ANADOLU ÜNİVERSİTESİ BİLİM VE TEKNOLOJİ DERGİSİ
ANADOLU UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
Cilt/Vol.:10-Sayı/No: 2 : 549-556 (2009)

ARAŞTIRMA MAKALESİ /**RESEARCH ARTICLE**

KAYIP GÖZLEM OLDUĞUNDA KİTLE ORTALAMASININ TAHMİNİ

Esra SATICI [1], Cem KADILAR

ÖZ

Bu çalışmada, kayıp gözlem olması durumunda kitle ortalamasına ilişkin, Singh ve Horn (2000), Singh ve Deo (2003) ve Rueda vd. (2005) tarafından sunulan tahmin ediciler incelenmiş ve bunlara alternatif iki yeni tahmin edici önerilmiştir. Önerilen tahmin edicilerin Hata Kareler Ortalamaları (HKO) bulunarak önerilen tahmin ediciler Singh ve Horn (2000), Singh ve Deo (2003) tahmin edicileriyle karşılaştırılmıştır. Elde edilen koşullar altında önerilen tahmin edicilerin daha etkin olduğu gözlenmiştir.

Anahtar Kelimeler : Kayıp gözlem, Hata kareler ortalaması, Yardımcı değişken bilgisi.

ESTIMATION OF THE POPULATION MEAN WHEN SOME OBSERVATIONS ARE MISSING

ABSTRACT

In this paper, the estimators in Singh and Horn (2000), Singh and Deo (2003) and Rueda *et al.* (2005) are investigated in the case of missing data and two new estimators alternative to these estimators are proposed. The mean squared error expressions (MSE) for the proposed estimators are obtained and compared analytically with MSE of the estimators in Singh and Horn (2000) and Singh and Deo (2003). By this way, we show the efficient condition for the proposed estimators.

Keywords: Missing data, Mean square error, Auxiliary information.

1. GİRİŞ

Kayıp gözlem 1970'lerden bu yana tartışılan bir problemdir. Çeşitli alternatif yaklaşımlar geliştirilmesine rağmen, birçok çalışmada kayıp gözlemler silinmekte ve geriye kalan eksiksiz gözlemler üzerinden istatistiksel analizler uygulanmaktadır. Bu yöntem, bazı araştırmalarda "amputasyon" olarak adlandırılmaktadır. Bu durum sadece önemli bilgilerin kaybına değil, aynı zamanda örneklem büyüklüğünün küçülmesine neden olduğundan tahminlerin güvenilirliğinin azalmasına, yanın artmasına neden olmaktadır.

Kayıp gözleme yaklaşım 4 ana başlık altında toplanabilir (Little ve Rubin, 2002): Yukarıda bahsedilen amputasyon yönteminin yanında, ağırlıklandırılmış yöntem, imputasyon ve model tabanlı çözümlemeler de mevcuttur. Ağırlıklandırılmış yöntem, cevapsız birimin ağırlığının değiştirilmesi esasına dayanır. Model tabanlı çözümler de, gözlenen veriler üzerinden oluşturulan modele göre kayıp gözlemlerin tamamlanmasıdır. Buna karşılık, imputasyon teknikleri, kayıp gözlemin tahmini ile tam veri kümesi oluşturarak bilinen klasik istatistiksel yöntemlerin uygulanmasına olanak sağlamaktadır. Çok sayıda imputasyon yöntemi önerilmiştir: Örneklemdeki diğer gözlem değerlerini kullanan "hot-deck imputasyonu", cevaplıların oluşturduğu alt örneklem ortalamasını kullanan ortalama imputasyonu ve regresyon modelinden tahmin değerleri elde eden regresyon imputasyonu bunlardan başlıcalarıdır. Ayrıca bu tekli

[1,] Hacettepe Üniversitesi Fen Fakültesi İstatistik Bölümü, 06800, Beytepe/ANKARA, Tel: (0312)2977900.
E-posta: eelagoz@hacettepe.edu.tr

Geliş: 5 Mayıs 2008; **Düzeltme:** 27 Kasım 2008; **Kabul:** 3 Haziran 2009

imputasyon yöntemlerine karşılık, herbir kayıp değer için birden fazla değer üreterek alternatif tam veri kümesini oluşturma esasına dayanan çoklu imputasyon yöntemi de mevcuttur.

