SIRALI KÜME ÖRNEKLEMESİNDE YARDIMCI DEĞİŞKEN KULLANILARAK YIĞIN ORTALAMASININ TAHMİNİ

Yaprak Arzu ÖZDEMİR*

ÖZET

Sıralı Küme Örneklemesi birimlerin ilgilenilen değişken bakımından ölçümünün maliyet, zaman ve emek anlamında pahalı olduğu, ancak birimleri pahalı ölçüm gerektirmeyen bazı tekniklerle aynı değişken bakımından sıralamanın mümkün olduğu durumlarda sıkça kullanılan bir örnekleme yöntemidir. Bu gibi durumların ortaya çıktığı çevre, tarım ve tıp gibi alanlarda, sıralı küme örneklemesi yığın ortalamasını tahmin etmede basit tesadüfi örneklemeye göre daha etkindir. Bu çalışmada, sıralı küme örneklemesinde birimlerin yardımcı değişken kullanılarak sıralandığı durum ele alınarak, yığın ortalamasının tahmini incelenmiştir. Ayrıca ladin ağaçlarının yükseklik ve gövde çapı verisi kullanılarak, simülasyon çalışması ile sıralı küme örneklemesi ve basit tesadüfi örnekleme farklı örnek çapları için göreli etkinlikleri bakımından karşılaştırılmıştır.

Anahtar Kelimeler: *Sıralı Küme Örneklemesi, Basit Tesadüfi Örnekleme, Yardımcı Değişken, Göreli Etkinlik, Doğrusal Regresyon Modeli*

1. GİRİŞ

Günümüzde özellikle çevre, ekoloji, tarım ve tıp gibi alanlara yönelik çalışmalarda, ilgilenilen değişkenin ölçümünün maliyet bakımından oldukça pahalı ya da zaman ve emek bakımından oldukça zor olduğu durumlarla sıkça karşılaşılmaktadır. Dolayısıyla bu tür alanlarda, en az örnek çapı ile yığını en iyi şekilde temsil edecek bir örnekleme yönteminin kullanılmasına ihtiyaç duyulmaktadır. Bu amaca yönelik bir örnekleme yöntemi ilk olarak 1952 yılında McIntyre tarafından sıralı küme örneklemesi (SKÖ) adıyla önerilmiştir (McIntyre, 1952). McIntyre SKÖ ile elde edilen örnekten meralardaki ortalama ürün miktarını tahmin ederek SKÖ yönteminin yığın ortalamasını tahmin etmede basit tesadüfi örneklemeye (BTÖ) göre oldukça etkin bir örnekleme yöntemi olduğunu göstermiştir.

SKÖ'de örnek seçimi iki adımda yapılmaktadır. Örnek seçim işleminde öncelikle; herhangi bir $F(y)$ dağılım fonksiyonuna sahip bir yığından, m çaplı m tane tesadüfi örnek BTÖ ile seçilir ve bu örneklerin her biri "küme" olarak isimlendirilir. Bu işlem yığından seçilecek m^2 çaplı basit tesadüfi örneğin her biri m çaplı m kümeye tamamen tesadüfi olarak paylaştırmasıyla da yapılabilir. Örnek seçim işleminin birinci

* Gazi Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü 06500 Beşevler/ ANKARA

adımında, her bir küme ilgilenilen Y değişkeni bakımından hassas olmayan bir ölçümle küçükten büyüğe doğru sıralanır. Bu ölçüm, yüksek maliyet gerektirmeyen düşük düzeyli bir ölçüm olup, birimlerin sıralanması; daha önceki deneyimler, görsel bir ölçüm veya yardımcı bir değişken kullanılarak yapılabilir. İkinci adımda, birinci kümeden ilk sıradaki birim, ikinci kümeden ikinci sıradaki birim ve bu şekilde devam edilerek m'nci kümeden m'nci sıradaki birim alınarak, istenilen hassaslığı sağlayan yüksek düzeyli bir ölçümle Y değişkeni bakımından ölçülür. Buradan elde edilecek $Y_{(1)}, Y_{(2)}, \ldots, Y_{(m)}$ örneği, m çaplı sıralı küme örneğini oluşturur. Ayrıca $Y_{(i)}$, i=1,2,...,m olmak üzere, i'nci kümeden hassas ölçüm yapılarak alınan i'nci sıradaki örnek birimini ifade eder. Birimlerin hassas olmayan ölçümle sıralanmasında hata yapılmadığı varsayımı altında, $Y_{(i)}$ aynı zamanda, m çaplı sıralı küme örneğinde i'nci kümedeki i'nci sıra istatistiğini göstermektedir. Ayrıca $Y_{(i)}$ sıra istatistiği, herhangi bir $F(y)$ dağılım fonksiyonuna sahip m çaplı tesadüfi bir örnekteki i'nci sıra istatistiği $Y_{i:m}$ ile aynı dağılıma sahiptir. Ancak, her bir m çaplı tesadüfi küme, birbirinden bağımsız olduğundan, $Y_{(i)}$ sıra istatistikleri de birbirinden bağımsız olacaktır. SKÖ ile örnek seçim işlemi, istenilen n örnek çapını elde etmek amacıyla, r kez tekrarlanabilir. Böylece r tekrar sonunda $n=mr$ çaplı sıralı küme örneği elde edilir ve n çaplı sıralı küme örneğinden elde edilecek yığın ortalamasına ilişkin tahmin edici,

