

UYARLANABİLİR KÜME ÖRNEKLEMESİNDE ÖN ÖRNEKLEM HACMİNE GÖRE ETKİNLİK

Olca BOZKURT*

Şanslı ŞENOL**

ÖZET

Bu çalışmada uyarlanabilir küme örneklemesinin (UKÖ) ön örneklem hacmine bağlı olarak rasgele örnekleme (RÖ) göre etkinliği araştırılmıştır. Bu amaçla farklı boyutlardaki çalışma alanlarında, farklı ön örneklem hacimleri ile benzetim çalışmaları hazırlanmış ve sonuçlar sunulmuştur.

Anahtar Kelimeler: Rasgele örnekleme, Göreceli etkinlik, Uyarlanabilir küme örnekleme.

1. GİRİŞ

Uyarlanabilir Küme Örnekleme (UKÖ) ilk kez 1990 yılında Steven K. Thompson (1990, 1992) tarafından tanıtılmıştır. Bu yöntem özellikle birimlerin bir arada olma yani kümelenme eğilimine sahip olduğu kitlelerde, kitle karakteristiği ve kitlenin büyüklüğü hakkında önemli bilgileri tahmin etmek için kullanılan alternatif bir örnekleme yöntemidir. Bu tipteki kitleler hakkında tahminde bulunurken UKÖ'nin kullanılması geleneksel örnekleme yöntemlerine nazaran göreceli etkinlik sağlamaktadır (Connors, 2002).

Yöntemin işleyişi özetlenecek olursa; ön örneklem hacmine (n_0) göre birimler Rasgele Örnekleme (RÖ) (iadedeli veya iadesiz seçim) ile seçilir ve çalışmada ilgilenilen birimler için kritik bir eşik değeri olarak ön görülen koşulu sağlayıp sağlamadığının kontrol edilmesi ile süreç başlar. Seçilen birimlerin verilen koşulu sağlayıp sağlamadığı kontrol edilir ve koşulu sağlayan birimlerin komşuluğundaki birimlerde örnekleme alınır ve koşulu sağlamayan birim seçilinceye kadar süreç devam eder (Thompson ve Collins, 2002).

Örnekleme yöntemlerinde, örneklem hacminin önemli bir rolü vardır. Uygun bir örneklem hacmi kitle hakkında son derece değerli bilgiler verir (Baskan, 1998). Temel örnekleme yöntemlerinde örneklem hacminin belirlenmesine ilişkin kriterler olmasına rağmen UKÖ çalışmalarında araştırmacılar ön örneklem hacmini keyfi olarak seçerek çalışmalarını sürdürmektedirler (Consigilo,2000 ve Khan, 2002). Bu nedenle UKÖ'nde ön örneklem hacminin belirlenmesi yapılan çalışma için son derece önemlidir. Zaman ve para kaybını önlemek amacıyla araştırmacı, ne büyüklükte bir ön örneklem hacmi ile başlayacağına karar vermeli ve yöntemi uygulamalıdır.

Hazırlanan bu çalışmada UKÖ'nin RÖ'ye göre daha düşük varyansa sahip olmasını ve kitledeki toplam birim sayısını gerçeğe yakın tahmin edebilmesini sağlamak amacıyla seçilebilecek ön örneklem hacmi hakkında öngörüle bulunabilmek amaçlanmıştır.

* Dr., Ege Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: olcaybozkurt@gmail.com

** Prof. Dr., Ege Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: sansli.senol@ege.edu.tr

2. YÖNTEM

Uyarlanmamış yöntemler için klasik tahmin ediciler örneğin; örneklem ortalaması, kitle ortalamasının yansız tahmin edicisidir. Fakat uyarlanabilir yöntemler kullanıldığında bu tahmin ediciler yansız değildirlir. Bu nedenle UKÖ için geliştirilmiş başka tahmin ediciler vardır. Bunlardan en önemlileri ise değiştirilmiş Hansen-Hurwitz (HH) ve değiştirilmiş Horvitz-Thompson (HT) ile geliştirilmiş tahmin edicilerdir. Bazı durumlarda HT tahmin edicisi varyansı negatif hesaplamaktadır bu nedenle, Thompson 1990'da da belirttiği gibi hangi koşullar altında varyans tahmin edicilerinin negatif değer aldıklarına dair araştırma yapılması gerekmektedir. Çalışmanın uygulama aşamasında HT tahmin edicisi varyansı negatif hesapladığı için sadece değiştirilmiş HH tahmin edicisi kullanılmıştır.

n adet birimin iadeli çekildiği ve i .nci birimin çekim olasılığı olan p_i 'nin, her çekim sırasında ve tüm birimler için bilindiği planlarda, her bir y -değerinin ilgili seçim olasılığına bölüdüğü ve birimin seçilme sırasıyla çarpılarak hesaplandığı HH tahmin edicisi, kitle toplamı için yansız bir tahmin edicidir.