Bu çalışmanın 2. bölümünde son yıllarda Singh ve Horn (2000), Singh ve Deo (2003) ve Rueda vd. (2005) tarafından literatüre kazandırılan tahmin ediciler incelenmiş; 3. bölümünde önerilen tahmin ediciler tanıtılarak literatürdeki tahmin edicilerle etkinlik yönünden karşılaştırmış ve 4. bölümünde ise yapılabilecek yeni çalışmalar için öneride bulunulmuştur.

2. LİTERATÜRDEKİ TAHMİN EDİCİLER

Örnekleme kuramında, ilgilenilen değişken ile ilgili kitle karakteristikleri tahmininde, ilgilenilen değişken ile ilişkili bir başka değişken (yardımcı değişken) bilgisini kullanmak tahmin güvenilirliğini artırmaktadır. Kayıp gözlem olması durumunda kitle ortalaması tahmini ile ilgili literatürde yer alan çalışmaları, yardımcı değişken kullanma ve kullanmama, yardımcı değişkende de kayıp gözlem olması veya olmaması şeklinde sınıflandırmak mümkündür. Bu sınıflandırmada yardımcı değişken bilgisi kullanılmadan kitle ortalaması tahminine ilişkin çalışmaya Gamrot (2007) örnek verilebilir. Singh ve Horn (2000), Singh ve Deo (2003) ve Rueda vd. (2005) tarafından yapılan çalışmalar ise kayıp gözlem olması durumunda yardımcı değişken bilgisi kullanılarak kitle ortalaması tahmin edicisi elde etmeye yönelik çalışmalardan bazılarıdır. Fakat Singh ve Horn (2000), Singh ve Deo (2003) yardımcı değişkende kayıp gözlem olmadığını varsayarken, Rueda vd. (2005) yardımcı değişkende de kayıp gözlem olması durumunu dikkate almıştır.

Sonlu *N* birimli kitleden, yerine koymadan basit rasgele örnekleme ile *n* birimlik örneklem çekilsin. Cevaplı birimlerin yer aldığı küme *A* ve cevapsızların kümesi A^c olsun. Buna göre eleman sayısı *r* olan *A* kümesinde, her $i \in A$ için y_i değeri gözlenmektedir. Bununla birlikte, $i \in A^c$ birimi için y_i değeri kayıptır ve bu birimler için imputasyon yöntemine başvurulabilir.

Ortalama imputasyon ve oransal imputasyon yöntemlerine göre kitle ortalaması tahmin edicileri sırasıyla aşağıda verilmiştir:

$$\bar{y}_m = \frac{1}{r}\sum_{i=1}^{r} y_i \tag{1}$$

$$\bar{y}_{oran} = \frac{\bar{y}_r}{\bar{x}_r}\bar{x}_n \tag{2}$$

Burada, $\bar{x}_n = (1/n)\sum_{i=1}^{n} x_i$, $\bar{x}_r = (1/r)\sum_{i=1}^{r} x_i$ ve $\bar{y}_r = (1/r)\sum_{i=1}^{r} y_i$ göstermektedir.

X yardımcı değişken bilgisini kullanan Singh ve Horn (2000) ve Singh ve Deo (2003), ortalama ve oransal imputasyon yöntemlerinden faydalanarak, sırasıyla aşağıdaki imputasyon yöntemini ve buna bağlı olarak ilgilenilen değişkenin kitle ortalaması ($\bar{Y}$) tahmin edicilerini önermişlerdir:

$$y_{.i} = \begin{cases} \dfrac{\gamma n y_i}{r} + (1-\gamma)\hat{b}x_i, & i \in A \\ (1-\gamma)\hat{b}x_i, & i \in A^c \end{cases} \tag{3}$$

$$\bar{y}_{SH} = \gamma\,\bar{y}_r + (1-\gamma)\bar{y}_r\frac{\bar{x}_n}{\bar{x}_r} \tag{4}$$

$$y_{.i} = \begin{cases} y_i, & i \in A \\ \bar{y}_r\left[n\left(\dfrac{\bar{x}_n}{\bar{x}_r}\right)^{\alpha} - r\right]\dfrac{x_i}{\sum_{i \in A} x_i}, & i \in A^c \end{cases} \tag{5}$$

$$\bar{y}_{SD} = \bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\alpha} \tag{6}$$

Burada $\bar{y}_r$, cevaplı birimlerin oluşturduğu *A* kümesindeki ilgilenilen değişkenin örneklem ortalamasını, $\bar{x}_r$ *A* kümesindeki yardımcı değişkeninin örneklem ortalamasını; $\bar{x}_n$, yardımcı değişkenin *n* birimli örnekleminin ortalamasını; $\hat{b} = \bar{y}_r / \bar{x}_r$; α ve γ, tahmin edicilerin, Hata Kareler Ortalamalarını (HKO) en küçükleyen sabitleri göstermektedir.

