Süleyman Demirel Üniversitesi
Fen Bilimleri Enstitüsü Dergisi

Suleyman Demirel University
Journal of Natural and Applied Science

17(3), 9-17, 2013

Çok Aşamalı Örnekleme Yöntemlerinde Örneklem Büyüklüğünün Belirlenmesi : Bir Uygulama

Sevil BACANLI*[1], Pınar UÇAR[2]
[1]Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü-Beytepe/ANKARA
[2]Katar İstatistik Kurumu, Ekonomi İstatistikleri Bölümü-Doha/Katar

(Alınış Tarihi: 06.03.2013, Kabul Tarihi: 10.12.2013)

Anahtar Kelimeler
İki aşamalı örnekleme
Küme örneklemesi
Üç aşamalı örnekleme
Örneklem büyüklüğü.

Özet: Bu çalışmada, iki aşamalı, iki aşamalı küme ve üç aşamalı örnekleme yöntemleri hakkında kısaca bilgi verilmiş ve yöntemlere ilişkin örneklem büyüklüğünün belirlenmesi amaçlanmıştır. Türkiye'nin 7 coğrafi bölge, 78 il ve toplam 828 ilçesinden elde edilen 1997 yılı buğday üretim miktarlarına ait veriler kullanılarak basit rasgele, tabakalı rasgele, iki aşamalı, iki aşamalı küme ve üç aşamalı örnekleme yöntemleri için uygulama yapılmıştır. Uygulama sonucuna göre, tabakalı rasgele örnekleme ile daha duyarlı tahmin sonuçları elde edilmesine karşın, aşamalı örnekleme yöntemleri kullanıldığında daha az örneklem birimi ile çalışıldığı için; emek, maliyet ve zaman açısından avantaj sağlandığı söylenebilir .

Sample Size Determination in Multi-Stage Sampling Methods: A Application

Keywords
Two stage sampling
Cluster sampling
Three stage sampling
Sample size.

Abstract: In this study, two stage, two stage cluster, three stage sampling are mentioned briefly and determination of sample size in these techniques are examined. Applications of simple random, stratified random, two stage, two stage cluster and three stage sampling were utilized on data of wheat production amounts from 828 towns of 78 cities of 7 geographical regions of Turkey in 1997. According to this application, although more precise estimation results were obtained by stratified random sampling, in can be stated that multistage sampling methods are advantageous of effort, cost, and time due to lower sampling units.

1.Giriş

Örnekleme yöntemleri, örneklem birimlerinin örnekleme alınma aşamasına göre tek ve çok aşamalı yöntemler olmak üzere iki gruba ayrılmaktadır. Çok aşamalı örnekleme yöntemlerinde, kitle birimlerinin örnekleme alınması birden çok aşamada gerçekleşmektedir (Singh, 2003).

Çok aşamalı örnekleme yöntemi, ön birimlere ilişkin bir çerçevenin bulunması durumunda yaygın olarak kullanılmaktadır. Çok aşamalı örnekleme yönteminin istatistik olarak etkinliği, basit rasgele örneklemeye göre daima daha düşüktür. Ancak, kümelerdeki örneklem birimleri arasında kısmi korelasyon bulunmaktadır ve basit rasgele örneklemede olduğu gibi bunlar birbirinden bağımsız kabul edilmezler. Dolayısıyla, kümelerdeki birimlerin maliyetinin düşmesi, istatistik etkinliğin düşmesine karşılık geliyorsa, çok aşamalı örnekleme yatırılan birim maliyet başına daha çok doğruluk sağlamaktadır.

Çok aşamalı örnekleme, özellikle homojenliğe etkili olan faktörlerin, doğal olarak kümelenebilmesi nedeniyle, geniş ölçekli çalışmalar için uygun bir yöntemdir. Örneğin Missisippi Alluvial Vadisinde sonbaharın son dönemlerinde, su kuşlarının yemek olarak kullandığı ekili alanlardaki pirinç tanelerinin tahmin edilmesi için yapılmış çok aşamalı örnekleme tasarımından elde edilen sonuçlar, simülasyon verileri ile elde edilmiş basit rasgele örnekleme sonuçlarıyla karşılaştırılmış ve çok aşamalı örneklemenin basit rasgele örneklemeye göre istatistik etkinliğinin 3.2 ile 9 kat arasında değiştiği ve çok aşamalı örnekleme birimlerinin kümelenmesi sonucunda maliyetin azalmasıyla, basit rasgele örnekleme tasarımına harcanan araştırma maliyetinin çok aşamalı örneklemeye göre ortalamadan 1.4 kat daha fazla olduğu tespit edilmiştir. Ancak, çok aşamalı örnekleme yönteminin göreli etkinliğinin, maliyet ve varyans bileşenlerindeki değişimlere karşı hassas olması nedeniyle basit rasgele örnekleme yöntemine göre

* *İlgili yazar: sevil@hacettepe.edu.tr*

avantajları belirtildiğinde bu değişimlere dikkat etmek gerekir (Stafford vd., 2006).

Günümüzde yapılan büyük ölçekli araştırmalarda, çok aşamalı örnekleme yönteminin yaygın olarak kullanıldığı söylenebilir. Bu nedenle,çalışmada, çok aşamalı örnekleme yöntemlerinden iki aşamalı, iki aşamalı küme ve üç aşamalı örnekleme yöntemleri ve bu yöntemlerde uygun örneklem büyüklüklerinin belirlenmesi incelenmiştir. Çalışmanın uygulama bölümünde ise 1997 yılında Türkiye'deki 828 ilçeye ait buğday üretim miktarına ait veriler kullanılarak; iki aşamalı, iki aşamalı küme ve üç aşamalı örnekleme yöntemleri için gerekli örneklem büyüklükleri belirlenmiş ve tahminler verilmiştir. Ayrıca basit ve tabakalı rasgele örnekleme yöntemleri ile elde edilen sonuçlar karşılaştırılmıştır.

