Süleyman Demirel Üniversitesi
Fen Bilimleri Enstitüsü Dergisi

Suleyman Demirel University
Journal of Natural and Applied Science

18(2), 54-63, 2014

Cevaplamama Hatası ve Ortalama Tahmini Üzerindeki Etkisinin İncelenmesi: İki Alt Grup Çalışması

Cenker Burak METİN[1], Yaprak Arzu ÖZDEMİR*[2]
[1]TÜİK, Örnekleme ve Analiz Teknikleri Daire Başkanlığı, 06100, Ankara
[2]Gazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06500, Ankara

(Alınış Tarihi: 25.11.2013, Kabul Tarihi: 08.09.2014)

Anahtar Kelimeler
Cevaplamama hatası
Cevaplamama sapması
Rasgele cevap modeli
Hata kare ortalaması
Göreli sapma.

Özet: Araştırmada örneğe alınan birimlerden çeşitli nedenlerden dolayı ölçüm alınamaması sonucunda cevaplamama hatası ortaya çıkar. İlgilenilen değişken bakımından ortak özelliğe sahip bazı gruplarda oluşan cevaplamama hatası tahmin edicilerde cevaplamama sapmasına neden olur. Bethlehem (2002), Särndal ve Lundström (2006) cevaplamama sapmasını birimlerin cevap verme eğilimlerine bağlı olarak tanımlamışlardır. Bu çalışmada, birimlerin cevap verme eğilimlerinden yararlanarak, yığın ortalamasının tahmin edilmek istendiği bir araştırma için cevaplamama hatasının ortalama tahmini üzerindeki etkisi, tahmin edicinin hata kare ortalaması (HKO) ve cevaplamama sapması bakımından incelenmiştir. Cevaplamama hatası durumunda tahmin edicinin hata kare ortalamasını teorik olarak elde etmek mümkün olmadığından Monte Carlo simülasyon çalışması yapılmıştır. Simülasyon çalışması ile farklı yığın yapıları, örnek çapları ve farklı cevap verme eğilimleri altında ortalamaya ilişkin hata kare ortalaması ve cevaplamama sapma değerleri elde edilmiştir. Elde edilen simülasyon sonuçlarından; ortalama tahminindeki cevaplamama sapmasının, cevaplama oranı yüksek ve iki alt grup arası cevap verme eğilimi farklılığının büyük olduğu yığınlarda bu oranın, cevap verme eğilimi farklılığının düşük olduğu yığınlara göre daha büyük olduğu görülmüştür.

Nonresponse Error and Examining the Effects on The Estimation of Population Mean: Two Subpopulation Study

Keywords
Nonresponse error
Nonresponse bias
Random response model
Mean square error
Relative bias

Abstract: Nonresponse error appears due to various reasons as a result of lack of measurement for taken units in survey. Nonresponse error which takes place in some groups has common characteristics with respect to target variable causes nonresponse bias for estimators. Bethlehem (2002), Särndal and Lundström (2006) defined nonresponse bias subject to response propensities. The purpose of this study is to examine the effect of nonresponse error on the estimation of population mean with respect to mean square error (MSE) and nonresponse bias by using response propensities. Monte Carlo simulation was conducted since in theoretical way it is impossible to evaluate mean square error of the estimator when nonresponse occurs. By simulation study, mean square error and nonresponse bias values observed for the estimation of population mean at different population structures, sample sizes and response propensities. It is recognized by simulation results indicated that when response rate is high and the difference of propensities between two sub groups is large in a population, nonresponse bias for estimation of population mean is greater when it is compared with the case in which the response rate is low and the difference of propensities between two groups is small.

* *İlgili yazar: yaprak@gazi.edu.tr*

1. Giriş

Bilimsel araştırmalarda amaç, yığının bilinmeyen bazı karakteristikleri hakkında bilgi elde etmektir. Bu amaçla yapılacak bir araştırma, başlangıçta eksiksiz olarak planlansa da, uygulama aşamasında bazı zorluklarla karşılaşılabilir. Bu zorluklar çeşitli hataların ortaya çıkmasına neden olur. Araştırmalarda karşılaşılan bu hatalar, örnekleme hatası ve örnekleme dışı hatalar olmak üzere iki gruba ayrılır. Hataların bu iki başlık altında detaylı sınıflandırması, Bethlehem (1999) tarafından yapılmıştır. Örnekleme dışı hatalardan biri olan cevaplamama hatası hemen her araştırmada karşılaşılan bir durumdur. Cevaplamama hatası, araştırmada örneğe alınan birimlerden çeşitli nedenlerden dolayı ölçüm alınamaması veya diğer bir ifade ile cevaplama yapılmaması durumunda ortaya çıkar (Särndal ve Lundström, 2006). Birimle iletişim kurulamaması, birimin cevap vermek istememesi, cevap verecek yeterli bilgiye sahip olmaması veya birimin erişilmez olması cevaplamamanın en önemli nedenleridir. Cevaplamama genel olarak birim cevaplamama ve madde cevaplamama olmak üzere ikiye ayrılır.