ψ_k , k birimini içeren ağı, m_k bu ağıdaki birimlerin sayısını, \bar{y}_k^* ise, ön örneklemin k .nci birimini içeren ağıdaki gözlemlerin ortalamalarını ifade etmek üzere;

$$\bar{y}_k^* = \frac{1}{m_k} \sum_{j \in \psi_k} y_j \quad (1)$$

kitle ortalaması için yansız tahmin edici ise

$$t_{HH}^* = \frac{1}{n} \sum_{k=1}^{n_1} \bar{y}_k^* \quad (2)$$

olur.

t_{HH}^* 'in varyansı eğer ön örnekleme iadesiz seçilmişse:

$$\text{var}(t_{HH}^*) = \left(\frac{N-n}{Nn} \right) \sum_{i=1}^N \frac{(\bar{y}_i^* - \mu)^2}{(N-1)} \quad (3)$$

olur. Tahmin edicisi ise;

$$\hat{\text{var}}(t_{HH}^*) = \left(\frac{N-n}{Nn} \right) \sum_{k=1}^{n_1} \frac{(\bar{y}_k^* - t_{HH}^*)^2}{(n-1)} \quad (4)$$

olacaktır. Eğer iadeli seçilmişse varyansı;

$$\text{var}(t_{HH}^*) = \left(\frac{1}{n} \right) \sum_{i=1}^N \frac{(\bar{y}_i^* - \mu)^2}{N} \quad (5)$$

tahmin edicisi ise;

$$\hat{\text{var}}(t_{HH}^*) = \left(\frac{1}{n} \right) \sum_{k=1}^{n_1} \frac{(\bar{y}_k^* - t_{HH}^*)^2}{(n-1)} \quad (6)$$

olur (Thompson, 1990).

3. UYGULAMA

Uygulama için C++ programlama dilinde bir benzetim yazılım programı hazırlandı. Bu program RÖ ve UKÖ (HH tahmin edicisi) için varyans tahminini ve toplam birim sayısını hesaplamıştır. Türetilerek elde edilen ve kitleyi oluşturulan birimler (10×10), (20×20), (30×30), (40×40) ve (50×50) birimlik karesel çalışma alanlarına Diggle'in Poisson Kümeleme Süreci (Diggle, 2003) kullanılarak ana küme ve alt kümeler oluşturacak şekilde Poisson dağılımı ile dağıtıldı.

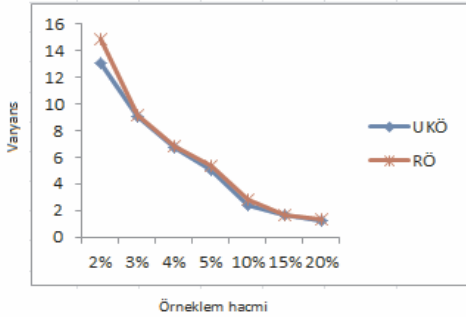
Diğer taraftan tahmin edicileri hesaplamak için ön örneklem hacimleri alanların yüzde 2, 3, 4, 5, 10, 15 ve 20'si olarak seçildi ve program iadeli ve iadesiz seçim yöntemi kullanılarak her bir n_0 değeri için 500 kez tekrarlandı. Her bir alanda ön örneklem hacimlerine bağlı olarak hesaplanan göreceli etkinlik-GE (7) değerleri Tablo 1'de ve varyans değişimleri ise Şekil 1-5'te gösterilmiştir.