2. İki Aşamalı Örnekleme

Basit veya tabakalı rasgele örnekleme yöntemlerini kullanmak için populasyondaki tüm birimlere ait bir çerçeve olması gerekirken, iki aşamalı örnekleme yönteminde sadece örnekleme alınan ön birimler için çerçeve hazırlanması yeterlidir. Sadece örnekleme alınan ön birimler için çerçeveye ihtiyaç duyulması, örnekleme yönteminin uygulaması için gerekli olan emek ve maliyeti azaltır. Bu durum, iki aşamalı örnekleme yönteminin diğer yöntemlere göre daha düşük duyarlılığa sahip tahmin sonuçları verdiği durumlarda bile avantaj sağlamaktadır. İki aşamalı örnekleme, basit rasgele ve tabakalı rasgele örnekleme yöntemine göre uygulamada daha uygun ve maliyeti az olan bir yöntemdir (Tryfos, 1996).

2.1. Alt birim büyüklükleri eşit olan iki aşamalı örnekleme

Ön birimlerin eşit büyüklükte olduğu durumlarda uygulanan iki aşamalı örnekleme yöntemidir. Örneğin, bir sitedeki blokların ön birim, hane halklarının alt birim olarak kabul edildiği iki aşamalı bir hane halkı araştırmasında bloklarda yaşayan hane halklarının eşit sayıda olması durumunda alt birim büyüklükleri eşit olan iki aşamalı örnekleme planı uygulanabilir.

Yöntemde, N: Kitlede bulunan ön birimlerin sayısı (Küme sayısı), n: Belirlenen örneklemde bulunan ön birimlerin sayısı, M: Belirlenen ön birimde bulunan alt birim sayısı (Kümelerde yer alan birimlerin sayısı) ve m: Alınan örneklemde bulunan alt birim sayısı olarak tanımlanır. Ortalama tahmin ($\bar{\bar{y}}$):

$$\bar{\bar{y}} = \frac{1}{n}\sum \bar{y}_i = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij}}{nm} \qquad (1)$$

olarak tanımlanır. Eşitlikte, y_{ij}: i = 1, 2, …, n ve j = 1, 2, 3, …, m olmak üzere; örneklemde *i.* ön birimde yer alan *j.* alt birimin değeridir. Ortalama tahmininin varyansı,

$$V(\bar{\bar{y}}) = \frac{N-n}{N}\frac{S_1^2}{n} + \frac{M-m}{M}\frac{S_2^2}{mn} \qquad (2)$$

eşitliği ile elde edilir. Burada, $S_1^2 = \frac{\sum_{i=1}^{N}(\bar{Y}_i - \bar{\bar{Y}})^2}{N-1}$ ön birim ortalamaları arası varyans ve $S_2^2 = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}(y_{ij} - \bar{Y}_i)^2}{N(M-1)}$ ise alt birimler arası varyansdır. Birinci aşamada örnekleme oranı $f_1 = n/N$ ile, ikinci aşamada $f_2 = m/M$ ile gösterilirse varyans,

$$V(\bar{\bar{y}}) = \frac{1-f_1}{n}S_1^2 + \frac{1-f_2}{mn}S_2^2$$

olarak tanımlanır. Varyansın tahmini ise,

$$v(\bar{\bar{y}}) = \frac{1-f_1}{n}s_1^2 + \frac{1-f_2}{mn}s_2^2 \qquad (3)$$

olur. Eşitlikte s_1^2 ve s_2^2 örneklem üzerinden elde edilen varyans tahminleridir (Çıngı, 1994).

2.1.1. Örneklem büyüklüğünün belirlenmesi

İki aşamalı örneklemede, ön birimin ortalama araştırma maliyeti, alt birimin ortalama araştırma maliyetinden çok daha yüksektir. Daha genel bir ifade ile daha çok ön birim ve belirlenen ön birimlerden daha az alt birimin örnekleme alınması, örnekleme hatalarını azaltır dolayısıyla güvenilirdir ancak maliyeti yükseltebilir.

En basit maliyet fonksiyonu (C); sabit maliyet ihmal edildiğinde ve c_1 ile c_2' nin sırasıyla ön birim ve alt birim seçme maliyetleri olduğunda ve aralarındaki ilişkinin de doğrusal olduğu varsayıldığında,

$$C = c_1 n + c_2 nm \qquad (4)$$

olarak tanımlanır. Eşitlik (2) 'den,

$$V(\bar{\bar{y}}) = \frac{1}{n}(S_1^2 - \frac{S_2^2}{M}) + \frac{1}{mn}S_2^2 - \frac{1}{N}S_1^2 \qquad (5)$$

olarak yazılabilir. Maliyet kısıtlayıcısı altında, varyansı en küçük yapmak için Lagrange çarpanları yönteminden yararlanılarak,

$$V(\bar{\bar{y}})+\lambda(c_1n+c_2nm-C)= \tag{6}$$
$$\frac{1}{n}(S_1^2-\frac{S_2^2}{M})+\frac{1}{mn}S_2^2-\frac{1}{N}S_1^2+\lambda(c_1n+c_2nm-C)$$

ifadesinin en küçük olmasını sağlayan λ çarpanının bulunması istenir. Belirlenmek istenen değişkenler n ve m olduğundan n ve m'ye göre kısmi türevler alınıp sıfıra eşitlenirse,

$$m=\sqrt{\frac{c_1}{c_2}\frac{S_2^2}{S_1^2-S_2^2/M}} \tag{7}$$

eşitliği elde edilir (Cohran, 1977; Uçar, 2009). m'nin değerinin elde edilmesinden sonra n'nin değerinin belirlenmesi için iki yöntem bulunmaktadır. Birinci yöntem; n'yi, belirli bir maliyet kısıtı altında minimum varyans için maliyet fonksiyonunda m'nin yerine yazarak belirlemektir. Maliyet fonksiyonu (4), n'ye göre düzenlendiğinde, $C\le C_0$ durumu için,

$$n=\frac{C}{c_1+c_2m}$$

$$n=\frac{C_0}{c_1+c_2\left(\frac{c_1}{c_2}\frac{S_2^2}{S_1^2-S_2^2/M}\right)} \tag{8}$$

eşitliği elde edilir.