Singh-Horn ve Singh-Deo tahmin edicilerini ($\bar{y}_{SH}$ ve $\bar{y}_{SD}$) en küçükleyen sırasıyla γ_{opt} , α_{opt} değerleri ve bunlara bağlı olarak en küçük HKO eşitlikleri aşağıda verilmiştir:

$$\gamma_{opt} = 1 - \rho \frac{C_y}{C_x} \Rightarrow HKO_{\min}(\bar{y}_{SH}) = \left(\frac{1}{r} - \frac{1}{N}\right)\sigma_y^2 - \left(\frac{1}{r} - \frac{1}{n}\right)\rho^2\sigma_y^2, \tag{7}$$

$$\alpha_{opt} = \rho \frac{C_y}{C_x} \Rightarrow \quad HKO_{\min}(\bar{y}_{SD}) = \left(\frac{1}{r} - \frac{1}{N}\right)\sigma_y^2 - \left(\frac{1}{r} - \frac{1}{n}\right)B\sigma_{xy}. \tag{8}$$

Burada $C_x = \frac{\sigma_x}{\bar{X}}$ ve $C_y = \frac{\sigma_y}{\bar{Y}}$, sırasıyla yardımcı değişkene ve ilgilenilen değişkene ait kitle değişim katsayılarını; σ_x^2 ve σ_y^2, sırasıyla yardımcı değişkene ve ilgilenilen değişkene ait kitle varyanslarını; $\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$ kitle için yardımcı değişken ve ilgilenilen değişken arasındaki ilişki katsayısını ve $B = \frac{\sigma_{xy}}{\sigma_x^2}$ kitleye ait regresyon katsayısını göstermektedir.

Bu iki tahmin edicinin etkinlik yönünden birbirlerine göre üstünlükleri yoktur, *HKO* ifadeleri eşittir. Fakat bu iki tahmin edici de, (1) ve (2) de sırasıyla verilen ortalama ve oransal tahmin edicilerine göre daha etkin tahmin edicilerdir (Kadılar ve Çıngı, 2008).

Uygulamada her zaman ilgilenilen değişkene ilişkin tam gözlem yapılamadığı gibi yardımcı değişkene ilişkin de tam gözlem yapılamayabilir. Yardımcı değişkende kayıp gözlem olması durumunda, çekilen s örneklemini, s_1, s_2 ve s_3 olmak üzere üç parçada inceleyebiliriz. Buna göre, (n-p-r) birimli s_1 örneklem parçasında, tam gözlemin elde edilebildiği varsayılmaktadır. p birimli s_2 örneklem parçasında yardımcı değişkene ait gözlemler elde edilebilir. Fakat ilgilenilen değişkende buna karşılık gelen gözlemler kayıptır. Aynı şekilde, s_3 örneklem parçasında, ilgilenilen değişkene ait r gözlem bilgisi mevcut iken yardımcı değişkende bunlara karşılık değerler kayıptır. Yukarıda bahsedilen s örneklem birimlerinin görüntüsü, aşağıda gösterildiği gibidir:

$s_1=\{i \in s/ x_i, y_i$ elde edilebilir$\}$ → n-p-r birimli,
$s_2=\{i \in s/ x_i$ elde edilebilir, y_i kayıp$\}$ → p birimli,
$s_3=\{i \in s/ y_i$ elde edilebilir, x_i kayıp$\}$ → r birimli.

p ve r'nin 0<p, r<n/2 koşullarını sağlayan tamsayı oldukları varsayımı altında Rueda vd. (2005), yardımcı değişkende de kayıp gözlem olması durumunda, $\bar{Y}$ kitle ortalaması tahmini için aşağıdaki tahmin ediciyi önermişlerdir:

$$\bar{y}_{rueda} = k_1\bar{y}_{s1} + k_2\bar{y}_{s3} + k_3\hat{B}\,\bar{x}_{s2} \tag{9}$$