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

şeklinde tanımlanır (McIntyre, 1952). Ayrıca yukarıda açıklanan örnek seçim işleminde her bir kümeden farklı sıradaki birimler ölçülerek, çeşitli sıralı küme örneklemeleri geliştirilmiştir. Bu konu ile ilgili bir tanıtım Çıngı ve Ünyazıcı (2002) tarafından yapılmıştır.

Küme içindeki birimlerin, yüksek maliyet gerektirmeyen düşük düzeyli ölçüm teknikleriyle sıralanmasında, herhangi bir sıralama hatası yapılmadığı varsayımı altında, $\bar{Y}_{SKÖ}$'ne ilişkin varyans,

$$Var(\bar{Y}_{SKÖ}) = \frac{1}{mr} \sum_{i=1}^{m} \frac{\sigma^2_{i:m}}{m} = Var(\bar{Y}_{BTÖ}) - \frac{1}{m^2 r} \sum_{i=1}^{m} \left(\mu_{i:m} - \mu_y\right)^2 \tag{2}$$

şeklinde oluşur. Burada μ_y, Y değişkenine ilişkin yığın ortalamasını, $\mu_{i:m}$ ve $\sigma^2_{i:m}$ ise sırasıyla $Y_{(i)}$ sıra istatistiğine ilişkin yığın ortalamasını ve varyansını ifade eder. Ayrıca aynı n örnek çaplı BTÖ'den elde edilen ortalamaya ilişkin varyans

$$Var(\bar{Y}_{BTÖ}) = \frac{1}{n} \sigma_y^2$$

olmak üzere, σ_y^2, Y değişkenine ilişkin yığın varyansıdır. Buradan ortalamaya ilişkin varyanslar karşılaştırıldığında, etkinlik bakımından (2)'nin daha üstün olduğu görülmektedir. Dolayısıyla SKÖ'nün BTÖ'ye Göreli etkinliği (GE) (Relative Efficiency)

$$GE_{SKÖ} = \frac{Var(\overline{Y}_{BTÖ})}{Var(\overline{Y}_{SKÖ})} = \frac{1}{1-\{\frac{1}{m}\sum_{i=1}^{m}(\mu_{i:m}-\mu_y)^2/\sigma_y^2\}} \qquad (3)$$

şeklinde tanımlanır (Patil vd, 1994). $GE_{SKÖ}$'nin büyüklüğü yığının dağılımına, sıralamanın doğruluğuna ve küme çapı m'ye bağlıdır.

SKÖ yönteminin başarısı, özellikle, örnek seçim işleminde birimlerin ilgilenilen Y değişkeni bakımından doğru sıralanıp sıralanmadığına bağlıdır. Sıralama işlemi, ilgilenilen Y değişkenine göre hassas ölçüm gerçekleştirilmeden görsel bir sıralama veya yardımcı değişken kullanılarak sıralama gibi daha ucuz ve kolay bir teknikle yapıldığından, Y değişkeni bakımından sıralama hatası yapılması olasıdır. Bu konuyla ilgili olarak, Dell ve Clutter (1972), David ve Levine (1972), Stokes (1977), Ridout ve Cobby (1987), sıralama hatası durumunda, SKÖ altında yığın ortalamasına ilişkin tahmin edicinin özelliklerini incelemişlerdir. Yapılan bu çalışmalar, yığından birimlerin SKÖ ile farklı teknikler ile sıralanarak seçilmesi durumunda, yığın ortalamasına ilişkin tahmin edicinin sapmasızlık özelliğini koruduğunu ve örnek birimlerinin sıralanmasında tesadüfi bir sıralama oluşturacak kadar hata yapılmadığı sürece, SKÖ'nün BTÖ'ye göre daha etkin bir tahmin yöntemi olduğunu göstermektedir.

SKÖ'de örnek seçim işlemi gerçekleştirilirken, Y değişkenine göre sıralama işlemi, aynı birime ait farklı bir değişken olan X değişkeni kullanılarak yapılabilir. Bu durumda sıralama için kullanılan X değişkenine yardımcı değişken (concomitant variable) adı verilir. Yardımcı değişken ile ilgilenilen değişken arasındaki ilişki ilk olarak Stokes tarafından modellenerek, regresyon modeli yardımıyla yığın ortalaması tahmin edilmeye çalışılmıştır (Stokes, 1977). Ayrıca Yu ve Lam (1997) yardımcı değişkene ilişkin yığın ortalaması biliniyorken, ilgilenilen yığın ortalamasına ilişkin regresyon tahmin edicisini ele almışlardır. Demir ve Çıngı (2000) ise Yu ve Lam (1997)'in çalışmasını gerçek bir veri üzerinde incelemişlerdir.