$$\hat{GE} = \frac{\hat{\text{var}}(UKÖ)}{\hat{\text{var}}(RÖ)} \quad (7)$$

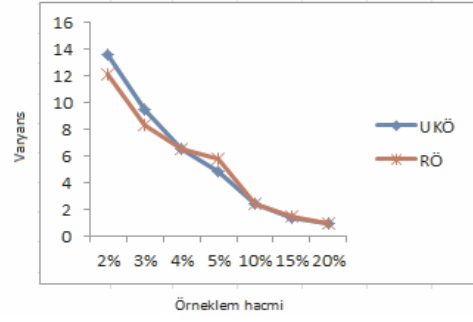
Tablo 1. Varyansların göreceli etkinlikleri

	Yöntem	Örneklem Hacmi						
		%2	%3	%4	%5	%10	%15	%20
10×10	İadeli	0,88	0,99	0,98	0,93	0,86	1,03	0,93
	İadesiz	1,12	1,15	1,00	0,84	1,00	0,89	0,96
20×20	İadeli	1,00	0,98	1,01	0,94	0,99	0,98	0,99
	İadesiz	0,91	0,90	0,86	0,84	0,90	0,80	0,72
30×30	İadeli	0,99	1,02	0,99	0,98	0,98	1,02	1,00
	İadesiz	0,96	0,93	0,91	0,85	0,73	0,64	0,57
40×40	İadeli	0,99	1,01	0,99	0,98	1,00	0,99	1,00
	İadesiz	0,91	0,91	0,85	0,80	0,68	0,56	0,48
50×50	İadeli	1,00	0,98	0,97	1,00	0,99	1,00	1,00
	İadesiz	0,91	0,88	0,82	0,79	0,65	0,54	0,46

Tablo 1 incelendiğinde, çalışma alanlarının çoğunluğunda UKÖ, RÖ'den daha düşük varyans değerlerine sahiptir. Böylece UKÖ'nin daha etkin bir tahmin edici olduğu görülmektedir. İadeli seçim yapılması durumunda UKÖ ve RÖ varyans tahmin edicileri neredeyse birbirine eşittir. Ancak iadesiz seçim yapıldığına kitledeki birim sayısı ve çalışma alanı büyüdükçe UKÖ'nin etkinliği RÖ'den daha iyi olmakta ve RÖ'ye göre tercih edilebilir konuma gelmektedir.

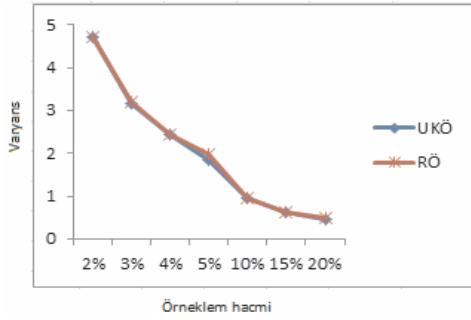


(1)

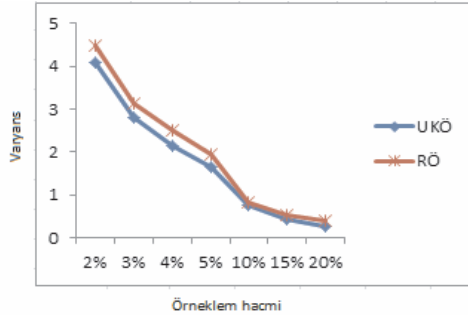


(2)

Şekil 1. 10×10 Alandaki varyans değişimleri

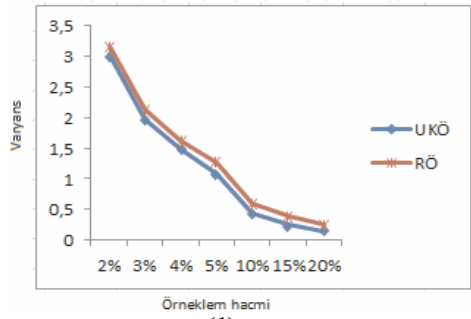


(1)

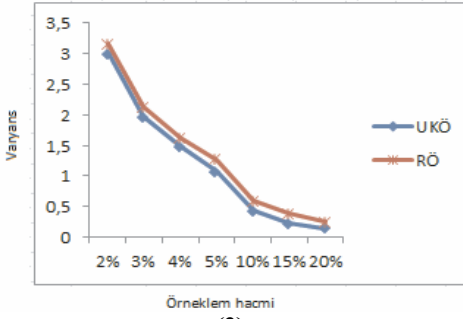


(2)