Burada $\bar{y}_{s1}$, s_1 örneklem parçasında ilgilenilen değişkene ait örneklem ortalamasını; $\bar{y}_{s3}$, s_3 örneklem parçasında ilgilenilen değişkene ait örneklem ortalamasını; $\bar{x}_{s2}$, s_2 örneklem parçasında yardımcı değişkene ait örneklem ortalamasını; $\hat{B}$, örnekleme ait regresyon katsayısını; $k_1 = \frac{n-p-r(N-p)}{N(n-p)}$, $k_2 = \frac{r(N-p)}{N(n-p)}$ ve $k_3 = \frac{p}{N}$, sabitleri göstermektedir. $\bar{y}_{rueda}$ tahmin edicisinin HKO eşitliği ise aşağıda verildiği gibidir:

$$\begin{aligned} HKO(\bar{y}_{rueda}) &\cong \sigma_y^2\left(k_1^2 a + k_2^2 b + 2k_1k_2c\right) \\ &+ B^2k_3^2\sigma_x^2 d + 2k_3B\sigma_{xy}\left(k_1e + k_2 f\right) \end{aligned} \tag{10}$$

Burada a, b, c, d, e, f katsayıları, N kitle birimine ve yukarıda ifade edilen n, p ve r örneklem birimlerine bağlı olarak aşağıdaki eşitliklerden elde edilmektedir:

$$a = \left(\frac{1}{n-p-r} - \frac{1}{N}\right),$$

$$b = \left(\frac{1}{r} - \frac{1}{N}\right),$$

$$c = \begin{cases} \frac{1}{n-p-r} - \frac{1}{N}, & r \le n-p-r \\ \frac{1}{r} - \frac{1}{N}, & r > n-p-r \end{cases}$$

$$d = \left(\frac{1}{p} - \frac{1}{N}\right),$$

$$e = \begin{cases} \dfrac{1}{n-p-r} - \dfrac{1}{N}, & p \le n-p-r \\ \dfrac{1}{p} - \dfrac{1}{N}, & p > n-p-r \end{cases}$$

$$f = \begin{cases} \dfrac{1}{p} - \dfrac{1}{N}, & p \ge \mathrm{r} \\ \dfrac{1}{r} - \dfrac{1}{N}, & p < \mathrm{r} \end{cases}$$

3. ÖNERİLEN TAHMİN EDİCİLER

Bu çalışma kapsamında kitle ortalaması tahmini için, yardımcı değişkende kayıp gözlem olmadığı varsayılmıştır. Buna göre, ilgilenilen değişkende *p* gözlem kayıptır. Daha önceki tanımlamalar ışığında, *r* birim cevaplı olmaktadır. Bu durumda, *n-p=r* eşitliği olacağı açıktır. Yardımcı değişkene ilişkin kitle ortalamasının bilindiği durumda, (6)'da verilen tahmin edici düzenlenerek, basit rasgele örneklemede aşağıdaki tahmin edici önerilmiştir:

$$\bar{y}_{öneri1} = \bar{y}_r \left(\frac{\bar{X}}{\bar{x}_r} \right)^{\alpha} \tag{11}$$

Burada $\bar{X} = \dfrac{\sum_{i=1}^{N} x_i}{N}$, yardımcı değişkene ait kitle ortalamasını göstermektedir. α, $HKO(\bar{y}_{öneri1})$ eşitliğini en küçük yapan sabittir.

Teorem 1: (11)'de verilen tahmin edicinin, α_{opt} için en küçük HKO eşitliği aşağıda verilmiştir:

$$\alpha_{opt} = \rho \frac{C_y}{C_x} \Rightarrow HKO_{\min}(\bar{y}_{öneri1}) = \left(\frac{1}{n-p} - \frac{1}{N} \right) \sigma_y^2 \left(1-\rho^2\right) \tag{12}$$

Tanıt: Birinci dereceden Taylor serisi yaklaşımı kullanılarak,

$$HKO(\bar{y}_{öneri1}) \cong \left(\frac{1}{n-p} - \frac{1}{N} \right) \sigma_y^2 + \alpha^2 \left(\frac{1}{n-p} - \frac{1}{N} \right) \frac{\bar{Y}^2}{\bar{X}^2} \sigma_x^2 - 2\alpha \left(\frac{1}{n-p} - \frac{1}{N} \right) \frac{\bar{Y}}{\bar{X}} \sigma_{xy} \tag{13}$$

yazılabilir (Ayrıntılar EK 1'de verilmiştir).

$\dfrac{\partial HKO(\bar{y}_{öneri1})}{\partial \alpha} = 0$ eşitliğinden α_{opt} değeri aşağıdaki gibi elde edilir:

$$\alpha_{opt} = \rho \frac{C_y}{C_x} \tag{14}$$

(14), (13)'de yerine yazıldığında (12)'de verilen *HKO* eşitliğine ulaşılır (Ayrıntılar EK 2'de verilmiştir).