<u>Birim cevaplamama:</u> Örneğe alınan birim hiçbir bilgi vermez, yani soru formundaki maddelerin hiçbiri cevaplanmazsa buna birim cevaplamama denir.

<u>Madde cevaplamama:</u> Örneğe alınan birim soru formunda yer alan bazı maddelere cevap verirken, bazılarına cevap vermezse (özellikle hassas sorulara) soru formu kısmen cevaplanmış olur ve bu duruma madde cevaplamama adı verilir.

Birim cevaplamama, araştırmanın başlangıçta planlanan örnek çapından daha küçük bir örnek çapı ile sonuçlanmasına neden olur. Eğer cevaplamama, yığındaki birimler arasında rastgele gerçekleşmişse, tahmin edicilerin varyansında bir artış söz konusu olacak, ancak yine geçerli tahminler elde edilebilecektir. Burada asıl problem, yığında yer alan ve ilgilenilen değişken bakımından ortak özelliğe sahip bazı gruplarda oluşan cevaplamamadır. Bu durumda yığındaki bazı gruplar örnekte daha az veya daha fazla temsil edilecektir. Buna seçici cevaplamama adı verilir (Bethlehem vd., 2011). Bu tür bir cevaplamama sonucunda yığın parametrelerinin tahminleri sapmalı olacaktır. Örneğin yüksek gelir seviyesine sahip bireyler gelirlerini beyan etmeme veya daha az beyan etme eğilimindedir. Bu durumda, elde edilecek tahmin değeri yığın parametresinin gerçek değerinin altında bir tahmin verecektir. Bunun sonucunda başlangıçta belirlenen güvenilirlik seviyesi geçerli olmayacaktır.

Aksi durum ispatlanmadıkça, araştırmalarda genellikle, birim veya madde cevaplamama durumlarında, seçici cevaplamamanın da olduğu ve sapmalı tahminlerin elde edildiği varsayılır. Bethlehem ve Kersten (1985) seçici cevaplamamanın ele alındığı çeşitli araştırma örnekleri vermişlerdir. Bunun yanı sıra, cevaplamama sapması ile ilgili çok sayıda çalışma bulunmaktadır. Groves (1989), ortalamaya ilişkin cevaplamama sapması için cevaplayan ve cevaplamayanların örnek ortalaması farklarına dayalı olarak bir formül vermiştir. Bu formüle göre, cevaplama oranı arttıkça cevaplamama sapması da azalmaktadır. Ayrıca Groves (1989), bireylerin cevaplama eğilimlerinin aynı olduğunu varsaymaktadır. Bethlehem (2002) ise ortalamaya ilişkin cevaplamama sapmasını, cevaplama eğilimi ile araştırma değişkeni arasındaki kovaryansın ortalama cevaplama eğilimine oranı olarak tanımlamıştır. Ancak bu formülle cevaplama oranı arttıkça cevaplamama sapmasının azalacağını söylemek her zaman mümkün değildir. Ayrıca Groves (2006) ile Groves ve Peytcheva (2008) cevaplamama sapması ile ilgili çalışmaların bir meta analizini yapmışlardır. Bu çalışma sonucunda cevaplama oranının cevaplamama sapması için zayıf bir gösterge olduğuna ve cevaplamama sapmasının araştırmada tahmin edilecek parametreye göre değiştiğine işaret etmişlerdir. Peytchev (2013) cevaplamamanın ortaya çıkardığı sonuçları yapılan araştırmalar üzerinden detaylı biçimde ele almıştır. Särndal ve Lundström (2006) cevaplamama sapmasını birimlerin cevap verme eğilimlerine bağlı olarak incelemiştir.

Bu çalışmada, Särndal ve Lundtsröm (2006)'un yaklaşımından yararlanarak yığın ortalamasının tahmin edilmek istendiği araştırmalarda, farklı yığın yapıları ve farklı cevap verme eğilimleri için cevaplamama etkileri, cevaplamama sapması ve HKO bakımından incelenmiştir. Çalışmanın ikinci bölümünde cevaplamama mekanizması ve cevap verme eğilimi, üçüncü bölümde ise rasgele cevap modeli tanıtılmıştır. Ayrıca birimlerin cevap verme eğilimlerine bağlı olarak ortalamaya ilişkin tahmin edici verilmiştir. Dördüncü bölümde ise tahmin edicinin cevaplamama durumunda ortaya çıkan sapma ve hata kare ortalama değerleri çeşitli örnek çapları için simetrik, sağa ve sola çarpık dağılımlarda Monte Carlo simülasyon çalışması ile elde edilmiştir. Son bölümde simülasyon çalışması sonucu elde edilen bulgular değerlendirilmiştir.