Bu çalışmada, SKÖ'de örnek seçim işleminin yardımcı değişken kullanılarak yapılması durumunda yığın ortalamasının tahmini tanıtılmıştır. Bu amaçla ikinci bölümde, ilgilenilen değişken Y ile yardımcı değişken X arasındaki ilişki doğrusal regresyon modeli yardımıyla incelenerek, (Y,X)'in iki değişkenli normal dağılıma sahip olduğu varsayımı altında farklı m küme çapları ve ρ_{xy} korelasyon katsayıları için GE değerleri elde edilmiştir. Ayrıca üçüncü bölümde Forest Biometrics (Prodan, 1968)'de yer alan 295 ladin ağacının yükseklik ve gövde çapı verisi kullanılarak simülasyon çalışması ile n=4 ve 8 örnek çapını sağlayan mümkün (m,r) ikilileri için yığın ortalamasına ilişkin varyanslar hesaplanarak SKÖ ve BTÖ, GE bakımından karşılaştırılmıştır.

2. YARDIMCI DEĞİŞKENE GÖRE SIRALAMANIN MODELLENMESİ

SKÖ'nün uygulanmasında, ilgilenilen Y değişkeni bakımından birimleri sıralamak amacıyla kullanılabilecek önemli yöntemlerden biri de yardımcı değişkene göre sıralamadır. Yardımcı değişken, aynı birime ait farklı bir değişken olmak üzere; bu değişkenin Y değişkeni ile sıkı ilişkili, hassas ölçümünün kolay ve ihmal edilebilecek kadar maliyetli olması tercih edilir. Dolayısıyla, SKÖ'ye göre örnek seçme işleminin ilk adımında, birimler Y değişkeni yerine, yardımcı değişken X'e göre hassas ölçüm yapılarak sıralanır. İkinci adımda ise, X değişkeni bakımından sıralanan birimlerden, i'nci kümedeki i'nci sıradaki birim (i=1,2,...,m) Y değişkeni bakımından hassas ölçümle ölçülür. Bu işlem istenilen örnek çapını elde etmek üzere r kez tekrarlanabilir. Tekrarlar birbirinden bağımsız olup r=1 iken örnek seçiminin adımları Tablo 1'de gösterilmiştir. Böylece Y değişkenine göre ölçülecek birimler öncelikle X değişkenine göre sıralandığından, Y değişkeni için sıralama X değişkenine göre oluşacaktır. Bu durumda, X değişkenine ilişkin i'nci kümedeki i'nci sıra istatistiği $X_{(i)}$'ye karşılık gelen Y değişkeni, $Y_{[i]}$ ile ifade edilir ve $Y_{[i]}$, i'nci kümedeki i'nci indirgenmiş sıra istatistiği (induced order statistics) olarak isimlendirilir. X ve Y değişkenleri arasındaki korelasyon katsayısı $\rho_{xy}=1$ iken $Y_{[i]}=Y_{(i)}$ olacaktır.

Tablo 1. Yardımcı Değişkene Göre Sıralama Yapılması Durumunda m çaplı sıralı küme örneğinin oluşumu

	Küme				
Yığından seçilen örnek birimleri	1	(Y_{11}, X_{11})	(Y_{12}, X_{12})	$\cdots$	(Y_{1m}, X_{1m})
	2	(Y_{21}, X_{21})	(Y_{22}, X_{22})	$\cdots$	(Y_{2m}, X_{2m})
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	m	(Y_{m1}, X_{m1})	(Y_{m2}, X_{m2})	$\cdots$	(Y_{mm}, X_{mm})
			Birinci Adım $\Downarrow$		
	Küme				
Sıralanan örnek birimleri	1	$(Y^1_{[1]}, X^1_{(1)})$	$(Y^1_{[2]}, X^1_{(2)})$	$\cdots$	$(Y^1_{[m]}, X^1_{(m)})$
	2	$(Y^2_{[1]}, X^2_{(1)})$	$(Y^2_{[2]}, X^2_{(2)})$	$\cdots$	$(Y^2_{[m]}, X^2_{(m)})$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	m	$(Y^m_{[1]}, X^m_{(1)})$	$(Y^m_{[2]}, X^m_{(2)})$	$\cdots$	$(Y^m_{[m]}, X^m_{(m)})$
			İkinci Adım $\Downarrow$		
	Küme				
Örneğe alınan örnek birimleri (Sıralı Küme Örneği)	1	$(Y_{[1]}, X_{(1)})$	*	$\cdots$	*
	2	*	$(Y_{[2]}, X_{(2)})$	$\cdots$	*
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	*
	m	*	*	$\cdots$	$(Y_{[m]}, X_{(m)})$