(9)'da verilen tahmin edici, yardımcı değişkende kayıp gözlem olmadığı varsayımı altında düzenlenerek aşağıdaki ikinci tahmin edici önerilmiştir:

$$\bar{y}_{öneri2} = f_1 \bar{y}_{s1} + f_2 \hat{B} \bar{x}_{s2} \tag{15}$$

Burada, yardımcı değişkende kayıp gözlem olmadığı durumda örneklem iki parçada incelenmektedir. Gözlenen ilgili değişken birimlerinin oluşturduğu 1. örneklem parçası, *n-p* elemanlı, gözlenmeyen ilgili değişken birimlerinin oluşturduğu 2. örneklem parçası *p* elemanlıdır. $\bar{y}_{s1}$, 1. örneklem parçasındaki ilgilenilen değişkenin ortalamasını, $\bar{x}_{s2}$ ise 2. örneklem parçasındaki yardımcı değişkenin ortalamasını, $f_1 = \dfrac{N-p}{N}$ ve $f_2 = \dfrac{p}{N}$ göstermektedir.

Teorem 2: (15)'de verilen tahmin edicinin, *HKO* eşitliği aşağıda verilmiştir:

$$HKO(\bar{y}_{öneri2}) \cong f_1^2 \mu \sigma_y^2 + f_2^2 \lambda B^2 \sigma_x^2 + 2 f_1 f_2 \tau B \sigma_{xy} \tag{16}$$

Burada σ_{xy}, ilgilenilen değişken ile yardımcı değişken arasındaki kitle kovaryansını göstermekte ve μ, λ, τ katsayıları da aşağıdaki gibi tanımlanmaktadır:

$$\mu = \left(\frac{1}{n-p} - \frac{1}{N}\right),$$

$$\lambda = \left(\frac{1}{p} - \frac{1}{N}\right),$$

$$\tau = \begin{cases} \dfrac{1}{n-p} - \dfrac{1}{N} & , p \leq n/2 \\ \dfrac{1}{p} - \dfrac{1}{N} & , p > n/2 \end{cases}$$

Tanıt: $V(\bar{y}_{s1}) = \left(\frac{1}{n-p} - \frac{1}{N}\right)\sigma_y^2$, $V(\bar{x}_{s2}) = \left(\frac{1}{p} - \frac{1}{N}\right)\sigma_x^2$ ve

$$Cov(\bar{y}_{s1}, \bar{x}_{s2}) = \begin{cases} \left(\dfrac{1}{n-p} - \dfrac{1}{N}\right)\sigma_{xy} & , p \leq n/2 \\ \left(\dfrac{1}{p} - \dfrac{1}{N}\right)\sigma_{xy} & , p > n/2 \end{cases}$$

olmak üzere birinci dereceden Taylor serisi yaklaşım yöntemi kullanılarak,

$$HKO(\bar{y}_{öneri2}) \cong f_1^2 V(\bar{y}_{s_1}) + f_2^2 B^2 V(\bar{x}_{s2}) + 2 f_1 f_2 B Cov(\bar{y}_{s1}, \bar{x}_{s2})$$

yazılabilir. Burada varyans ve kovaryans ifadeleri yerine yazıldığında (16)'da verilen *HKO* eşitliği elde edilir.

4. ETKİNLİK KARŞILAŞTIRMASI

Burada, (11) ve (15)'de önerilen tahmin ediciler (6)'da verilen Singh ve Deo (2003) tahmin edicisi ile karşılaştırılmıştır: (11)'de önerilen birinci tahmin edicinin, (6)'da verilen Singh ve Deo (2003) tahmin edicisinden dolayısıyla aynı zamanda (4)'de verilen Singh ve Horn (2000) tahmin edicisinden her zaman daha etkin olduğu, (12)'de verilen *HKO* eşitliğinin (8)'de verilen *HKO* eşitliği ile teorik karşılaştırması yapılarak görülmüştür (Ayrıntılar Ek 3'de verilmiştir).