İkinci yöntemde n belirli bir varyans altında minimum maliyet için (5)' de verilen varyans eşitliğinde m'nin değeri yerine yazarak elde etmektir. $V\le V_0$ için,

$$n=\frac{S_1^2-\frac{S_2^2}{M}+\frac{S_2^2}{m}}{V_0+\frac{S_1^2}{N}} \tag{9}$$

olarak elde edilir (Sukhatme vd., 1970; Wang vd., 2006).

2.2. Alt birim büyüklükleri farklı olan iki aşamalı örnekleme

Alt birim büyüklükleri farklı olan iki aşamalı örneklemede, kitle ortalamasının tahmini oransal tahminden yararlanılarak elde edilebilmektedir. Bu durumda yardımcı değişken, alt birim büyüklükleridir ve kitle toplamı M_0 bilindiğinde kitle ortalaması

$$\bar{\bar{y}}_o=\frac{\sum_{i=1}^{n}M_i\bar{y}_i}{\sum_{i=1}^{n}M_i} \tag{10}$$

dır. Oransal tahmine ilişkin hata kareler ortalaması ise,

$$HKO(\bar{\bar{y}}_o)=\frac{1-f_1}{n}S_b^2+\frac{1-f_2}{n\bar{m}}S_2^2 \tag{11}$$

eşitliği ile elde edilir. Eşitlikte,

$$S_b^2=\frac{\sum_{i=1}^{N}M_i^2(\bar{Y}_i-\bar{\bar{Y}})^2}{\bar{M}^2(N-1)},\ S_2^2=\sum_{i=1}^{N}\frac{M_i}{M_0}S_i^2 \quad \bar{M}=\frac{1}{N}\sum_{i=1}^{N}M_i \text{ ve}$$

$\bar{m}=\frac{1}{n}\sum_{i=1}^{n}m_i$ olarak tanımlanır.

2.2.1. Örneklem büyüklüğünün belirlenmesi

Alt birim büyüklükleri farklı olan iki aşamalı örneklemede, örneklem büyüklüğünün belirlenmesi eşit olan duruma benzer olarak elde edilir. Ancak bu durum için maliyet fonksiyonu,

$$C=c_1n+c_2n\bar{m} \tag{12}$$

biçiminde tanımlanır. Burada c_1 ön birim, c_2 ise alt birim başına örnekleme maliyetidir.

(11) eşitliği ile verilen HKO, maliyet kısıtlayıcısı altında en küçük yapmak için Lagrange çarpanları yöntemi kullanılarak,

$$HKO(\bar{\bar{y}})+\lambda(c_1n+c_2n\bar{m})= \tag{13}$$
$$\frac{1}{n}(S_b^2-\frac{S_2^2}{\bar{M}})+\frac{1}{n\bar{m}}S_2^2-\frac{1}{N}S_b^2+\lambda(c_1n+c_2n\bar{m})$$

şeklinde yazmak mümkündür. Belirlenmek istenen değişkenler n ve $\bar{m}$ olduğundan, n ve $\bar{m}$'ye göre kısmi türevler alıp sıfıra eşitlendiğinde,

$$\bar{m}=\sqrt{\frac{c_1}{c_2}\frac{S_2^2}{S_b^2-S_2^2/\bar{M}}} \tag{14}$$

eşitliği elde edilir. $\bar{m}$'nin değerinin elde edilmesinden sonra n'nin değerinin belirlenmesi için izlenecek iki yöntem bulunmaktadır. n'nin elde edilmesi birinci yöntem için, maliyet fonksiyonunun n'e göre düzenlenmesi ve $\bar{m}$'nın yerine yazılmasıyla,

($C \le C_0$ için),

$$n = \frac{C_0}{c_1 + c_2\sqrt{\dfrac{c_1}{c_2}\dfrac{S_2^2}{S_1^2 - S_2^2/M}}} = \frac{C_0}{c_1 + S_2\sqrt{\dfrac{c_1 c_2}{S_1^2 - S_2^2/M}}} \tag{15}$$

olarak elde edilir. İkinci yöntem için ise n'yi, belirli bir varyans altında minimum maliyet için, (11) eşitliğinde $\overline{m}$'nin yerine yazılmasıyla ($V \le V_0$) için,

$$n = \frac{S_b^2 - \dfrac{S_2^2}{\overline{M}} + \dfrac{S_2^2}{\overline{m}}}{V_0 + \dfrac{S_b^2}{N}} \tag{16}$$

olarak elde edilir (Sukhatme vd., 1970). $\overline{m}$'nın belirlenmesinden sonra seçilmiş i. (birinci) aşama birimindeki ikinci aşama biriminin büyüklükleri, $m_i = M_i \dfrac{\overline{m}}{\overline{M}}$ i=1, 2,…,n eşitliği ile elde edilir (Wang vd., 2006).

3. İki Aşamalı Küme Örneklemesi

İki aşamalı küme örneklemesinde kitlenin N sayıda birinci aşama biriminden oluştuğu varsayılır.i, (i = 1,2, …, N) birinci aşama birimi, M_i sayıda ikinci aşama birimi, *i.* birinci aşama biriminde yer alan *j* ikinci aşama birimi ise K_{ij} (i = 1, 2, …,N ve j =1, 2, …, M) sayıda üçüncü aşama birimi içermektedir. Her bir birinci aşama birimi ortalama $\overline{M}$ sayıda ikinci aşama birimi içerir. Her bir ikinci aşama birimi ise ortalama $\overline{K}$ sayıda üçüncü aşama birimi içerir. Kitle toplamda K sayıda üçüncü aşama biriminden oluşmaktadır.
İlk aşamada, kitleden n sayıda birinci aşama birimi, ikinci aşamada, alınan *i.* birinci aşama biriminden m_i sayıda ikinci aşama birimi alınır (i = 1,2…n). Alınan her bir birinci aşama biriminden ortalama $\overline{m}$ sayıda ikinci aşama birimi alınmıştır. Üçüncü aşama birimlerinin hepsi ikinci aşama birimlerinden araştırılarak elde edilmiştir.