Tablo 1'de $i = 1, 2, ..., m$ olmak üzere, $\{(Y_{i1}, X_{i1}), (Y_{i2}, X_{i2}), ..., (Y_{im}, X_{im})\}$, i'nci küme için yığından tesadüfi olarak seçilen örnek birimlerini, $\{(Y^i_{[1]}, X^i_{(1)}), (Y^i_{[2]}, X^i_{(2)}), ..., (Y^i_{[m]}, X^i_{(m)})\}$ i'nci küme için X değişkenine göre hassas ölçümle küçükten büyüğe doğru sıralanmış örnek birimlerini ve $\{(Y_{[i]}, X_{(i)})\}$, m çaplı sıralı küme örneğinde, i'nci kümedeki i'nci örnek birimini ifade eder.

X ve Y değişkenleri arasındaki ilişkinin doğrusal olduğu varsayımı altında, $X_{(i)}$ koşulu altında $Y_{[i]}$'nin beklenen değeri aşağıdaki doğrusal regresyon modeli yardımıyla modellenir (Stokes, 1977).

$$E\left(Y_{[i]} / X_{(i)}\right)=\mu_y+\rho_{xy}\frac{\sigma_y}{\sigma_x}\left(X_{(i)}-\mu_x\right) \qquad i=1,2,...,m \tag{4}$$

Bu modelin beklenen değerinin alınması ile i'nci kümedeki i'nci indirgenmiş sıra istatistiğinin beklenen değeri aşağıdaki gibi elde edilir:

$$E\left(E\left(Y_{[i]} / X_{(i)}\right)\right)=E\left(Y_{[i]}\right)=\mu_y+\rho_{xy}\frac{\sigma_y}{\sigma_x}\left(E(X_{(i)})-\mu_x\right) \tag{5}$$

Yardımcı değişken X'e göre sıralama yapılması durumunda, $\rho_{xy}=1$ olmadıkça, Y değişkeninin sıralamasında sıralama hatası ortaya çıkacaktır. Bu durum SKÖ'de hatalı sıralama (SKÖH) ile ifade edilmek üzere, yığın ortalamasına ilişkin tahmin edici $r>1$ iken,

$$\overline{Y}_{SKÖH}=\frac{1}{mr}\sum_{i=1}^{m}\sum_{j=1}^{r}Y_{[i]j} \tag{6}$$

şeklinde ifade edilir. Bu tahmin edicinin beklenen değeri,

$$E(\overline{Y}_{SKÖH})=\frac{1}{mr}\sum_{i=1}^{m}\sum_{j=1}^{r}E(Y_{[i]j})=\frac{1}{\mathrm{m}}\sum_{i=1}^{m}\left(\mu_y+\rho_{xy}\frac{\sigma_y}{\sigma_x}\left(E(X_{(i)})-\mu_x\right)\right)=\mu_y \tag{7}$$

olmak üzere, $\overline{Y}_{SKÖH}$ yığın ortalaması için sapmasız bir tahmin edicidir. Ayrıca bu tahmin edicinin varyansı, r tekrar birbirinden bağımsız olduğundan,

$$Var(\overline{Y}_{SKÖH})=\frac{1}{m^2r^2}\sum_{i=1}^{m}\sum_{j=1}^{r}Var(Y_{[i]j})=\frac{1}{m^2r}\sum_{i=1}^{m}Var(Y_{[i]}) \tag{8}$$

olacaktır. Burada $Var(Y_{[i]})$ ifadesi, Eşitlik (4)'den faydalanılarak,

$$\begin{aligned}Var(Y_{[i]}) &= E\left[Var\left(Y_{[i]} / X_{(i)}\right)\right]+Var\left[E\left(Y_{[i]} / X_{(i)}\right)\right] \\ &= \sigma_y^2(1-\rho_{xy}^2)+\rho_{xy}^2\frac{\sigma_y^2}{\sigma_x^2}Var(X_{(i)})\end{aligned} \tag{9}$$

şeklinde tanımlanır (Stokes, 1977). Böylece, $Var(\overline{Y}_{SKÖH})$ değişkenler arasındaki korelasyon katsayısı ρ_{xy} cinsinden Eşitlik (10)'daki gibi yazılabilir:

$$Var\left(\overline{Y}_{SKÖH}\right)=\frac{1}{m^2r}\sum_{i=1}^{m}\left(\sigma_y^2\left(1-\rho_{xy}^2\right)+\rho_{xy}^2\frac{\sigma_y^2}{\sigma_x^2}V\left(X_{(i)}\right)\right)$$

$$
=\frac{1}{m^2 r}\left[m\sigma_y^2\left(1-\rho_{xy}^2\right)+\rho_{xy}^2\frac{\sigma_y^2}{\sigma_x^2}\left(m\sigma_x^2-\sum_{i=1}^{m}\left(E(X_{(i)})-\mu_x\right)^2\right)\right]
$$

$$
=\frac{\sigma_y^2}{mr}-\frac{\rho_{xy}^2\sigma_y^2}{m^2 r\sigma_x^2}\sum_{i=1}^{m}\left(\mu_{i:m(x)}-\mu_x\right)^2 \tag{10}
$$