2. Cevap Verme Eğilimi ve Cevaplamama Mekanizması

Cevaplamama hatasının yığın ortalamasına ilişkin tahmin edici üzerindeki etkisini ortaya çıkarmak amacıyla iki ayrı cevap modeli yaklaşımı benimsenmiştir. Bu modeller rasgele cevap modeli ve sabit cevap modeli olarak adlandırılır (Lindström vd., 1979; Cassel vd., 1983). Rasgele ve sabit cevap modelleri yardımı ile hangi koşullar altında cevaplamama sapmasının ortaya çıktığı daha açık olarak görülebilir. Sabit cevap modelinde yığının cevaplayan ve cevaplamayanlar olmak üzere iki tabakaya ayrıldığı varsayılır. Bu durumda cevaplayanlar tabakasından herhangi bir birim örneğe çıktığında takdirde sorulara cevap verilmiş, cevaplamayanlar tabakasından herhangi bir birim örneğe çıktığında ise cevap verilmemiş olacaktır. Yığın ortalamasına ilişkin tahmin edicinin sapma miktarı, cevaplayan ve cevaplamayanların tabaka yığın ortalamaları arasındaki fark ve cevaplamayan birim sayısının yığındaki birim sayısına oranı arttıkça artmaktadır.

Bu çalışmada cevap verme eğilimine dayalı olan rasgele cevap modeli ele alınmış ve bu modele geçmeden önce cevap verme eğilimi ve cevaplamama mekanizması tanıtılmıştır.

Araştırmalarda birimin cevaplamama durumu, birimin cevap verme eğilimine bağlı olarak ortaya çıkar. Birimin cevap verme eğilimi ise ilgilenilen değişken Y ve bilinen yardımcı değişken vektörü $\underline{X}$'e göre değişir. Cevaplama gösterge değişkeni R olmak üzere,

$$R = \begin{cases} 1 \text{ birim cevaplarsa} \\ 0 \text{ birim cevaplamazsa} \end{cases} \tag{1}$$

şeklinde tanımlanır. Buna göre i. birimin ilgilenilen değişken değerine ve yardımcı değişken bilgisine bağlı olarak cevap verme eğilimi,

$$\rho_{x,y} = \Pr\left[R_{i=}1|\underline{X}_{i=}\underline{x}\,, Y_{i=}y\right] \tag{2}$$

olarak tanımlanır. $\rho_{x,y}$, aynı zamanda birimin $\underline{x}$ ve y değerleri bilindiğinde cevap verme olasılığı olarak adlandırılır (Schouten, 2007). Cevap verme eğilimini modellemek amacıyla cevaplamama mekanizmasından yararlanılır. Cevaplamama mekanizması için üç önemli varsayım önerilmiştir (Little ve Rubin 2002). Bu varsayımlar ve temel özellikleri kısaca aşağıdaki gibi özetlenmiştir.

Tamamen Rasgele Kayıp (TRK): $\rho_{x,y}$; $\underline{X}$ ve Y değişkenlerinden, diğer bir ifadeyle araştırmadan tamamen bağımsızsa cevaplamama mekanizması Tamamen Rasgele Kayıp olarak adlandırılır. Bu varsayım altında seçici cevaplamama yoktur ve tahminler sapmasızdır. Bu varsayım altında $\rho_{x,y}$ deterministiktir. Yani, $\rho_{x,y} = \rho_R$ olup, bu tüm $\underline{x}$ ve y değerleri için sabittir.

Rasgele Kayıp (RK): $\rho_{x,y}$; $\underline{X}$'e bağlı, Y'ye bağlı değilse cevaplamama mekanizması Rasgele Kayıp olarak adlandırılır. RK varsayımı TRK varsayımından daha zayıftır ve seçici cevaplamama vardır. Bu varsayım altında elde edilen tahminler sapmalı olacaktır. Ancak sapmasız tahminleri elde etmek için yardımcı değişken bilgisi kullanılabilir. Ayrıca, $\rho_{x,y} = \rho_x$ olup, belirli $\underline{x}$ değerleri için sabittir.

Rasgele Olmayan Kayıp (ROK): $\rho_{x,y}$; tamamen ilgilenilen değişken Y'ye bağlı ve bu ilişki yardımcı değişken $\underline{X}$ tarafından açıklanamıyorsa, cevaplamama mekanizması Rasgele Olmayan Kayıp olarak adlandırılır. Bu varsayım altında elde edilen tahminler sapmalıdır.

3. Rasgele Cevap Modeli

Rasgele cevap modelinde yığındaki her birimin bilinmeyen bir cevap verme eğiliminin, ρ_i'ye sahip olduğu varsayılır. Bu çalışmada, sapmasız tahminleri elde etmek için yardımcı değişken bilgisini kullanabilmesi nedeniyle, cevap verme eğiliminin RK cevaplamama mekanizması ile modellendiği varsayılmıştır. Bu modele göre, Eşitlik (1)'den, R_i, i=1,2,…,N olmak üzere, i. birim için cevap verme göstergesini ifade eder. Bu durumda i. birim için cevap verme eğilimi, RK modelinden yararlanarak

$$\rho_{ix} = \Pr\left[R_{i=}1|\underline{X}_{i=}\underline{x}\right] \tag{3}$$

eşitliği ile tanımlanır.