İki aşamalı küme örneklemesinde kitle ortalaması tahmini,

$$\overline{\overline{y}} = \frac{N}{Kn}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{j=1}^{m_i}\sum_{u=1}^{K_{ij}} y_{iju} \tag{17}$$

olarak tanımlanır. Burada , y_{iju} : Örneklemde (*i.*) birinci aşama biriminde yer alan (*j.*) ikinci aşama biriminde bulunan (*u.*) üçüncü aşama biriminin değeri (u= 1, 2,…, k_{ij} ; j = 1, 2,..., m_i ; i = 1, 2,..., n).

Kitle ortalamasının tahminine ilişkin varyans,

$$V(\overline{\overline{y}}) = \frac{S_1^2}{n}\left(1 - \frac{n}{N}\right) + \frac{S_2^2}{n\overline{m}}\left(1 - \frac{\overline{m}}{\overline{M}}\right) = \frac{1}{n}\left(S_1^2 - \frac{S_2^2}{\overline{M}}\right) + \frac{S_2^2}{n\overline{m}} - \frac{S_1^2}{N} \tag{18}$$

olarak tanımlanır. Eşitlikte, ön birim ve alt birimler arası varyans,

$$S_1^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\overline{Y}_i - \overline{\overline{Y}})^2 \quad \text{ve}$$

$$S_2^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{(M_i - 1)}\sum_{j=1}^{M_i}(\overline{Y}_{ij} - \overline{\overline{Y}}_i)^2$$

dır. Kitle ortalamasının varyansının örneklemden tahmini ise

$$v(\overline{\overline{y}}) = \frac{s_1^2}{n}\left(1 - \frac{n}{N}\right) + \frac{s_2^2}{n\overline{m}}\left(1 - \frac{\overline{m}}{\overline{M}}\right) = \frac{1}{n}\left(s_1^2 - \frac{s_2^2}{\overline{M}}\right) + \frac{s_2^2}{n\overline{m}} - \frac{s_1^2}{N} \tag{19}$$

eşitliği ile verilir.

3.1. Örneklem büyüklüğünün belirlenmesi

İki aşamalı küme örneklemesinde kitle ortalaması için örneklem büyüklüğünün belirlenmesinde kullanılan maliyet fonksiyonu,

$$C = c_1 n + c_2 n\overline{m} + c_3 n\overline{m}\overline{K} = c_1 n + (c_2 + c_3\overline{K})n\overline{m} \tag{20}$$

dır. Eşitlikde; C, araştırmanın toplam maliyeti, c_1 birinci aşama birimi başına örnekleme maliyeti, c_2 ikinci aşama birimi başına örnekleme maliyeti, c_3 üçüncü aşama birimi başına araştırma maliyetidir.

Varyans eşitliğini (18), maliyet kısıtlayıcısı altında en küçük yapmak gerekir. Lagrange çarpanları yönteminden,

$$V(\overline{\overline{y}}) + \lambda\left(c_1 n + (c_2 + c_3\overline{K})n\overline{m}\right) = -\frac{S_1^2}{N} + \frac{1}{n}\left(S_1^2 - \frac{S_2^2}{\overline{\overline{M}}}\right) + \frac{1}{n\overline{m}}S_2^2 + \lambda\left(c_1 n + (c_2 + c_3\overline{K})n\overline{m}\right) \tag{21}$$

dır. Eşitliğin, n ve $n\overline{m}$' ye göre kısmi türevleri alınıp sıfıra eşitlenir ve gerekli düzenlenmeler yapılırsa ,

$$\overline{m} = \frac{S_2}{\sqrt{S_1^2 - S_2^2/\overline{M}}}\sqrt{\frac{c_1}{c_2 + c_3\overline{K}}} \tag{22}$$

olarak elde edilir.

Birinci yönteme göre n değerini belirlemek için maliyet fonksiyonunda $\overline{m}$' nın yerine yazılır ve düzenlenirse, $C \le C_0$ durumu için

$$n = \frac{C_0}{c_1 + c_2\left(\frac{S_2}{\sqrt{S_1^2 - S_2^2/\overline{M}}}\sqrt{\frac{c_1}{c_2 + c_3\overline{K}}}\right) + c_3\left(\frac{S_2}{\sqrt{S_1^2 - S_2^2/\overline{K}}}\sqrt{\frac{c_1}{c_2 + c_3\overline{K}}}\right)\overline{K}}$$

(23) olarak elde edilir. İkinci yöntem için ise n'nin, belirli bir varyans altında minimum maliyet için varyans eşitliği (16) da $\overline{m}$'nin değerinin yerine yazılıp düzenlendiğinde, $V \le V_0$ için,

$$n = \frac{S_1^2 - \frac{S_2^2}{\overline{M}} + \frac{S_2^2}{\overline{m}}}{V_0 + \frac{S_1^2}{N}} \tag{24}$$

olarak elde edilmesidir. $\overline{m}$'nın belirlenmesinden sonra belirlenmiş *i.* birinci aşama birimindeki ikinci aşama birimlerinin büyüklükleri, $m_i = M_i \frac{\overline{\overline{m}}}{\overline{M}}$ i= 1,..,n eşitliği ile elde edilir (Wang vd., 2006; Uçar, 2009).

4. Üç Aşamalı Örnekleme

İki aşamalı örneklemenin ikinci aşamasında örnekleme alınan birimler alt birimlere ayrılır. Bu birimler arasından yeniden örneklem belirlenmesi üç aşamalı örneklemedir.
Üç aşamalı örneklemede kullanılan formüller iki aşamalı örneklemede kullanılan formüllerden geliştirilmiştir. Dolayısıyla araştırmanın yapısına uygun olarak aşama sayısının artması mümkün olmaktadır. Ancak aşama sayısı artarken, tahminlerin varyanslarının hesaplanması giderek zorlaşmaktadır (İdil, 1980).

Üç aşamalı örneklemenin üstünlükleri, iki aşamalı örnekleme ile benzerlik göstermektedir. Birincisi sadece örnekleme alınmış olan birinci aşama birimleri, ikinci aşama birimleri ve üçüncü aşama birimleri için çerçevenin oluşturulmasına ihtiyaç duyulmaktadır. İkinci olarak, basit rasgele örnekleme yöntemi ile karşılaştırıldığında bu yöntem ulaşım ve yönetim maliyetlerini azaltmaktadır.

