sadece m tanesi ölçülmüştür. r kez tekrar yapılırsa, mr birim ölçülmüş olur. r tekrar sayısıdır. ÇSKÖ kullanılarak tüm dağılımlar için duyarlılıkta artış sağlanır. ÇSKÖ'nin BRÖ'ye etkinliği, m küme büyüklüğü arttıkça artar.

r =1 için (tek tekrar yapıldığında) ÇSKÖ ile kitle ortalamasının tahmin edicisi :

$$\bar{y}_{(k)} = \frac{1}{m}\left(\sum_{i=1}^{k}Y_{i(i)} + \sum_{i=1}^{k}Y_{i(m+1-i)}\right) \quad \text{, m çift ise} \qquad (5)$$

$$\bar{y}_{(L)} = \frac{1}{m}\left(\sum_{i=1}^{L} Y_{i(i)} + \sum_{i=1}^{L-1} Y_{i(m+1-i)}\right) \qquad \text{, m tek ise}$$

Burada $Y_{i(i)}$: i. kümedeki i. sıralı birim

$Y_{i(m+1-i)}$: i. kümedeki (m+1-i). sıralı birim

ÇSKÖ'nde kitle ortalaması tahmininin varyans değeri aşağıdaki eşitlik ile verilebilir.

$$Var(\bar{y}_{(k)}) = Var(\bar{y}_{(SK\ddot{O})}) + \frac{2}{m^2}\sum\sum Cov(Y_{(i)}, Y_{m+1-i)}) \qquad (6)$$

4. ORTANCAYA DAYALI SIRALI KÜME ÖRNEKLEMESİ (ODSKÖ) (MEDİAN RANK SET SAMPLİNG MRSS)

1997 yılında, Muttlak H.A tarafından sıralamadaki hataların azaltması amacıyla ortaya atılmıştır.Kitleden m birim rasgele olarak seçilir. Herhangi bir ölçüm yapılmadan, m seçilen m birim gözle sıralanır. Eğer m tek ise, kümeden, ölçüm için, kümede orta değeri ya da en ortada yer alan birim alınır. (m+1)/2. sıralı birim ölçülür. Eğer m çiftse, ilk m/2 kümede (kümelerin yarısında), (m/2). en küçük sıralı birim, ikinci m/2 kümede ise, (m+2)/2. sıralı birim alınır. m küme için bu şekilde ölçümü yapılacak birimler seçilir. m büyüklüklü m kümeden, herbirinde ortanca birim seçilerek, m birimin ölçümü yapılmış olur. m^2 birim ölçmek yerine sadece m birim ölçülmektedir. İşlem, r kez tekrarlandığında mr ölçüm elde edilir. Eğer y'nin olasılık yoğunluk fonksiyonu simetrik ise, tahmin ediciler kitle ortalamasının yansız tahmin edicileridir ve varyansları hem BRÖ tahmininden hem de SKÖ tahmininden daha küçüktür. Ayrıca iki aşamalı örneklemede, ikinci aşamada BRÖ yerine ODSKÖ kullanılarak hem kitle ortalamasının hem de kitle büyüklüğünün yansız tahminleri yapılabilir.

r=1 tekrar için ODSKÖ ile kitle ortalamasının tahmin edicisi :

$$\bar{y}_{(ODSK\ddot{O}1)} = \frac{1}{m}\sum_{i=1}^{m} Y_{i(\frac{m+1}{2})} \qquad \text{, m tek ise} \qquad (7)$$

$$\bar{y}_{(ODSK\ddot{O}2)} = \frac{1}{m}\left(\sum_{i=1}^{L} Y_{i(\frac{m}{2})} + \sum_{i=L+1}^{m} Y_{i(\frac{m+2}{2})}\right) \qquad \text{, m çift ise}$$

Burada $Y_{i(\frac{m+1}{2})}$, i. kümedeki $(\frac{m+1}{2})$. sıralı birimi göstermektedir. m hem küme sayısı hem de küme büyüklüğüdür.

Varyans değeri de m'nin tek ya da çift olmasına bağlı olarak bulunur.

$$Var(\bar{y}_{(ODSKÖ1)}) = Var(\bar{y}_{BRÖ}) - \frac{1}{m^2}\sum_{i=1}^{m}(\mu_{i(\frac{m+1}{2})} - \mu)^2 \text{ , m tek ise} \quad (8)$$

$$Var(\bar{y}_{(ODSKÖ2)}) = Var(\bar{y}_{BRÖ}) - \frac{1}{m^2}(\sum_{i=1}^{m/2}(\mu_{i(\frac{m}{2})} - \mu)^2 + \sum_{i=m/2+1}^{m}(\mu_{i(\frac{m+2}{2})} - \mu)^2) \text{ , m çift}$$

ise

Bu eşitlikte verilen $\mu_{i(\frac{m+1}{2})}$, (m+1)/2. birimin ölçüldüğü i. kümenin ortalamasıdır. Diğer ortalamalar da benzer şekilde tanımlanabilir. $\mu_{i(\frac{m}{2})}$, (m/2). birimin ölçüldüğü i. kümenin ortalaması ve $\mu_{i(\frac{m+2}{2})}$, (m+2)/2, birimin ölçüldüğü i. kümenin ortalamasıdır (Muttlak, 1997).