Şekil 2. 20×20 Alandaki varyans değişimleri

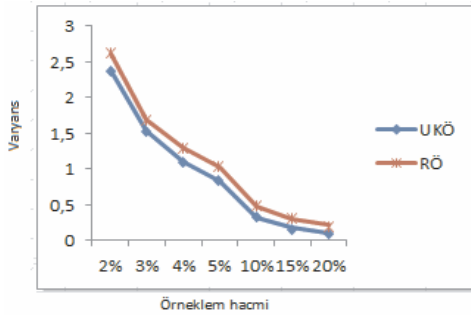


(1)

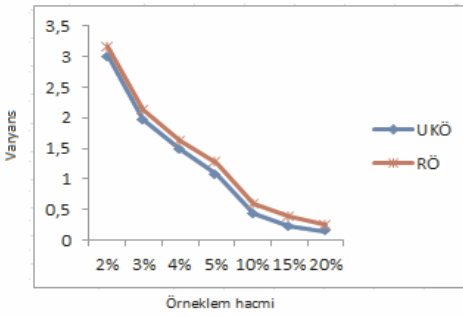


(2)

Şekil 3. 30×30 Alandaki varyans değişimleri

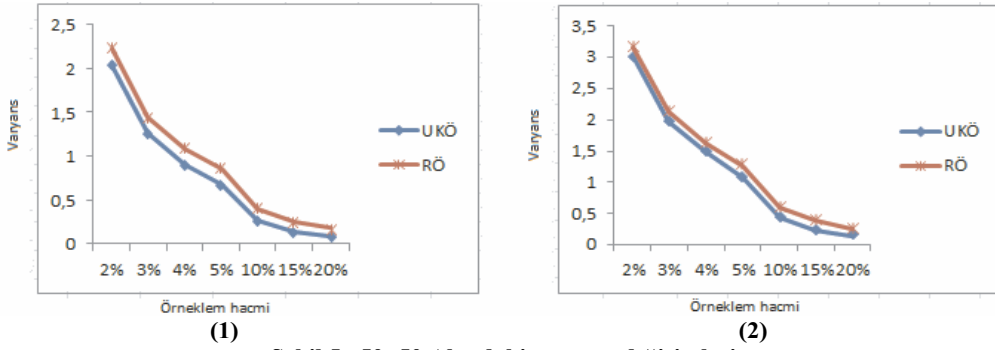


(1)



(2)

Şekil 4. 40×40 Alandaki varyans değişimleri



Şekil 5. 50×50 Alandaki varyans değişimleri
(1) iadelı seçım (2) iadesız seçım

Şekillerden de görüldüğü üzere kitle büyüklüğünün artmasıyla varyans değeri azalmaktadır. %10'luk ön örneklem hacmi değerinde varyans değişimi için bir kırılma söz konusu olup bu noktadan sonra varyansta önemli bir değişim görülmemektedir. Benzetim çalışmasının sonucunda 10×10'luk alanda iadesız seçım dışında UKÖ tüm alanlarda RÖ'den daha küçük varyans değerleri vermiştir. Tablo 2'de ise hesaplanan tahmini toplam birim sayıları verilmektedir.

Tablo 2. Alanlardaki tahmin edilen toplam birim sayıları

	Yöntem	Örneklem Hacmi						
		%2	%3	%4	%5	%10	%15	%20
10×10 (166*)	UKÖ (1)	148	176	179	155	154	163	163
	RÖ (1)	182	172	172	168	177	154	170
	UKÖ (2)	165	184	170	162	165	146	147
	RÖ (2)	160	160	172	192	168	168	155
20×20 (806*)	UKÖ (1)	804	812	836	782	813	807	806
	RÖ (1)	787	797	837	823	817	814	812
	UKÖ (2)	681	737	755	718	696	634	577
	RÖ (2)	789	816	874	865	777	795	804
30×30 (2273*)	UKÖ (1)	2289	2308	2291	2193	2248	2314	2272
	RÖ (1)	2326	2254	2278	2227	2294	2244	2266
	UKÖ (2)	2107	2073	2116	1927	1652	1448	1260
	RÖ (2)	2202	2266	2318	2300	2301	2311	2273
40×40 (4819*)	UKÖ (1)	4837	4913	4824	4756	4785	4833	4819
	RÖ (1)	4848	4811	4814	4816	4726	4849	4819
	UKÖ (2)	4339	4237	4098	3892	3206	2645	2206
	RÖ (2)	4805	4706	4844	4888	4807	4868	4784
50×50 (8688*)	UKÖ (1)	8775	8458	8630	8780	8591	8730	8706
	RÖ (1)	8675	8633	8920	8720	8664	8672	8710
	UKÖ (2)	8046	7516	7158	6857	5557	4610	3947
	RÖ (2)	8875	8587	8804	8780	8697	8660	8702