İkinci ve üçüncü aşama birimleri arasında homojenlik, birinci aşama birimleri arasında heterojenlik olması durumunda daha çok sayıda ikinci ve üçüncü aşama biriminin alınmasının araştırmaya katacağı bilgi ile daha az sayıda ikinci ve üçüncü aşama biriminin alınmasının araştırmaya katacağı bilgi arasında belirgin derecede farklılık olmayacağı sezgisel olarak söylenebilmektedir. Bu durumda kitle hakkında daha fazla bilgi edinmek için birinci aşama birimlerinin sayısını artırmak gerekli olacaktır. Ancak birinci aşama birimlerini elde etmek, ulaşım ve yönetim maliyetinin artmasına neden olacaktır ki bu durum örnekleme maliyetinin en aza indirgenmesi amacına uygun değildir (Yamane, 2001).

Üç aşamalı örnekleme için verilecek gösterimler, iki aşamalı örneklemenin devamı biçimindedir. Üç aşamalı örneklemede, kitlenin N sayıda birinci aşama biriminden oluştuğu varsayılır. i. (i=1, 2, ...,N) birinci aşama birimi M_i sayıda ikinci aşama birimi, i .birinci aşama biriminde yer alan j .ikinci aşama birimi ise K_{ij} (i=1, 2, ...,N, j = 1, 2, ..., M) sayıda üçüncü aşama birimi içermektedir. Her bir birinci aşama birimi ortalama $\overline{M}$ sayıda ikinci aşama birimi içerir. Her bir ikinci aşama birimi ise ortalama $\overline{K}$ sayıda üçüncü aşama birimi içerir. Kitle toplamda K sayıda üçüncü aşama biriminden oluşmaktadır.

İlk aşamada, kitleden n sayıda birinci aşama birimi alınır. İkinci aşamada, birinci aşamada alınan (i=1, 2,...,n) birimlerden m_i sayıda ikinci aşama birimi alınır. Alınan her bir birinci aşama biriminden $\overline{m}$ sayıda ikinci aşama birimi alınmıştır. Üçüncü aşamada, bir önceki aşamada alınmış (j) ikinci aşama birimlerinden k_{ij} sayıda üçüncü aşama birimleri alınır. Alınan her bir ikinci aşama biriminden ortalama $\overline{k}$ sayıda üçüncü aşama birimi alınmıştır. Kitle ortalamasının tahmini $\overline{\overline{\overline{y}}}$,

$$\overline{\overline{\overline{y}}} = \frac{N}{Kn}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{j=1}^{m_i}\frac{K_{ij}}{k_{ij}}\sum_{u=1}^{k_{ij}} y_{iju} \tag{25}$$

olarak tanımlanır. Eşitlikte, y_{iju}: Örneklemde i. birinci aşama biriminde yer alan j. ikinci aşama biriminde bulunan u. üçüncü aşama biriminin değeri $u = 1,2,..,k_{ij}$; $j = 1,2,..,m_i$; $i = 1,2,..,n$.

Kitle ortalamasının tahminine ilişkin varyans ise

$$V(\overline{\overline{\overline{y}}}) = \frac{S_1^2}{n}\left(1 - \frac{n}{N}\right) + \frac{S_2^2}{n\overline{m}}\left(1 - \frac{\overline{m}}{\overline{M}}\right) + \frac{S_3^2}{n\overline{m}\overline{k}}\left(1 - \frac{\overline{k}}{\overline{K}}\right) \tag{26}$$

olarak tanımlanır. Ön birim ve alt birimler arası varyans

$$S_1^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\overline{\overline{Y}}_i - \overline{\overline{\overline{Y}}})^2 \quad (27)$$

$$S_2^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{(M_i-1)}\sum_{j=1}^{M_i}(\overline{Y}_{ij} - \overline{\overline{Y}}_i)^2 \quad (28)$$

$$S_3^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M_i}\sum_{j=1}^{M_i}\frac{1}{K_{ij}-1}\sum_{k=1}^{K_{ij}}(y_{iju} - \overline{Y}_{ij})^2 \quad (29)$$

eşitlikleri ile verilmektedir. Varyansın örneklemden tahmini ise,

$$v(\overline{\overline{\overline{y}}}) = \frac{s_1^2}{n}\left(1-\frac{n}{N}\right) + \frac{s_2^2}{n\overline{m}}\left(1-\frac{\overline{m}}{\overline{M}}\right) + \frac{s_3^2}{n\overline{m}\overline{k}}\left(1-\frac{\overline{k}}{\overline{K}}\right) \quad (30)$$

eşitliği ile elde edilir. Eşitliklerdeki , $\overline{K}$ kitlede ve $\overline{k}$ ise örneklemde yer alan üçüncü aşama birimlerinin ortalama sayısı,

$$\overline{K} = \frac{1}{M_0}\sum_{i=1}^{M_i}K_{ij} \qquad \overline{k} = \frac{1}{m_0}\sum_{i=1}^{m_i}k_{ij}$$

olarak tanımlanır. M_0 kitlede, m_0 ise örneklemde yer alan ikinci aşama birimlerin toplam sayısıdır. (Wang vd., 2006).