Dolayısıyla sıralama hatası durumunda SKÖH'nin BTÖ'ye *GE*'liği,

$$
GE_{SKÖH}=\frac{Var(\bar{Y}_{BTÖ})}{Var(\bar{Y}_{SKÖH})}=\frac{1}{1-\left\{\frac{\rho_{xy}^2}{m\sigma_x^2}\sum_{i=1}^{m}\left(\mu_{i:m(x)}-\mu_x\right)^2\right\}} \tag{11}
$$

şeklinde tanımlanır. Burada μ_x ve σ_x^2, X değişkenine ilişkin sırasıyla yığın ortalamasını ve varyansını, $\mu_{i:m(x)}$ ise $X_{(i)}$ sıra istatistiğine ilişkin yığın ortalamasını ifade eder.

(Y, X)'in iki değişkenli normal dağılıma sahip oldukları varsayımı altında, farklı m örnek çapları ve ρ_{xy} korelasyon katsayıları için Eşitlik (11)'den yararlanarak hesaplanan *GE* değerleri Tablo 2'de verilmektedir. *GE* değerleri hesaplanırken, X değişkeni için standart normal dağılıma ilişkin sıra istatistiklerinin beklenen değerlerinden yararlanılması gerekir. Bu amaçla, standart normal değişken $Z=\frac{X-\mu_x}{\sigma_x}$ şeklinde tanımlanmak üzere; Z değişkenine ilişkin i'nci sıra istatistiği

$$
Z_{(i)}=\frac{X_{(i)}-\mu_x}{\sigma_x} \tag{12}
$$

şeklinde ifade edilir. X değişkeni, ortalaması μ_x ve varyansı σ_x^2 olan normal dağılıma sahip olmak üzere, $Z_{(i)}$ sıra istatistiği, ortalaması 0 varyansı 1 olan standart normal dağılımdan çekilen n çaplı örnek için i'nci sıra istatistiğini gösterir. Böylece $Z_{(i)}$ sıra istatistiğinin beklenen değeri

$$
E(Z_{(i)})=\frac{E(X_{(i)})-\mu_x}{\sigma_x}=\frac{\mu_{i:m(x)}-\mu_x}{\sigma_x} \tag{13}
$$

olur. Eşitlik (13) kullanılarak, Eşitlik (11) yeniden düzenlenirse, $GE_{SKÖH}$ aşağıdaki gibi elde edilecektir:

$$
GE_{SKÖH}=\frac{Var(\bar{Y}_{BTÖ})}{Var(\bar{Y}_{SKÖH})}=\frac{1}{1-\left\{\frac{\rho_{xy}^2}{m}\sum_{i=1}^{m}E\left(Z_{(i)}\right)^2\right\}} \tag{14}
$$

Standart normal dağılım için i'nci sıra istatistiğinin beklenen değeri $E(Z_{(i)})$, m=2(1)10 olmak üzere, Arnold vd. (1993) tarafından tablo halinde verilmiştir. Tablo 2'deki $GE_{SKÖH}$ değerleri bu tablo değerlerinden yararlanarak hesaplanmıştır.

Tablo 2. (Y,X) iki değişkenli normal dağılıma sahip olmak üzere, farklı m ve $|\rho_{xy}|$ değerleri için $GE_{SKÖH}$ değerleri

$\lvert\rho_{xy}\rvert$	$GE_{SKÖH}$								
	m								
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.20	1.01	1.02	1.02	1.03	1.03	1.03	1.03	1.03	1.03
0.30	1.03	1.04	1.05	1.06	1.07	1.07	1.07	1.07	1.08
0.40	1.05	1.08	1.10	1.11	1.12	1.13	1.14	1.14	1.15
0.50	1.09	1.14	1.17	1.19	1.21	1.22	1.23	1.24	1.25
0.60	1.13	1.21	1.26	1.30	1.33	1.35	1.37	1.39	1.40
0.70	1.18	1.31	1.39	1.46	1.51	1.55	1.58	1.61	1.63
0.80	1.26	1.44	1.58	1.69	1.78	1.86	1.92	1.98	2.03
0.90	1.35	1.63	1.87	2.07	2.25	2.41	2.55	2.67	2.79
1.00	1.47	1.91	2.35	2.77	3.19	3.60	4.00	4.40	4.79

Tablo 2'den görüldüğü gibi;