Teorem 3: (15)'de önerilen ikinci tahmin edici, $p<n/2$ varsayımı altında aşağıda verilen koşulda, (6)'da verilen Singh ve Deo (2003) tahmin edicisinden daha etkindir:

$$\rho^2 < \frac{n[N-(n-p)][2N-p]}{np(n-p) + N[N^2 + n(2N-n-p)]} \quad (17)$$

Tanıt: $\text{HKO}(\bar{y}_{\text{öneri2}}) < \text{HKO}(\bar{y}_{\text{SD}})$ eşitsizliğinde (16) ve (8) yazılarak,

$$\frac{(N-p)^2}{N^2}\left(\frac{1}{n-p} - \frac{1}{N}\right)\sigma_y^2 + \frac{p^2}{N^2}\left(\frac{1}{p} - \frac{1}{N}\right)\frac{\sigma_{xy}^2}{\sigma_x^2}$$
$$+2\frac{N-p}{N}\frac{p}{N}\left(\frac{1}{n-p} - \frac{1}{N}\right)\frac{\sigma_{xy}^2}{\sigma_x^2}$$
$$< \left(\frac{1}{n-p} - \frac{1}{N}\right)\sigma_y^2 - \left(\frac{1}{n-p} - \frac{1}{n}\right)\frac{\sigma_{xy}^2}{\sigma_x^2},$$

$$\frac{\sigma_{xy}^2}{\sigma_x^2}\left[\begin{array}{l} \dfrac{1}{n-p} - \dfrac{1}{n} + \dfrac{p^2}{N^2}\left(\dfrac{1}{p} - \dfrac{1}{N}\right) \\ +2\dfrac{p(N-p)}{N^2}\left(\dfrac{1}{n-p} - \dfrac{1}{N}\right) \end{array}\right]$$
$$< \sigma_y^2\left(\frac{1}{n-p} - \frac{1}{N}\right)\left[1 - \frac{(N-p)^2}{N^2}\right],$$

$$\rho^2\left[\begin{array}{l} \dfrac{1}{n-p} - \dfrac{1}{n} + \dfrac{p^2}{N^2}\left(\dfrac{1}{p} - \dfrac{1}{N}\right) \\ +2\dfrac{p(N-p)}{N^2}\left(\dfrac{1}{n-p} - \dfrac{1}{N}\right) \end{array}\right]$$
$$< \left(\frac{1}{n-p} - \frac{1}{N}\right)\left[1 - \frac{(N-p)^2}{N^2}\right]$$

elde edilir. Daha sonra gerekli sadeleştirmeler yapıldıktan sonra,

$$\rho^2\left[N^3 + 2N^2 n - Nn^2 - Nnp + np(n-p)\right]$$
$$< n\left[N - (n-p)\right](2N-p)$$

ve buradan

$$\rho^2 < \frac{n[N-(n-p)][2N-p]}{np(n-p) + N[N^2 + n(2N-n-p)]}$$

biçiminde bulunur.

5. SONUÇ

Bu çalışmada, Singh ve Horn (2000), Singh ve Deo (2003), Rueda vd. (2005) tarafından sunulan tahmin ediciler incelenerek, kayıp gözlem olması durumunda kitle ortalamasına ilişkin alternatif iki yeni tahmin edici önerilmiştir. Önerilen tahmin edicilere ait *HKO* eşitlikleri elde edilerek, bu tahmin ediciler teorik olarak Singh ve Horn (2000), Singh ve Deo (2003) tahmin edicileri ile karşılaştırılmıştır. (11)'de verilen birinci önerilen tahmin edici her zaman için daha etkin bir tahmin edici olarak bulunurken, (15)'de verilen ikinci tahmin edici, (17) koşulu altında literatürdeki tahmin edicilerden daha etkin olduğu Teorem 3 ile gösterilmeye çalışılmıştır. Bu çalışma ışığı altında bundan sonra yapılabilecek çalışmalar, kayıp gözlem olması durumunda farklı örnekleme yöntemleri için daha etkin kitle ortalaması tahmin edicileri elde etmeye yönelik olabilir.

EK 1

Taylor serisi açılımı

$$h\left(\hat{Y}_1,\hat{Y}_2,...,\hat{Y}_p\right)\cong h\left(Y_1,Y_2,...,Y_p\right)+\sum_{j=1}^{p} d_j\left(\hat{Y}_j-Y_j\right)+R\left(\hat{Y},Y\right)$$

biçiminde tanımlanmaktadır. Burada p, parametre sayısıdır ve $R\left(\hat{Y},Y\right)$ kalanı göstermekte olup birinci dereceden Taylor serisi yaklaşımında ihmal edilir. Önerilen tahmin edici için p=2 olmaktadır ve

$$d_j=\frac{\partial h(t)}{\partial \hat{Y}_j}\Big|_{t=T},\ h(t)=h\left(\bar{y}_r,\bar{x}_r\right)=\bar{y}_{öneri1}=\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)^{\alpha},$$