4.1. Örneklem büyüklüğünün belirlenmesi

Üç aşamalı örnekleme yönteminde örneklem büyüklüğünün belirlenmesi için maliyet fonksiyonu,

$$C = c_1 n + c_2 n\overline{m} + c_3 n\overline{m}\overline{k} \quad (31)$$

olarak tanımlanır. Eşitlikte, c_1, c_2, c_3 sırasıyla birinci, ikinci ve üçüncü aşama birimlerinin örnekleme alınma maliyetidir. C ise toplam maliyetdir. Varyans eşitliği (23), düzenlenmesi sonucunda

$$V(\overline{\overline{\overline{y}}}) = -\frac{S_1^2}{N} + \frac{1}{n}\left(S_1^2 - S_2^2/\overline{M}\right) + \frac{1}{n\overline{m}}\left(S_2^2 - S_3^2/\overline{K}\right) + \frac{S_3^2}{n\overline{m}\overline{k}}$$

(32) eşitliği ile verilir. Varyans eşitliğini maliyet kısıtlayıcısı altında en küçük yapmak için Lagrange çarpanları kulllanılır. Buradan,

$$V(\overline{\overline{\overline{y}}}) + \lambda(c_1 n + c_2 n\overline{m} + c_3 n\overline{m}\overline{k})$$
$$= -\frac{S_1^2}{N} + \frac{1}{n}(S_1^2 - \frac{S_2^2}{\overline{M}}) + \frac{1}{n\overline{m}}(S_2^2 - \frac{S_3^2}{\overline{K}}) + \frac{S_3^2}{n\overline{m}\overline{k}}$$
$$+ \lambda(c_1 n + c_2 n\overline{m} + c_3 n\overline{m}\overline{k}) \quad (33)$$

olarak elde edilir. Belirlenmek istenen değişkenler n, $\overline{m}$ ve $\overline{k}$ olduğundan n, $n\overline{m}$ ve $n\overline{m}\overline{k}$' ya göre kısmi türevler alınıp sıfıra eşitlendiğinde,

$$\overline{k} = \frac{S_3}{\sqrt{S_2^2 - S_3^2/\overline{K}}}\sqrt{\frac{c_2}{c_3}} \quad (34)$$

$$\overline{m} = \frac{\sqrt{S_2^2 - S_3^2/\overline{K}}}{\sqrt{S_1^2 - S_2^2/\overline{M}}}\sqrt{\frac{c_1}{c_2}} \quad (35)$$

olarak elde edilir. İki aşamalı örneklemede olduğu gibi birinci yöntem ile n'nin değeri maliyet fonksiyonunda $\overline{k}$ ve $\overline{m}$' nin yerine yazılmasıyla $C \le C_0$ durumu için

$$n = \frac{C_0}{c_1 + c_2\left(\frac{\sqrt{S_2^2 - S_3^2/\overline{K}}}{\sqrt{S_1^2 - S_2^2/\overline{M}}}\sqrt{\frac{c_1}{c_2}}\right) + c_3\left(\frac{\sqrt{S_2^2 - S_3^2/\overline{K}}}{\sqrt{S_1^2 - S_2^2/\overline{M}}}\sqrt{\frac{c_1}{c_2}}\right)\left(\frac{S_3}{\sqrt{S_1^2 - S_2^2/\overline{K}}}\sqrt{\frac{c_2}{c_3}}\right)}$$

(36) olarak elde edilir. İkinci yöntemde ise n, belirli bir varyans altında minimum maliyet için, varyans eşitliğinde (32), $\overline{m}$ ve $\overline{k}$ değerlerinin yerine yazılıp, düzenlenlenmesi ile $V \le V_0$ için,

$$n = \frac{S_1^2 - \frac{S_2^2}{\overline{M}} + \frac{1}{\overline{m}}(S_2^2 - \frac{S_3^2}{\overline{K}}) + \frac{S_3^2}{\overline{m}\overline{k}}}{V_0 + \frac{S_1^2}{N}} \quad (37)$$

olarak elde edilir. $\overline{m}$ değerinin belirlenmesinden sonra alınan (i)birinci aşama birimindeki, ikinci aşama biriminin büyüklüğü, m_i iki aşamalı örneklemede verildiği gibidir. $\overline{k}$ belirlenmesinden sonra alınmış birinci aşama biriminden alınan (j) ikinci aşama birimindeki üçüncü aşama biriminin büyüklüğü ise $k_{ij} = K_{ij}\frac{\overline{k}}{\overline{K}}$ $\quad i = 1,2,..,n \quad j = 1,2,..,m_i$ eşitliği ile elde edilir (Uçar, 2009).

5. Uygulama

Uygulamada, çok aşamalı örnekleme yöntemlerinde örneklem büyüklüğünün belirlenmesi için Demir (1999), çalışmasında kullanılan 1997 yılına ait Türkiye'deki 828 ilçeye ilişkin veriler kullanılmıştır. Bu veriler 7 coğrafi bölge, 78 il ve toplam 828 ilçeye ait buğday ürün (ton) miktarlarını içermektedir.
Alt birim büyüklükleri farklı iki aşamalı, iki aşamalı küme örneklemesi ve üç aşamalı örnekleme yöntemlerinde örneklem büyüklüklerinin belirlenmesini göstermek amacıyla ise öncelikle ön birim ve alt birim büyüklükleri için temsili değerler verilerek ön örneklemler oluşturulmuştur. Elde edilen bilgiler yardımıyla iki aşamalı, iki aşamalı küme ve üç aşamalı örnekleme yöntemleri için alt birim büyüklükleri ve ön birim büyüklükleri belirli bir varyans altında minimum maliyet için hesaplanmıştır. Daha sonra kitleden belirlenen ön birim ve alt birim büyüklüklerine göre üç ayrı yöntemle örneklemler alınmış ve Türkiye için ortalama buğday ürün miktarı tahminleri ve tahminlerin varyansları hesaplanmıştır.