* Alanlardaki gerçek toplam birim sayıları

(1) iadelı seçım (2) iadesız seçım

Tablo 2'deki değerler incelendiğinde tüm çalışma alanları için UKÖ (iadelı seçım) ve RÖ'nin birbirine ve gerçek değere çok yakın tahmin değerleri elde ettikleri görülmektedir. UKÖ'nde iadesız seçım ile elde edilen toplam birim sayılarına bakıldığında ise 10×10'luk alanda %10'dan sonra, 20×20'lik alanda %4, 30×30'luk alanda %4, 40×40'lık alanda %2 ve 50×50'lik alanda %2'den sonra HH tahmin edicisi gerçek toplam birim sayılarından oldukça uzak değerler tahmin etmektedir. Bu nedenle

UKÖ’nde iadesiz seçim yapılacak ise çalışma alanı büyüdükçe bu alanın çok küçük oranında bir ön örneklem hacmi ile çalışmaya başlandığı takdirde hem RÖ’ye göre tahmin edici etkinliği hem de tahmin edilen toplam birim sayısı gerçeğe yakın tahmin edilebilir.

4. TARTIŞMA VE SONUÇ

Sonuç olarak, bu çalışmada elde edilen sonuçlar nadir ve bir arada görülme eğilimindeki kitleler için elde edilen sonuçlardır. Burada 5 farklı alandan 7 farklı ön örneklem hacmine bağlı olarak UKÖ ve RÖ örneklemede iadeli ve iadesiz seçim yöntemleri ile varyans değerleri ve kitledeki toplam birim sayıları tahmin edilmiştir. Araştırmacının UKÖ’nde iadesiz seçim yöntemini kullanarak alandan birim seçecek olması durumunda, ön örneklem hacmini çalışma alanının çok küçük bir oranında seçerek çalışması hem küçük bir varyansa hem de gerçek toplam birim sayısını elde etmesine olanak sağlayabilecektir. Gerek tablolardan gerek ise şekillerden de görüldüğü üzere UKÖ’nde iadeli seçim RÖ’de iadeli seçim ile birbirine oldukça yakın sonuçlar vermektedir. Bu nedenle eğer araştırmacı iadeli bir seçim yöntemi kullanmayı tercih ediyor ise maliyet, zaman ve kolaylık açısından uygun olan yöntemi tercih edebilir.

5. KAYNAKLAR

- Baskan, S., 1998. Araştırma Yöntemleri ve Örnekleme Giriş, İzmir.
- Connors, M.E. ve Schwager, S.J., 2002. The Use of Adaptive Cluster Sampling for Hydroacoustic Surveys, ICES Journal of Marine Science, 59, 1314-1325.
- Consigilo, L.D. ve Scanu, M., 2000. Some Results on Asymptotic in Adaptive Cluster Sampling, Statistics and Probability Letters, 52, 189-197.
- Diggle, P. J., 2003. Statistical Analysis of Spatial Point Patterns, Arnold.
- Khan, A. ve Muttlak, H. A., 2002. Adjusted Two-Stage Cluster Sampling, Applied Mathematics and Computation, 126, 83-95.
- Thompson, S. K., 1990. Adaptive Cluster Sampling, Journal of the American Statistical Association, 85, 1050-1059.
- Thompson, S. K., 1992. Sampling, John Wiley & Sons, Inc.
- Thompson, S. K., Collins, L. M., 2002. Adaptive Sampling in Research on Risk-Related Behaviors, Drug and Alcohol Dependence, 68, 57-67.

EFFICIENCY IN ADAPTIVE CLUSTER SAMPLING ACCORDING TO INITIAL SAMPLE VOLUME

ABSTRACT

In this study, the efficiency of adaptive cluster sampling (ACS) with respect to random sampling (RS) is investigated in terms of the initial sample volume. For this purpose, simulation studies are conducted under different sizes and different initial sample volumes and results are presented.

Keywords: Simple random sampling, Relative efficiency, Adaptive cluster sampling.