Söz konusu yığından iadesiz n çaplı rassal bir örnek alınmak üzere, a_i , i=1,2,…,N, i. birimin örneğe alınma göstergesi olacaktır. i. birim örneğe alınmışsa $a_i = 1$, örneğe alınmamışsa $a_i = 0$ değerini alır. Buna göre örnekteki cevaplayan ve cevaplamayan birim sayısı sırasıyla,

$$n_C = \sum_{i=1}^{N} a_i R_i$$

$$n_{CS} = \sum_{i=1}^{N} a_i (1 - R_i)$$

olur. Toplam örnek çapı ise $n=n_C+n_{CS}$ dir. Burada önemli olan nokta cevaplayan ve cevaplamayan birim sayılarının da rassal değişken olmasıdır. Buna göre yığın ortalamasına ilişkin tahmin edici, cevaplayan birimlerin örnek ortalamasından yararlanarak,

$$\bar{y}_C = \frac{1}{n_C} \sum_{i=1}^{N} a_i R_i\, Y_i \tag{4}$$

şeklinde tanımlanır. Tahmin edicinin beklenen değeri ise yaklaşık olarak

$$E(\bar{y}_c) \approx \tilde{Y} \tag{5}$$

olur. Burada,

$$\tilde{Y} = \frac{1}{N} \sum_{i=1}^{N} \frac{\rho_{ix}}{\bar{\rho}} Y_i$$

ve

$$\bar{\rho} = \frac{1}{N} \sum_{i=1}^{N} \rho_{ix}$$

olarak tanımlanır. $\bar{\rho}$, yığındaki birimler için cevap verme eğilimlerinin ortalamasıdır. $\bar{\rho}$ aynı zamanda cevaplama oranı olarak da adlandırılır. Eşitlik (5)'de görüldüğü gibi cevaplayan birimlerin ortalamalarının beklenen değeri tahmin edilen yığın ortalamasına eşit değildir. Bu nedenle tahmin edici sapmalıdır ve cevaplamama sapması yaklaşık olarak aşağıdaki gibi ifade edilir.

$$B(\bar{y}_c) = \widetilde{Y} - \bar{Y} = \frac{R_{\rho Y} S_\rho S_Y}{\bar{\rho}} \tag{6}$$

Burada $R_{\rho Y}$, ilgilenilen değişken ile cevap verme eğilimi arasındaki korelasyonu, S_Y, Y'nin standart sapmasını, S_ρ ise cevap verme eğilimlerinin standart sapmasını gösterir. Eşitlik (6)'dan yararlanarak aşağıdaki çıkarsamalar yapılabilir.

1) İlgilenilen değişken ile cevap verme eğilimi arasında ilişki yoksa $(R_{\rho Y} = 0)$ sapma sıfırdır. İlişki güçlendikçe sapma artar.

2) Tüm cevap verme eğilimleri eşitse $(S_\rho = 0)$ sapma sıfırdır. Bu durumda seçici cevaplamama durumu ortaya çıkmaz ve bu durum sadece örnek çapının azalmasına neden olur.

3) $R_{\rho Y}$ nin değişmediği varsayımı altında, cevap verme eğilimlerinin ortalaması azaldıkça sapma artar. Uygulamada bu durum düşük cevaplama oranlarının daha büyük sapmalara sebep olacağı anlamına gelir (Bethlehem vd., 2011).

Eşitlik (6)'dan aynı cevaplama oranı altında farklı cevap verme eğilimlerinin etkisini görmek mümkün olmamaktadır. Bu nedenle simülasyon çalışması yapılarak, hem sapma üzerinde cevap verme eğilimlerinin etkisi hem de HKO değerleri elde edilerek, tahmin edici üzerinde cevaplamama etkisi tespit edilmeye çalışılmıştır.

4. Simülasyon Çalışması

Rasgele cevaplama modeli yaklaşımında, cevaplamama hatası ile oluşan sapmanın hesaplanabilmesi için Eşitlik (6)'dan yararlanılmıştır. Ancak cevaplamama hatasının tahmin edici üzerine etkisini ortaya çıkarmak bakımından, tahmin edicinin HKO değeri elde edilmelidir. Buna göre sapma miktarının HKO değerindeki payının büyüklüğü ile cevaplamama etkisinin büyüklüğü tespit edilebilir. Diğer yandan rasgele cevap modelinde $\bar{y}_c$ tahmin edicisinin HKO'nın hesaplanmasına ilişkin teorik bir eşitlik bulunmamaktadır. Bu nedenle Monte-Carlo simülasyon çalışmasından yararlanılmıştır. Monte Carlo simülasyon çalışması için gerekli program R 2.15 programında yazılmıştır. Simülasyon çalışmasında, cevap eğiliminin HKO ve sapma değeri üzerine etkisini daha açık biçimde görmek için yığının ilgilenilen değişken bakımından benzer özelliklere sahip iki ayrı gruptan oluştuğu varsayılmıştır. Buna göre ilgilenilen değişken (Y), iki değer alan bir yardımcı değişkenden (X) etkilenecektir ve i. birim için cevap verme eğilimi ρ_{ix}, ρ_1 ve ρ_2 olmak üzere iki değerden oluşacaktır.