İki aşamalı örnekleme için iller ön birim, ilçeler ise alt birim olarak alınmıştır. Dolayısıyla, $N=78$, $M_0=828$ ve $\overline{M}=\frac{828}{78}=10.61$ dır. İllerdeki ilçe sayıları eşit olmadığı için alt birim büyüklükleri farklı olan iki aşamalı örnekleme için örneklem büyüklüğü hesaplanmış ve tahminler elde edilmiştir. Buna göre, hesaplamalar için gerekli değerler, kitleden ön örneklem alınarak, ön birim ve alt birimler arası varyansın örneklemden tahmini ve kitle ortalaması tahmininin varyansı $s_b^2=657382172.5$ $s_2^2=967359169.7$ ve $v(\bar{\bar{y}})=51616087$ olarak hesaplanmıştır. Elde edilen bu bilgiler yardımıyla, illeri belirleme maliyetinin ilçeleri belirleme maliyetine oranının 4 olduğu kabul edilerek **(**Wang vd., 2006**)**, öncelikle ikinci aşamada, belirlenmiş illerden alınacak ortalama ilçe sayısı (14) eşitliğinden,

$$\overline{m}=\frac{\sqrt{967359169.7}}{\sqrt{657382172.5-967359169.7/10.61}}\sqrt{4}=2.61$$

olarak belirlenmiştir. (16) eşitliğinden ise birinci aşamada Türkiye'deki tüm illerden alınması gereken il sayısı,

$$n=\frac{657382172.5-967359169.7/10.61+967359169.7/2.61}{51616087+657382172.5/78}=15.59$$

olarak bulunur. Daha sonra *i.* ilden alınacak ilçe sayısı hesaplanabilir. Örneğin birinci aşamada örnekleme alınmış olan Aydın ilinde yer alan 17 ilçeden ikinci aşamada örnekleme alınacak ilçe sayısı $m_A=17x\frac{2.61}{10.61}=4$ olmalıdır.

Örnekleme alınacak il sayısı ve illerden alınacak ilçe sayıları belirlendikten sonra kitleden Basit rasgele örnekleme ile örneklem alınmıştır. Alınan örneklemden elde edilen kitle ortalaması tahmini ve tahminin varyansı Tablo 1'de verilmiştir.

İki aşamalı küme örneklemesi için bölgeler birinci aşama birimi, iller ise ikinci aşama birimi olarak alınmıştır. Bu durumda, $N=7$, $M_0=78$ ve $\overline{M}=\frac{78}{7}=11.14$ dır. Ön örneklemden, birinci aşama ve ikinci aşama birimleri arası varyansın tahminleri ve kitle ortalaması tahmini varyansı $s_1^2=340570550.8$, $s_2^2=501088664.6$ $v(\bar{\bar{y}})=83286660.2$ olarak hesaplanmıştır. Elde edilen bu bilgiler yardımıyla, ikinci aşamada, bölgeleri belirleme maliyetinin illeri belirleme maliyetine oranının ve illeri belirleme maliyetinin ilçeleri araştırma maliyetine oranının 4 olduğu kabul edilerek, öncelikle belirlenmiş bölgelerden alınacak olan ortalama il sayısı (22) eşitliğinden,

$$\overline{m}=\frac{\sqrt{501088664.6}}{\sqrt{340570550.8-501088664.6/11.14}}\sqrt{\frac{16}{4+10.61}}=1.36$$

(24) eşitliğinden ise, ilk aşamada Türkiye'deki 7 bölgeden alınması gereken bölge sayısı

$$n=\frac{340570550.8-501088664.6/11.14+501088664.6/1.36}{83286660.2+340570550.7/7}=5.02$$

olarak bulunur. Daha sonra *i.* bölgeden alınacak il sayısı hesaplanabilir. Örneğin, ilk aşamada örnekleme alınmış olan Doğu Anadolu bölgesinde yer alan 14 ilden, ikinci aşamada örnekleme alınacak il sayısı $m_{D.A}14x\frac{1.36}{11.14}=2$ olmalıdır. Örnekleme alınacak bölge sayısı ve bölgelerden alınacak il sayıları belirlendikten sonra kitleden örneklem alınmıştır. Alınan örneklemden elde edilen kitle ortalaması tahmini, birinci aşama ve tahminin varyansı Tablo 1'de verilmiştir.

Üç aşamalı örnekleme için bölgeler birinci aşama birimi, iller ikinci aşama birimi, ilçeler ise üçüncü aşama birimi olarak alınır ise $N=7$, $M_0=78$, $K=828$, $\overline{M}=\frac{78}{7}=11.14$ ve $\overline{K}=\frac{828}{78}=10.61$ dır. Ön örneklemden, birinci aşama birimleri, ikinci aşama birimleri ve üçüncü aşama birimleri arasındaki varyansın tahminleri ve kitle ortalaması tahmininin varyansı $s_1^2 = 704596829$ $s_2^2 = 1295772850$

$s_3^2 = 2656872030$ $V(\bar{\bar{\bar{y}}}) = 2182831116$ olarak

hesaplanmıştır. Elde edilen bu bilgiler yardımıyla, bölgeleri belirleme maliyetinin illeri belirleme maliyetine oranının ve illeri belirleme maliyetinin ilçeleri belirleme maliyetine oranının 4 olduğu kabul edilerek, ikinci aşamada belirlenen bölgelerden alınacak ortalama il sayısı ve üçüncü aşamada belirlenen illerden alınacak ortalama ilçe sayısı (34) ve (35) eşitliklerden,

$$\bar{m} = \frac{\sqrt{1295772850 - 2656872030/10.61}}{\sqrt{704596829 - 1295772850/11.14}} \sqrt{4} = 2.66$$

$$\bar{k} = \frac{\sqrt{2656872030}}{\sqrt{1295772850 - 265872030/10.61}} \sqrt{4} = 3.18$$

olarak bulunur. Eşitlik (37)' den ise birinci aşamada Türkiye'deki 7 bölgeden alınması gereken bölge sayısı $n = 4.05$ olarak belirlenmiştir. *i.* bölgeden alınacak il sayısı hesaplanabilir. Örneğin birinci aşamada örnekleme alınmış olan Doğu Anadolu bölgesinde yer alan 14 ilden ikinci aşamada örnekleme alınacak il sayisi $m_{D:A} = 14x\frac{2.66}{11.14} = 3$ olmalıdır. Daha sonra *i.* ilden alınacak ilçe sayısı hesaplanabilir. Örneğin ikinci aşamada örnekleme alınmış olan Erzurum ilinde yer alan 19 ilçeden, üçüncü aşamada örnekleme alınacak ilçe sayısı $k_E = 19x\frac{3.18}{10.61} = 6$ olmalıdır. Örnekleme alınacak bölge sayısı ve bölgelerden alınacak il sayıları ve illerden alınacak ilçe sayılarının belirlenmesinin ardından kitleden örneklem alınmıştır. Alınan örneklemden elde edilen kitle ortalaması tahmini ve kitle ortalaması tahmininin varyansı Tablo 1'de verilmiştir.