- m örnek çapı ve $|\rho_{xy}|$ korelasyon katsayısı arttıkça, $GE_{SKÖH}$ değerleri de artmaktadır.
- $|\rho_{xy}| = 0$ iken, SKÖ ile elde edilen örneğin basit tesadüfi bir örnekten farkı yoktur. Yani $Var(\overline{Y}_{BTÖ}) = Var(\overline{Y}_{SKÖ})$ olup tüm m değerleri için $GE_{SKÖH} = 1$ olmaktadır.
- $|\rho_{xy}| = 1$ iken $Var(\overline{Y}_{SKÖH}) = Var(\overline{Y}_{SKÖ})$ özelliğinden dolayı, $GE_{SKÖH}$ değerleri tüm m değerleri için en yüksek değerini almaktadır.
- $|\rho_{xy}| < 0.50$ iken incelenen tüm m değerleri için GE, 1'e yakın değerler alırken, özellikle $|\rho_{xy}| > 0.70$ iken, m arttıkça $GE_{SKÖH}$ değerlerinin diğer korelasyon durumlarına oranla daha fazla arttığı görülmektedir.

3. UYGULAMA

Uygulamada SKÖ'de yardımcı değişken kullanılarak yığın ortalamasının tahminini incelemek amacıyla Tablo 3'de verilen Almanya'nın Kattenbühl bölgesindeki 295 ladin ağacının yükseklik ve gövde çapı verisi kullanılmıştır (Prodan,1968). Özdemir (2005) aynı veriden yararlanarak n=6 (m=3, r=2) için ortalama ağaç yüksekliğini tahmin etmiştir. Ancak bu çalışmada, örnek çapları n=4 ve 8'e genişletilerek örnek çapını sağlayan farklı m ve r değerleri ele alınmıştır. Burada problemin, bölgede bulunan ladin ağaçlarına ilişkin ortalama ağaç yüksekliğini, örnek çapları n=4 ve 8 için tahmin etmek olduğunu varsayalım. 295 ladin ağacının ortalama yüksekliğini tahmin etmek için kullanılacak bir örnekleme yöntemi ile örneğe çıkan ağaçların her biri üzerinde, yükseklik ölçümü yapılması gereklidir. Ancak bu ölçümün yapabilmesi, ağacın kesilmesi veya maliyetli ölçme tekniklerinin kullanılması ile mümkündür. Ağaçları keserek ölçüm yapmak ya da yüksek maliyetli ölçme teknikleri kullanmak yerine, yükseklik değişkeni ile sıkı ilişkili, ancak daha düşük maliyetle ve daha kolay ölçülebilecek bir değişken yardımıyla, ağaçların yükseklikleri sıralanabilir. Orman araştırmalarında ağaç gövde çapı, yükseklik ve hacim değişkenleri arasında fonksiyonel bir ilişki olduğu bilinmektedir(Prodan, 1968). Özellikle gövde çapı ve yükseklik değişkenleri arasında pozitif yönde sıkı bir korelasyon vardır. Bu nedenle, ağaçların yükseklikleri, ağaç gövde çapı ölçümlerine göre sıralanabilir. Böylece SKÖ yöntemi ile, örneğe çıkan ağaçların tümü, gövde çapı değişkeni (X) bakımından ölçülerek sıralanır ve sıralanan bu ağaçlardan seçilecek ağaçlar üzerinde yükseklik değişkeni (Y) bakımından maliyetli hassas ölçüm gerçekleştirilir. Bu durumda, yükseklik değerlerinin sıralanmasında, gövde çapı değişkeni yardımcı değişken olacak ve örneğe çıkan ağaçların yalnızca bir kısmı üzerinde yükseklik değişkeni ölçülecektir.

Tablo 3. 295 ladin ağacının gövde çapı ve yüksekliklerine göre dağılımı

Yükseklik (m)	Gövde Çapı (cm) 8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Toplam
26												3	1	1		5
25								1	4	9	12	11	3	2	1	43
24						1	1	14	23	23	19	7				88
23					1	10	24	25	20	5	3	2				90
22				1	7	13	12	7	5		1					46
21			1	2	6	3	1	1								14
20			2	2	3											7
19		1														1
18																-
17	1															1
Toplam	1	1	3	5	17	27	38	48	52	37	35	23	4	3	1	295