Simülasyon çalışmasında, cevaplama oranları $\bar{\rho}$ =%70, %80, %90 ve %100 (tam cevaplı) olmak üzere, bu değerleri sağlayan farklı cevap verme eğilimleri ($\rho_1=\rho_2$ ve $\rho_1\neq\rho_2$ için) n=500, 1000, 2000 olan örnek çapları ele alınmıştır. Simetrik, sağa ve sola çarpık dağılımları temsil etmek üzere, sırasıyla Normal, Gamma ve Beta dağılımlarından elde edilen verilere dayalı olarak, ortalama tahmin edicisinin HKO, varyans ve sapma değerleri elde edilmiştir. Ayrıca sapma miktarının HKO'ndaki payını ortaya çıkarmak bakımından göreli sapma değerleri de hesaplanmıştır. Göreli sapma

$$Göreli\ Sapma = 100 * \frac{B(\bar{y}_c)^2}{HKO(\bar{y}_c)} \tag{7}$$

eşitliği ile tanımlanmıştır.

Normal dağılımda $Y_1 \sim N(2500,600)$ ve $Y_2 \sim N(2000,600)$ olmak üzere eşit varyanslı iki grup ve $Y_1 \sim N(2500,600)$ ve $Y_2 \sim N(2000,1500)$ olmak üzere farklı varyanslı iki grup dikkate alınmıştır. Gamma dağılımında $Y_1 \sim Gamma(2,3)$, $Y_2 \sim Gamma(2,2)$ ve Beta dağılımında $Y_1 \sim Beta(9,3)$, $Y_2 \sim Beta(9,2)$ parametreleri dikkate alınmıştır. Bu dağılımlardan n=500, 1000 ve 2000 çaplı örnekler alınarak, 1000000 tekrarlı Monte Carlo simülasyon çalışması ile $\bar{y}_c$ tahmin edicisi ve tahmin edicinin HKO, varyans, sapma ve göreli sapma değerleri elde edilmiştir. Gamma ve Beta dağılımından üretilen veriler oldukça düşük değerli olduklarından HKO ve göreli sapma değerlerini kolaylıkla karşılaştırmak için veriler 1000 değeri ile çarpılmıştır. Elde edilen sonuçlar Tablo 1-4 de verilmiştir.

Tablo 1. $Y_1 \sim N(2500,600)$, $Y_2 \sim N(2000,600)$ olmak üzere ortalama tahmin edicisinin HKO, sapma, varyans ve göreli sapma değerleri

		n=500				**n=1000**				**n=2000**			
		HKO	Sapma	Varyans	Göreli Sapma	HKO	Sapma	Varyans	Göreli Sapma	HKO	Sapma	Varyans	Göreli Sapma
$\bar{\rho}=1$	$\rho_1=\rho_2=1$	819,5692	-0,0082	819,5691	0	374,662	0,0244	374,6614	0,0002	166,4523	0,0211	166,4518	0,0003
$\bar{\rho}=0.7$	$\rho_1=0.5$ $\rho_2=0.9$	6275,989	-73,7089	842,9841	86,5681	5630,645	-72,1962	418,3491	92,5701	5612,811	-73,7542	173,129	96,9155
	$\rho_1=0.6$ $\rho_2=0.8$	2169,779	-36,7569	818,7084	62,2677	1588,385	-34,8158	376,2479	76,3126	1398,402	-35,0626	169,0183	87,9135
	$\rho_1=0.7$ $\rho_2=0.7$	1189,08	0,0035	1189,08	0	552,0083	0,0463	552,0062	0,0004	255,6198	0,0098	255,6197	0
$\bar{\rho}=0.8$	$\rho_1=0.65$ $\rho_2=0.95$	2877,022	-45,287	826,1123	71,2859	2921,111	-50,2332	397,7352	86,3841	2299,684	-46,1759	167,469	92,7177
	$\rho_1=0.7$ $\rho_2=0.9$	1796,468	-31,3284	814,9992	54,6333	1399,28	-31,841	385,43	72,4551	1065,699	-29,9206	170,4596	84,0049
	$\rho_1=0.75$ $\rho_2=0.85$	1086,545	-16,2581	822,2181	24,3273	609,2229	-15,3264	374,3251	38,557	426,267	-16,0577	168,4171	60,4902
	$\rho_1=0.8$ $\rho_2=0.8$	1026,897	0,0184	1026,897	0	478,9631	-0,0206	478,9627	0,0001	216,2103	0,0015	216,2103	0
$\bar{\rho}=0.9$	$\rho_1=0.8$ $\rho_2=1$	1661,444	-29,2543	805,6273	51,5104	1119,975	-27,096	385,783	65,5543	931,0403	-27,6018	169,1798	81,829
	$\rho_1=0.85$ $\rho_2=0.95$	976,3757	-13,2145	801,7537	17,8847	556,0694	-13,7752	366,3136	34,1245	356,719	-13,5829	172,2241	51,72
	$\rho_1=0.9$ $\rho_2=0.9$	893,6387	0,0286	893,6379	0,0001	429,3543	0,0327	429,3533	0,0002	192,5299	-0,0248	192,5293	0,0003