Demir (1999) çalışmasında, ortalama buğday üretim miktarı tahmini ve tahmine ilişkin varyans değerlerini basit rasgele ve 7 bölgenin tabaka olarak alındığı tabakalı rasgele örnekleme yöntemleri için hesaplamıştır. Bu durum için N=828 ilçeden basit rasgele örnekleme için hoşgörü miktarı (d) 2400 olarak belirlenmiş ve n=100 ilçe alınmıştır.7 bölgenin tabaka olarak alındığı tabakalı rasgele örnekleme için ise n=100 ilçe bölgelere en iyi dağıtım kullanılarak dağıtılmıştır. Örnekleme yöntemlerine ilişkin sonuçlar Tablo 1'de verilmiştir.

Dolayısıyla, basit rasgele örnekleme ile örneklemin belirlenmesi, 828 ilçeye ilişkin buğday üretim miktarlarını içeren çerçeve yardımıyla yapılmıştır. Tabakalı rasgele örneklemede örneklem, çerçevedeki kitle birimlerinin 7 bölgeye göre seçim yapılmasıyla oluşturulmuştur. İki aşamalı ve iki aşamalı küme örneklemesinde ise çerçeve sadece ilk aşamada seçilen ön birimler için hazırlanırken, üç aşamalı örneklemede çerçeve, alınan birinci aşama birimleri ve birinci aşama birimlerinden alınan ikinci aşama birimleri için hazırlanmıştır.

Tablo 1. Örnekleme yöntemleri için örneklem büyüklükleri ve ortalama buğday ürün miktarı tahmini sonuçları

Yöntem	Örneklem büyüklüğü	Ortalama Tahmini	Varyans
Basit Rasgele	n=100	28893.00	10336122.5
Tabakalı Rasgele	n=100	25295.17	7884633.8
İki-Aşamalı	n=15 m_T=49	17062.40	14801232.0
İki - Aşamalı Küme	n=15 m_T=10 K_T=76	17395.60	45513910.0
Üç - Aşamalı	n=15 m_T=11 k_T=41	15175.09	12116520.9

Tablo 1 incelendiğinde,tabakalı rasgele örneklemede ortalama tahmin varyansı diğer yöntemlere göre küçüktür dolayısıyla daha duyarlı olduğu söylenebilir. Tabakalı rasgele örneklemeyi basit rasgele örnekleme, üç aşamalı örnekleme, iki aşamalı örnekleme ve iki aşamalı küme örneklemesi takip etmektedir.

6. Sonuç

Örneklemenin amacı, çok sayıda birimden oluşan kitle hakkındaki tahmin ve incelemelerin kısa zamanda ve düşük maliyetle gerçekleştirilmesini sağlamaktır. Tek aşamalı örnekleme yöntemlerinin uygulanması için tüm örneklem birimlerini içeren bir çerçeve hazırlanması gerekmektedir. Ancak çoğu araştırmada tüm örneklem birimlerini içeren böyle bir çerçevenin hazırlanması mümkün değildir ya da çerçevenin hazırlanması için yeterli mali kaynak bulunamamaktadır. Örneklem çerçevesinin, kitlenin tüm birimleri için elde edilemediği ve ayrıca çerçeve hazırlamanın aşırı emek ve yüksek bir maliyet gerektirdiği durumlarda, çok aşamalı örnekleme yaygın olarak kullanılmaktadır. Bu çalışmada, çok aşamalı örnekleme yöntemlerinden iki aşamalı, iki aşamalı küme ve üç aşamalı örnekleme yöntemlerinde örneklem büyüklüklerinin elde edilmesi incelenmiş ve bir uygulama verilmiştir. Uygulamada kullanılan veriler için tabakalı rasgele örnekleme ile daha duyarlı tahmin sonuçları elde edilmesine rağmen, aşamalı örnekleme yöntemlerinde daha az örneklem birimi ile çalışıldığından emek, maliyet ve zaman açısından avantaj sağlandığı söylenebilir.

Kaynaklar

Cohran, W.G. 1977. Sampling Techniques, John Wiley and Sons, New York, 428 p.

Çıngı, H. 1994. Örnekleme Kuramı, H.Ü. Fen Fakültesi Yayınları, Ankara, 346 s.

Demir, S. 1999. Çeşitli Örnekleme Yöntemlerinde Basit Doğrusal Regresyon Tahmin Edicileri ve Uygulamaları, Yükseks Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü. Ankara, 90 s.

İdil, O. 1980. Örnekleme Teorisi ve İşletme Yönetiminde Uygulanması, İstanbul Üniversitesi, İstanbul. 102 s.

Singh, S. 2003. Advanced Sampling Theory with Applications: How Micheal Selected Amy, Volume II, Kluwer Academic Publishers, Dordrecht, 1218 p.

Sukhatme, P.V., Sukhatme, B.V. 1970. Sampling Theory of Surveys with Applications. Ames, Iowa State University Press.

Stafford, J.D., Kaminski, R.M. , Reinecke K.J., Gerard, P.D. 2006. Multi-stage sampling for large scale natural resources surveys: a case study of rice and waterfowl, Journal of Environtmental Management,78, 353-361 p.

Uçar, P. 2009. İki ve Üç Aşamalı Örnekleme Yöntemlerinde Örneklem Büyüklüğünün Belirlenmesi. Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü. Ankara, 87 s.

Tryfos, P. 1996. Sampling Methods for Applied Research, John Wiley and Sons, New York, 123 p.

Yamane, T. 2001. Temel Örnekleme Yöntemleri, (çev: A.Esin, C.Aydın, M.A.Bakır, E.Gürbüzel), Literatür yayınları, ISBN 975-8431-34-X, İstanbul, 509 s.

Wang, J., Gao, G., Fan, Y., Chen, L., Liu, S., Jin, Y., Yu, J. 2006. The estimation of sample size in multi-stage sampling and its application in medical survey, Applied Mathematics and Computation, 178, 243-249 p.