Tablo 3'deki 295 birimli yığının Y değişkeni bakımından ortalaması $\mu_Y = 23.28$ ve X ve Y değişkenleri arasındaki korelasyon değeri $\rho_{xy} = 0,8130$'dur. Yığının dağılımının iki değişkenli normal dağılıma uyup uymadığını araştırmak üzere yapılan Kolmogorov Simirnov uyum iyiliği testi ile yığının iki değişkenli normal dağılıma uymadığı ortaya çıkmıştır. Buna göre SKÖ, SKÖH ve BTÖ'ye göre yığın ortalamasına ilişkin tahmin edicileri karşılaştırmada kullanılacak en uygun yol simülasyon çalışması ile tahmin edicilerin varyanslarını elde ederek, *GE* değerleri bakımından karşılaştırmaktır. Bu amaçla, simülasyon çalışması ile n=4 ve 8 olmak üzere, bu örnek çaplarını sağlayan mümkün (m,r) ikilileri (n=4 iken (m,r)=(2,2), n=8 iken (m,r)=(2,4), (4,2)) için toplam 3 durumda, SKÖ, SKÖH ve BTÖ'ye göre yığın ortalamasına ilişkin tahmin edicilerin beklenen değerleri ve varyansları hesaplanmıştır. Simülasyon çalışmasında Matlab 6.5 paket programı kullanılarak aşağıdaki adımlar 3 durum için ayrı ayrı uygulanmıştır:

Adım 1: SKÖ'ye göre belirlenen n örnek çapını sağlayan (m,r) ikilisi için, 295 çaplı yığından r kez m^2 çaplı tesadüfi bir örnek yerine koyarak örnekleme yöntemi ile seçilerek, yine tesadüfi olarak, her bir tekrar için m çaplı m küme oluşturulmuştur.

Adım 2: Sıralama hatası olmadığı duruma karşılık gelmek üzere, ağaçların yüksekliklerine ilişkin Y değişken değerlerine göre sıralama yapılarak, her bir tekrardan m çaplı örnek elde edilmek üzere n=mr çaplı sıralı küme örneği elde edilmiştir.

Adım 3: Sıralama hatasının olduğu duruma karşılık gelmek üzere, ağaçların gövde çapına ilişkin X değişken değerlerine göre sıralama yapılarak, her bir tekrardan m çaplı örnek elde edilmek üzere n=mr çaplı sıralı küme örneği elde edilmiştir.

Adım 4: BTÖ'ye göre n=mr çaplı tesadüfi bir örnek yerine koyarak örnekleme yöntemi ile seçilmiştir.

Adım 5: 2'nci, 3'üncü ve 4'üncü adımda elde edilen örneklerden, sırasıyla yığın ortalamasına ilişkin tahmin değerleri hesaplanmıştır.

Adım 6: 1-5 arası adımlar 100000 kez tekrarlanarak, 100000 tekrar sonunda elde edilen tahmin değerlerinin ortalaması (tahmin edicinin beklenen değeri) ve varyansı bulunmuştur. Varyans değerlerinden yararlanılarak SKÖ'nün BTÖ'ye ve SKÖH'ın BTÖ'ye GE'liği hesaplanmıştır. Elde edilen sonuçlar Tablo 4'de verilmiştir.

Tablo 4. Simülasyon çalışmasından elde edilen beklenen değer, varyans ve *GE* değerleri

	n=4	n=8	
	m=2 r=2	m=2 r=4	m=4 r=2
$E(\bar{Y}_{BTÖ})$	23,29	23,29	23,28
$E(\bar{Y}_{SKÖ})$	23,28	23,28	23,29
$E(\bar{Y}_{SKÖH})$	23,28	23,28	23,29
$Var(\bar{Y}_{BTÖ})$	0,42	0,21	0,21
$Var(\bar{Y}_{SKÖ})$	0,30	0,15	0,10

$Var(\bar{Y}_{SKÖH})$	0,34	0,17	0,14
$GE_{SKÖ}$	1,41	1,41	2,09
$GE_{SKÖH}$	1,25	1,25	1,57

Tablo 4'den görüldüğü gibi; BTÖ, SKÖ ve SKÖH' ye göre yığın ortalamasına ilişkin elde edilen tahmin edicilerin beklenen değerleri yığın ortalaması $\mu_Y = 23.28$'e yakın değerler verdiğinden tahmin ediciler sapmasızdır. Varyans değerlerine bakıldığında ise, aynı (m,r) değeri için SKÖH'a göre bulunan varyans değeri, SKÖ'ye göre bulunan varyans değerinden daha büyüktür. Ancak hem SKÖ hem de SKÖH'a göre bulunan varyans değerleri aynı (m,r) değeri için BTÖ'den daha küçük değer almaktadırlar. Bunun yanında, aynı (m,r) değeri için sıralama hatasının yapıldığı durumdaki $GE_{SKÖH}$ değerleri, sıralama hatasının yapılmadığı durumdaki $GE_{SKÖ}$ değerlerinden daha küçüktür. Ancak $\rho_{xy} = 0,8130$ olduğundan, $GE_{SKÖH}$ değerleri de 1'den büyük çıkmaktadır. Örnek çapları bakımından GE değerlerine bakıldığında ise, aynı m küme çapına sahip (m,r) çiftleri için ((m,r)=(2,2), (2,4)) $GE_{SKÖ}$ ve $GE_{SKÖH}$ aynı değerleri almaktadır. Dolayısıyla r tekrar sayısının $GE_{SKÖ}$ ve $GE_{SKÖH}$ üzerinde etkili olmadığını gösteren Eşitlik (11) ve (14)'deki sonuçlar, bu uygulama ile de desteklenmektedir. Ayrıca en yüksek $GE_{SKÖ}$ ve $GE_{SKÖH}$ değerleri, n=4 için m=4 ve r=2 iken gerçekleşmektedir.