Eşitlik (8) ile verilen varyansın tahmini de aşağıdaki şekilde verilebilir.

$$\hat{Var}(\bar{y}_{(ODSKÖ1)}) = \hat{Var}(\bar{y}_{BRÖ}) - \frac{1}{m^2}\sum_{i=1}^{m}(\mu_{i(\frac{m+}{2})} - \bar{y}_{ODSKÖ1})^2 \text{, m tek ise} \quad (9)$$

$$\hat{Var}(\bar{y}_{(ODSKÖ1)}) = \hat{Var}(\bar{y}_{BRÖ}) - \frac{1}{m^2}(\sum_{i=1}^{m/2}(\mu_{i(\frac{m+1}{2})} - \bar{y}_{ODSKÖ2})^2 + \sum_{i=m/2+1}^{m}(\mu_{i(\frac{m+1}{2})} - \bar{y}_{ODSKÖ2})^2),$$

m çift ise

5. EŞİT ARALIKLI SIRALI KÜME ÖRNEKLEMESİ (EASKÖ) (EQUAL INTERVAL RANK SET SAMPLİNG EIRSS)

2001 yılında Syed S. Hossain, tarafından önerilmiştir. Dikdörtgensel, normal ve üstel dağılımlar düşünülerek diğer yöntemlerle (SKÖ, ODSKÖ, BRÖ) karşılaştırmalar yapılmıştır. m büyüklüklü m rasgele küme oluşturmak yerine k tane m büyüklüklü küme oluşturulur (k<m). Her kümenin m_i. birimi ölçülür. Bu teknik ile, özellikle kitle dağılımı simetrik ise, duyarlılığı yüksek tahmin ediciler elde edilmiştir.

Kitle ortalamasının tahmin edicisi :

$$\bar{y}_{(EASKÖ)} = \frac{1}{k}\sum_{i=1}^{k} Y_{(m_i:m)j} \quad (10)$$

Burada $Y_{(mi:m)j}$: j. tekrarda m büyüklüklü i. kümedeki m_i. sıralı birimdir.

Ortalamanın varyansı ise,

$$Var(\bar{y}_{(EASKÖ)}) = \frac{\theta^2}{k^2} \sum_{i=1}^{k} \beta_{m_i:m_i:m} \tag{11}$$

Burada $\beta_{m_i:m_i:m} = \dfrac{Var(Y_{m_i:m})}{\theta^2}$.

$$\alpha_{m_i:m} = \frac{E(Y_{m_i:m}) - \gamma}{\theta}$$

l,t : dağılıma bağlı sabitler

γ,θ : yer ve durum parametreleridir.

6. RASGELE SIRALI KÜME ÖRNEKLEMESİ (RSKÖ) (RANDOM RANK SET SAMPLİNG RRSS)

SKÖ, ÇSKÖ, ODSKÖ, EASKÖ gibi tekniklerde sabit bir küme büyüklüğü ve sabit bir tekrar sayısı vardır. Ancak gerçekte, küme büyüklüğü raslantı değişkeni olabilir. Bu düşünceyle 2001'de I. Rahimov ve H.A. Muttlak tarafından rasgele sıralı küme örneklemesi (Random Rank Set Sampling RRSS) geliştirilmiştir. Diğer yöntemlerle karşılaştırıldığında, duyarlılıkta artış sağladığı görülmüştür.

1.) Rasgele Küme Büyüklüğü ve Tek Tekrar Olma Durumu :

$$\bar{y}_{(m)} = \frac{1}{m} \sum_{i=1}^{m} y_i^{(m)} \tag{12}$$

Kitle ortalamasının yansız tahmin edicisidir.

$$Var(\bar{y}_{(m)}) = E\left[\frac{1}{m}\sigma^2_{(m)}\right] \quad , \quad \sigma^2_{(m)} = \frac{1}{m} \sum_{i=1}^{m} \sigma^2_{m_i} \tag{13}$$

Burada $\bar{y}_{(m)}$: Kitle ortalaması μ'nün yansız tahmincisidir

$y_i^{(m)}$: Bağımsız ve aynı dağılımlı sıralı birimler

$\sigma^2_{m_i}$: m_i büyüklüklü i. küme varyansıdır.