Tablo 2. $Y_1 \sim N(2500,600)$, $Y_2 \sim N(2000,1500)$ olmak üzere ortalama tahmin edicisinin HKO, sapma, varyans ve göreli sapma değerleri

		n=500				**n=1000**				**n=2000**			
		HKO	Sapma	Varyans	Göreli Sapma	HKO	Sapma	Varyans	Göreli Sapma	HKO	Sapma	Varyans	Göreli Sapma
$\bar{\rho}=1$	$\rho_1=\rho_2=1$	2589,442	0,0356	2589,441	0	1216,564	0,0279	1216,563	0,0001	555,0772	0,0412	555,0755	0,0003
$\bar{\rho}=0.7$	$\rho_1=0.5$ $\rho_2=0.9$	8787,851	-70,6745	3792,973	56,8385	7982,612	-78,8272	1768,885	77,8408	5983,301	-71,8925	814,7724	86,3826
	$\rho_1=0.6$ $\rho_2=0.8$	4432,664	-33,8545	3286,54	25,8563	2813,572	-36,5678	1476,371	47,5268	1906,758	-35,2401	664.8956	65,1295
	$\rho_1=0.7$ $\rho_2=0.7$	3765,293	0,0521	3765,29	0,0001	1798,466	0,0214	1798,466	0	852,7065	-0,0122	852,7064	0
$\bar{\rho}=0.8$	$\rho_1=0.65$ $\rho_2=0.95$	5297,352	-44,1351	3349,442	36,7714	4025,302	-49,5293	1572,148	60,9434	2642,337	-44,1316	694,7408	73,7073
	$\rho_1=0.7$ $\rho_2=0.9$	3996,854	-31,017	3034,802	24,0703	2321,039	-30,0721	1416,706	38,9624	1586,668	-30,567	652,324	58.8872
	$\rho_1=0.75$ $\rho_2=0.85$	3062,519	-15,0135	2837,114	7,3601	1569,492	-15,2861	1335,826	14,888	870,2044	-16,0954	611,1431	29,7702
	$\rho_1=0.8$ $\rho_2=0.8$	3276,587	0,0213	3276,587	0	1562,952	-0,0185	1562,952	0	705,7172	-0,029	705,7164	0,0001
$\bar{\rho}=0.9$	$\rho_1=0.8$ $\rho_2=1$	3701,73	-27,7363	2932,428	20,7822	2241,593	-27,9445	1460,699	34,8366	1409,011	-27,8427	633,7951	55,0184
	$\rho_1=0.85$ $\rho_2=0.95$	3051,074	-15,3716	2814,789	7,7443	1494,053	-13,378	1315,084	11,9788	790,0031	-13,8988	596,827	24,4526
	$\rho_1=0.9$ $\rho_2=0.9$	2813,502	-0,0067	2813,502	0	1409,266	-0,0414	1409,264	0,0001	620,2763	0,0417	620,2746	0,0003

Tablo 3. $Y_1/1000$~Gamma(2,3), $Y_2/1000$~Gamma(2,2) olmak üzere ortalama tahmin edicisinin HKO, sapma, varyans ve göreli sapma değerleri

		n=500				**n=1000**				**n=2000**			
		HKO	Sapma	Varyans	Göreli Sapma	HKO	Sapma	Varyans	Göreli Sapma	HKO	Sapma	Varyans	Göreli Sapma
$\bar{\rho}=1$	$\rho_1=\rho_2=1$	758,5167	-0,0137	758,5165	0	344,4174	0,0181	344,417	0,0001	155,4705	-0,0069	155,4705	0
$\bar{\rho}=0.7$	$\rho_1=0.5$ $\rho_2=0.9$	3353,083	48,8454	967,2151	71,1545	3039,742	51,008	437,9221	85,5934	2445,102	47,4122	197,1855	91,9355
	$\rho_1=0.6$ $\rho_2=0.8$	1285,876	22,4593	781,4545	39,2278	962,6229	23,985	387,3421	59,7618	784,7802	24,6505	177,1352	77,4287
	$\rho_1=0.7$ $\rho_2=0.7$	1099,089	0,0016	1099,089	0	505,8537	0,0004	505,8537	0	239,322	0,0178	239,3217	0,0001
$\bar{\rho}=0.8$	$\rho_1=0.65$ $\rho_2=0.95$	1659,866	28,9175	823,645	50,3788	1295,164	30,036	393,004	69,656	1154,585	31,2541	177,7683	84,6033
	$\rho_1=0.7$ $\rho_2=0.9$	1358,739	22,5462	850,4061	37,4121	814,7223	20,7914	382,4384	53,059	612,7225	21,0591	169,2368	72,3795
	$\rho_1=0.75$ $\rho_2=0.85$	889,34	10,7209	774,402	12,924	469,1222	10,26	363,8543	22,4393	276,8395	10,7396	161,5002	41,6629
	$\rho_1=0.8$ $\rho_2=0.8$	903,8843	-0,0088	903,8842	0	436,857	0,0523	436,8543	0,0006	200,755	-0,0051	200,7549	0
$\bar{\rho}=0.9$	$\rho_1=0.8$ $\rho_2=1$	1188,057	19,3784	812,5346	31,6081	698,9439	17,9497	376,7529	46,0968	502,7039	18,2587	169,3248	66,3172
	$\rho_1=0.85$ $\rho_2=0.95$	844,5641	9,1598	760,6627	9,9343	444,1861	9,298	357,734	19,463	248,1542	9,3003	161,659	34,8554
	$\rho_1=0.9$ $\rho_2=0.9$	832,0764	-0,0659	832,072	0,0005	390,0064	-0,0055	390,0063	0	176,7979	0,0003	176,7979	0