4. SONUÇ

SKÖ'de ilgilenilen değişkeni sıralamak için kullanılacak yardımcı değişken X ile ilgilenilen değişken Y arasındaki korelasyon katsayısı ρ_{xy} mutlak değerce ne denli yüksek ise, SKÖ altında yığın ortalamasının tahminine ilişkin varyans da o denli küçük çıkacaktır. Bu durum Eşitlik (11)'den de görülmektedir. Sonuç olarak, uygulamada, ilgilenilen değişkenle yüksek derecede korelasyonlu, aynı birime ait bir başka değişkenle sıralama yapılması veya hatasız sıralama yapmayı sağlayacak teknikler kullanılması durumunda, $GE_{SKÖH}$ yüksek değerler alacaktır. Bunun yanında, teorik değerlerin yer aldığı Tablo 2'den ve uygulama sonuçlarının yer aldığı Tablo 4'den görüldüğü gibi, *n* örnek çapı ne olursa olsun, *m* küme çapı aynı olan *n* örnek çapları için $GE_{SKÖ}$ ve $GE_{SKÖH}$ değerleri birbirine eşit olacaktır. Dolayısıyla, yığın ortalamasını tahmin etmede SKÖ'nün ve SKÖH'nin BTÖ'ye göreli etkinliği *m* küme çapı arttıkça artmakta, *r* tekrar sayısının ise $GE_{SKÖ}$ ve $GE_{SKÖH}$ üzerinde herhangi bir etkisi olmamaktadır. Bu anlamda, uygulamada belirlenen *n* örnek çapı için, en büyük *m* küme çapının seçilmesi durumunda, $GE_{SKÖ}$ ve $GE_{SKÖH}$ en yüksek değerini alacaktır.

KAYNAKLAR

ARNOLD, B., BALAKRISHNAN, N., AND NAGARAJA, H.N.(1993), *A First Course in Order Statistics,* John Wiley, New York, 1-107.

DAVID, H.A., LEVINE, D.N.(1972), *Ranked set sampling in the presence of judgment error,* Biometrics, 28, 553-555.

DELL, D.R., CLUTTER, J.L.(1972), *Ranked set sampling theory with order statistics background*, Biometrics, 28, 545-555.

DEMİR, S., ÇINGI, H., (2000), *An application of the regression estimator in ranked set sampling*, Hacettepe Bulletin of Natural Sciences and Engineering, Series (B) , 29, 93-101.

MCINTYRE, G.A. (1952), *A method of unbiased selective sampling using ranked sets*, Australian Journal of Agricaltural Research, 3, 385-390.

ÖZDEMİR, Y. A. (2005), *Sıralı Küme Örneklemesiyle Doğrusal Regresyon Modelinde Parametre Tahminlerinin İncelenmesi,* Doktora Tezi, Ankara.

PATIL,G.P., SINHA, A.K. AND TAILLIE,C.(1994), *Ranked set sampling*, A Handbook of Statistics, Elsevier Science B.V., vol.12, pp 167-200.

PRODAN, M.(1968), *Forest Biometrics,* Pergamon Pres Ltd. Oxford, pp.45.

RIDOUT, M. S. COBY, J.M.(1987), *Ranked set sampling with non-random selection of sets and errors in ranking*, Applied Statistics, 36, 145-152.

STOKES, L.S.(1977), *Ranked set sampling with concomitant variables*, Commun. Statist. Theor. Meth., A6(12), 1207-1211.

YU, P.L.H., LAM, K. (1997), *Regression estimator in ranked set sampling,* Biometrics, 53, 1070-1080.

ESTIMATION OF THE POPULATION MEAN USING CONCOMITANT VARIABLE IN RANKED SET SAMPLING

ABSTRACT

Ranked set sampling is frequently used sampling technique where the measurements of sampling units according to the variable under consideration is expensive in all sense, but if it is possible to rank sampling units according to the same variable by means of a method which is not expensive at all. The areas such as environment, agriculture and medicine ranked set sampling are more efficient than the simple random sampling as a sampling method in the sense of estimation the population mean. In this study, estimation the population mean is considered in the case of the units ranking by concomitant variable in ranked set sampling. Moreover using the data about height and diameter of spruce trees ranked set sampling compared with the simple random sampling in the sense of relative efficiency for different sample size by a simulation study.

Key Words: *Ranked set sampling, Simple random sampling, Concomitant variable, Relative efficiency, Lineer regression model*