$$\hat{Y}_1=\bar{y}_r,\ \hat{Y}_2=\bar{x}_r.$$

Buna göre,

$$d_1=\frac{\partial h(t)}{\partial \hat{Y}_1}\Big|_{\bar{X},\bar{Y}}=\left(\frac{\bar{X}}{\bar{x}_r}\right)^{\alpha}\Big|_{\bar{X},\bar{Y}}=1$$

$$d_2=\frac{\partial h(t)}{\partial \hat{Y}_2}\Big|_{\bar{X},\bar{Y}}=\bar{y}_r\bar{X}^{\alpha}\frac{-\alpha}{\bar{x}_r^{\alpha+1}}\Big|_{\bar{X},\bar{Y}}=-\alpha\frac{\bar{Y}}{\bar{X}}$$

olarak bulunur. Değerler yerine yazıldığında,

$$\bar{y}_{öneri1}\cong\bar{Y}+\left(\bar{y}_r-\bar{Y}\right)-\alpha\frac{\bar{Y}}{\bar{X}}\left(\bar{x}_r-\bar{X}\right),$$

$$\bar{y}_{öneri1}-\bar{Y}\cong\left(\bar{y}_r-\bar{Y}\right)-\alpha\frac{\bar{Y}}{\bar{X}}\left(\bar{x}_r-\bar{X}\right)$$

olur ve her iki tarafın karesi alınarak beklenen değer alındığında,

$$E\left(\bar{y}_{öneri1}-\bar{Y}\right)^2\cong V\left(\bar{y}_r\right)+\alpha^2\frac{\bar{Y}^2}{\bar{X}^2}V\left(\bar{x}_r\right)-2\alpha\frac{\bar{Y}}{\bar{X}}Cov\left(\bar{x}_r,\bar{y}_r\right)$$

biçiminde elde edilir. Burada,

$$V\left(\bar{y}_r\right)=\left(\frac{1}{n-p}-\frac{1}{N}\right)\sigma_y^2,$$

$$V\left(\bar{x}_r\right)=\left(\frac{1}{n-p}-\frac{1}{N}\right)\sigma_x^2 \text{ ve}$$

$$Cov\left(\bar{x}_r,\bar{y}_r\right)=\left(\frac{1}{n-p}-\frac{1}{N}\right)\sigma_{xy}$$

eşitlikleri yerine yazıldığında,

$$HKO\left(\bar{y}_{öneri1}\right)\cong\left(\frac{1}{n-p}-\frac{1}{N}\right)\sigma_y^2$$
$$+\alpha^2\frac{\bar{Y}^2}{\bar{X}^2}\left(\frac{1}{n-p}-\frac{1}{N}\right)\sigma_x^2$$
$$-2\alpha\frac{\bar{Y}}{\bar{X}}\left(\frac{1}{n-p}-\frac{1}{N}\right)\sigma_{xy}$$

elde edilir.

EK 2

$HKO(\bar{y}_{öneri1})$ ifadesinde $\bar{Y}^2$ ortak parantezine alınırsa,

$$HKO\left(\bar{y}_{öneri1}\right)\cong\bar{Y}^2\left\{\begin{array}{l}\left(\frac{1}{n-p}-\frac{1}{N}\right)\frac{\sigma_y^2}{\bar{Y}^2}\\+\alpha^2\left(\frac{1}{n-p}-\frac{1}{N}\right)\frac{\sigma_x^2}{\bar{X}^2}\\-2\alpha\left(\frac{1}{n-p}-\frac{1}{N}\right)\frac{\sigma_{xy}}{\bar{Y}\bar{X}}\end{array}\right\}$$

$$= \bar{Y}^2 \left\{ \begin{array}{l} \left(\frac{1}{n-p} - \frac{1}{N}\right) C_y^2 \\ +\alpha^2 \left(\frac{1}{n-p} - \frac{1}{N}\right) C_x^2 \\ -2\alpha \left(\frac{1}{n-p} - \frac{1}{N}\right) \rho C_x C_y \end{array} \right\}$$

biçiminde elde edilir. Burada $C_x = \frac{\sigma_x}{\bar{X}}$, $C_y = \frac{\sigma_y}{\bar{Y}}$ ve $\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$ göstermektedir.