Tablo 4. $Y_1/1000 \sim Beta(9,3)$, $Y_2/1000 \sim Beta(9,2)$ olmak üzere ortalama tahmin edicisinin HKO, sapma, varyans ve göreli sapma değerleri

		n=500				**n=1000**				**n=2000**			
		HKO	Sapma	Varyans	Göreli Sapma	HKO	Sapma	Varyans	Göreli Sapma	HKO	Sapma	Varyans	Göreli Sapma
$\bar{\rho}=1$	$\rho_1=\rho_2=1$	27,3526	0,0036	27,3526	0	12,9287	-0,0034	12,9287	0,0001	5,8956	0,0014	5,8956	0
$\bar{\rho}=0.7$	$\rho_1=0.5$ $\rho_2=0.9$	124,2652	9,8151	27,9283	77,5253	108,5262	9,7634	13,2021	87,8351	103,6772	9,8844	5,9759	94,2361
	$\rho_1=0.6$ $\rho_2=0.8$	50,1293	4,7486	27,5803	44,9818	38,0151	5,0363	12,651	66,7211	30,4874	4,9714	5,7731	81,0641
	$\rho_1=0.7$ $\rho_2=0.7$	39,6315	0,0075	39,6314	0,0001	19,1784	-0,0033	19,1784	0,0001	9,0589	-0,0009	9,0589	0
$\bar{\rho}=0.8$	$\rho_1=0.65$ $\rho_2=0.95$	65,6518	6,143	27,9156	57,4793	55,1646	6,4846	13,1148	76,2261	51,4716	6,7501	5,9083	88,5212
	$\rho_1=0.7$ $\rho_2=0.9$	45,9202	4,2541	27,823	39,4101	28,5665	3,9474	12,9844	54,5467	25,3531	4,3973	6,0172	76,2663
	$\rho_1=0.75$ $\rho_2=0.85$	31,4727	2,0625	27,2186	13,5167	17,1279	2,0549	12,9053	24,6534	10,0716	2,0489	5,8735	41,6824
	$\rho_1=0.8$ $\rho_2=0.8$	35,2852	0,0014	35,2852	0	16,8192	0,0036	16,8192	0,0001	7,8018	0,0019	7,8018	0
$\bar{\rho}=0.9$	$\rho_1=0.8$ $\rho_2=1$	41,9813	3,8285	27,3236	34,9149	27,2816	3,8031	12,8179	53,0162	20,316	3,8137	5,7718	71,5901
	$\rho_1=0.85$ $\rho_2=0.95$	30,3344	1,9185	26,6538	12,1332	16,1612	1,8649	12,6833	21,5202	8,8985	1,783	5,7195	35,7252
	$\rho_1=0.9$ $\rho_2=0.9$	30,7712	0,0015	30,7712	0	14,482	-0,0001	14,482	0	6,6397	-0,0017	6,6397	0

Tablo 1-4 den görüldüğü gibi, tam cevaplı durumda ortalama tahmin edicisinin sapması sıfıra oldukça yakındır. Simülasyondaki tekrar sayısı arttırıldığında sapma değerinin sıfır olması beklenmektedir. Dolayısıyla tam cevaplı durumda ortalama tahmin edicisi yığın ortalamasının sapmasız bir tahmin edicisidir. Benzer şekilde, cevap verme eğilimlerinin iki grup için eşit olduğu durumlarda (ρ_1=ρ_2=0.7,0.8,0.9), $\bar{y}_c$ tahmin edicisi incelenen dağılımların hepsinde oldukça düşük sapma miktarına sahiptir. Göreli sapma değeri ise sıfıra oldukça yakındır. Simülasyondaki tekrar sayısı arttırıldığı takdirde tahmin edicinin sapma miktarının sıfır olması beklenmektedir. Böylece, cevap verme eğilimlerinin iki grup için eşit olduğu durumlarda $\bar{y}_c$ tahmin edicisinin $\bar{Y}$ için sapmasız bir tahmin edici olduğu söylenebilir. Bu sonuç Eşitlik (6)'ya ilişkin olarak yapılan ikinci çıkarsamayı desteklemektedir. Tahmin edicinin HKO değerlerine bakıldığında ise, tam cevaplı durum için HKO değerlerinin diğer durumlardan daha düşük olduğu görülmektedir. Özellikle düşük cevaplama oranlarına sahip yığınlar için HKO değerleri diğer cevaplama oranlarına göre hesaplanan HKO değerlerine göre daha yüksektir.