2.) Rasgele Küme Büyüklüğü ve Sabit Tekrar Sayısı Olma Durumu :

$$\bar{y}_{(r)} = \frac{1}{r} \sum_{i=1}^{r} \bar{y}_{m_i i} \tag{14}$$

Kitle ortalamasının yansız tahmin edicisidir.

$$\mathrm{Var}(\bar{y}_{(r)}) = \frac{1}{r}E\left[\frac{1}{m_i}\sigma^2_{(m_i)}\right] \quad (15)$$

$\bar{y}_{m_i i}$: m_i büyüklüklü i. küme ortalaması

$\sigma^2_{(m_i)}$: m_i büyüklüklü i. küme varyansıdır.

1) Giriş Bölümünde üçüncü satırda belirtilen 'gözlemlerin tek tek ölçümlerinin yapılması' ile anlatılmak istenen, örnekleme m^2 birim seçilmektedir (tek tekrar için). Ancak bu birimler için herhangi bir ölçüm yapılmamaktadır.Bu seçilen birimler, bir ölçüm gerektirmeden, sadece görsel olarak sıralanmaktadırlar. Daha sonra seçilen m birim ölçülmektedir.
2) 3. sayfa, 2. paragrafta ise, m küme büyüklüğü, r ise tekrar sayısıdır. m büyüklüklü m tane küme oluşturulur. Her kümedeki birimler, ilgilenilen değişkene göre, ölçüm yapmadan (görsel olarak) sıralanır. En küçük birim ölçülür, diğer birimler kitleye iade edilir. İkinci kez m birimlik bir örneklem seçilir ve gözle sıralama yapılır. 2. en küçük birim ölçülür, diğer birimler kitleye iade edilir. Bu şekilde işlemler m küme için yapılır. m. kümede m birim sıralanır ve en büyük birim ölçülür . Eğer gerekli görülürse, işlem r kez tekrarlanır. r kez tekrar yapıldığında, toplam mr tane birim ölçülmüş olur ve ölçümü yapılmış olan bu birimler sıralı küme örneklemini oluşturur.
3) Birimler ilgilenilen değişkene göre sıralandıklarında, herhangi bir ölçüm yapılmamakta, birimler göz ile,ölçüm yapılmadan, sıralanmaktadırlar.
4) Çiftli sıralı küme örneklemesi için, m küme büyüklüğü, k veya L (m'nin tek ya da çift olmasına bağlı) küme sayısıdır. r tekrar sayısıdır.
5) 4. Sayfada, $Y_{i(i)}$ ile i. kümedeki i. sıralı birim gösterilmektedir. Bu birim, ölçüm için seçilen birimdir. Küme içindeki birimler sıralandıktan sonra (ölçüm yapılmadan), i. kümedeki i. sıralı birim ölçüm için seçilir. Terim sözcüğüne daha önce yer verilmediğinden, burada da tercih edilmemiştir.

KAYNAKLAR

BHOJ, S.D & AHSANULLAH, M., (1996), '*Estimation of Parameters of the Generalized Geometric Distribution Using Rank Set Sampling*', Biometrics, 52, 658-694

HOSSAİN SYED, S., (2001), '*Non-parametric Selected Rank Set Sampling*', Biometrical Journal, 43, 1, 97-105

MUTTLAK, H.A., (1995), '*Parameters Estimation in a Simple Linear Regression Using Rank Set Sampling*', Biometrical Journal, 37 ,7, 799-810

MUTTLAK, H.A., (1996), '*Pair Rank Set Sampling*' , Biometrical Journal, 38, 7, 879-885

MUTTLAK, H.A & SAMAWİ, H.M., (1996), '*Estimation of Ratio Using Rank Set Sampling*', Biometrical Journal, 38, 6, 753-764

MUTTLAK, H.A., (1998), *'Median Rank Set Sampling With Size Biased Probability of Selection'*, Biometrical Journal, 40 ,4, 455-465

MUTTLAK, H.A & RAHİMOV, I., (2001), *'Random Rank Set Samples', Pakistan Journal of Statistics*, Vol. 17(1) 51-66

Various Kinds Of Ranked Set Sampling

ABSTRACT

Ranked set sampling was first suggested by McIntyre (1952) as an alternative way of simple random sampling. Later, in order to reduce the ranking errors, some other methods based on rank set sampling are suggested. By using rank set sampling method, population mean can be estimated. The variation values for the mean estimates are obtained.

***Key Words** : Ranked set sampling, median rank set sampling, random rank set sampling, ranking errors*