Optimal α değeri için,

$$\frac{\partial HKO(\bar{y}_{öneri1})}{\partial \alpha} = \bar{Y}^2 \left\{ \begin{array}{l} 2\alpha \left(\frac{1}{n-p} - \frac{1}{N}\right) C_x^2 \\ -2\left(\frac{1}{n-p} - \frac{1}{N}\right) \rho C_x C_y \end{array} \right\} = 0$$

eşitliği çözüldüğünde,

$$\alpha_{\mathrm{opt}} = \rho \frac{\mathrm{C_y}}{\mathrm{C_x}}$$

olarak bulunur. Bu ifade $HKO(\bar{y}_{öneri1})$'de yerine yazılırsa,

$$HKO_{\min}(\bar{y}_{öneri1}) \cong \left(\frac{1}{n-p} - \frac{1}{N}\right) \sigma_y^2$$
$$+\rho^2 \frac{C_y^2}{C_x^2} \frac{\bar{Y}^2}{\bar{X}^2} \left(\frac{1}{n-p} - \frac{1}{N}\right) \sigma_x^2$$
$$-2\rho \frac{C_y}{C_x} \frac{\bar{Y}}{\bar{X}} \left(\frac{1}{n-p} - \frac{1}{N}\right) \sigma_{xy}$$

elde edilir. Gerekli sadeleştirmeler yapıldığında

$$HKO_{\min}(\bar{y}_{öneri1}) = \left(\frac{1}{n-p} - \frac{1}{N}\right) \sigma_y^2 (1-\rho^2)$$

biçiminde elde edilir.

EK 3

$HKO_{\min}(\bar{y}_{öneri1}) < HKO_{\min}(\bar{y}_{SD})$ eşitsizliğinde (12) ve (8) yerine yazılarak,

$$\left(\frac{1}{n-p} - \frac{1}{N}\right) \sigma_y^2 (1-\rho^2)$$
$$< \left(\frac{1}{r} - \frac{1}{N}\right) \sigma_y^2 - \left(\frac{1}{r} - \frac{1}{n}\right) B\sigma_{xy}$$

elde edilir. Yardımcı değişkende kayıp gözlem olmadığı varsayımı altında, tanımlamalara göre *n-p=r* olmaktadır. Buna göre,

$$\left(\frac{1}{n-p} - \frac{1}{n}\right) B\sigma_{xy} < \left(\frac{1}{n-p} - \frac{1}{N}\right) \sigma_y^2$$
$$-\left(\frac{1}{n-p} - \frac{1}{N}\right) \sigma_y^2 (1-\rho^2)$$

$$\left(\frac{1}{n-p} - \frac{1}{n}\right) \sigma_y^2 \rho^2$$
$$< \left(\frac{1}{n-p} - \frac{1}{N}\right) \sigma_y^2 (1-1+\rho^2)$$

$$\left(\frac{1}{n-p} - \frac{1}{n}\right) \sigma_y^2 \rho^2 < \left(\frac{1}{n-p} - \frac{1}{N}\right) \sigma_y^2 \rho^2$$

$$\left(\frac{1}{n-p} - \frac{1}{n}\right) < \left(\frac{1}{n-p} - \frac{1}{N}\right)$$

$$\frac{1}{n} > \frac{1}{N}$$

ve buradan,

$$n < N$$

koşulu elde edilir. Bu koşul her zaman sağlandığından, teorik olarak her zaman için $\bar{y}_{öneri1}$ tahmin edicisinin $\bar{y}_{SD}$ tahmin edicisinden daha etkin olduğu söylenebilir.

KAYNAKLAR

Gamrot, W. (2007). Mean value estimation using two-phase samples with missing data in both phases. *Acta Applicandae Mathematica* 96, 215-220.

Little, R.J.A. Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, Wiley-Interscience, New Jersey, USA.

Rueda, M.M., Gonzalez, S., Arcos, A., Roman, Y., Martinez, M.D. ve Munoz, J.F. (2005). Estimation of the population mean using auxiliary information when some observations are missing*, Proceedings of XIth International Symposium on Applied Stochastic Models and Data Analysis (ASMDA 2005)*, ss. 1493-1498, Brest, France, 17-20 May 2005.

Kadılar, C. ve Çıngı, H. (2008). Estimators for the Population Mean in the Case of Missing Data. *Communications in Statistics: Theory and Methods* 37, 14, 2226-2236.

Singh, S. ve Deo, B. (2003). Imputation by power transformation*, Statistical Papers* Vol.44, No.4, 555-579.

Singh, S. ve Horn, S. (2000). Compromised imputation in survey sampling. *Metrika* 51, 267-276.