İncelenen tüm dağılımlar için, cevaplama oranının yüksek ve gruplar arası cevap verme eğilimi farklılığının büyük olduğu durumda, ortalama tahmin edicisinin cevaplamama sapması, cevaplama oranının ve cevap verme eğilimi farklılığının düşük olduğu duruma göre daha büyüktür. Örneğin, Tablo 3'te %80 cevaplı durumunda ($\bar{\rho} = 0.8$) n=500 iken, $\rho_1 = 0.75$ $\rho_2 = 0.85$ için cevaplamama sapması 10.7209 iken aynı tabloda %90 cevaplı durumunda ($\bar{\rho} = 0.9$) n=500 iken $\rho_1 = 0.8$ $\rho_2 = 1$ için cevaplamama sapması daha büyük olup 19.3784'dür. Bu sonuç, Eşitlik (6)'ya ilişkin olarak yapılan üçüncü çıkarsamayla çelişiyor gibi gözükebilir. Ancak bu durum, ilgilenilen değişken ile cevaplama eğilimi arasındaki ilişki $R_{\rho Y}$'nin sabit kalmamasından kaynaklanmaktadır. Uygulamada aynı cevaplama oranı için farklı $R_{\rho Y}$ değerleri ortaya çıkabilir. Bu durum simülasyon çalışmasına yansıtıldığından, $R_{\rho Y}$ değeri aynı kalmadığı sürece, düşük cevaplama oranlarının daha büyük sapmalara sebep olacağını söylemek uygun değildir. Tablo 1-4 den görüldüğü gibi, sapma değerleri aynı zamanda grupların cevap verme eğilimleri arasındaki farktan da etkilenmektedir. Aynı cevaplama oranı için cevap verme eğilimleri arasındaki fark arttıkça göreli sapma değeri de artmaktadır. Aynı cevap verme eğilimleri için ise, örnek çapı arttıkça göreli sapma değerleri artmaktadır. Bunun nedeni örnek çapının artmasının sapmayı değiştirmezken tahmine ilişkin varyansı azaltmasıdır.

Tablo 1'de yığın varyanslarının iki grup için eşit olduğu Normal dağılım altında, aynı cevaplama oranı, cevaplama eğilimi ve örnek çapı için göreli sapma değerleri, Tablo 2'de verilen varyansların farklı olduğu duruma göre elde edilen göreli sapma değerlerinden daha büyüktür. HKO değerlerine bakıldığında ise, varyansların farklı olduğu durumdaki HKO değerleri, varyansların eşit olduğu duruma göre daha büyüktür.

Ayrıca Tablo 3 ve 4'den görüldüğü gibi, sırasıyla sağa ve sola çarpık olan Beta ve Gamma dağılımına sahip yığınlar için ortalamanın tahmininde, cevaplamama sapmasının tahminin etkinliği üzerindeki göreli etkisi iki grup için varyansların farklı olduğu durumdaki Normal Dağılıma göre daha fazladır. Bunun yanı sıra Gamma dağılımı için elde edilen göreli sapma değerleri Beta dağılımı için elde edilen göreli sapma değerlerinden daha düşüktür.

5. Tartışma ve Sonuç

Çalışma sonucunda elde edilen bulguları özetlersek, cevaplama oranı azaldıkça yığın ortalamasına ilişkin tahmin edicinin sapmasının artacağını her durumda söylemek mümkün değildir. Tahmin edicinin sapması cevaplama oranının yanı sıra, ilgilenilen değişken ile cevap verme eğilimi arasındaki ilişkiye ve yığındaki birimlerin cevap verme eğilimleri arasındaki farklılıklara da bağlıdır. Örneğin yüksek bir cevaplama oranının gerçekleştiği bir araştırmada, yığını oluşturan iki grup arasındaki cevap verme eğilimleri farklılaştıkça tahmin edicinin sapması artacaktır. Bu sonuç, uygulamada cevap verme oranlarının yanı sıra; yığındaki birimlerin cevap verme eğilimlerinin de dikkate alınmasının gerekliliğine işaret etmektedir.

Bu çalışmada, araştırmalarda karşılaşılan cevaplamama durumunun ilgilenilen değişkenin tahminini olumsuz yönde etkilediği gösterilmiştir. Cevaplamama durumunun ilgilenilen değişkenin tahmini üzerindeki olumsuz etkisini azaltmak için cevaplama sapmasının daha düşük olduğu tahmin ediciler üzerinde çalışmalar yapmak faydalı olacaktır. Ayrıca bundan sonraki çalışmalarda, yardımcı değişken bilgisinin çok değişkenli olması durumunda ortaya çıkacak etkiler incelenebilir.

Kaynaklar

Bethlehem, J.G., Kersten H.M.P., 1985. On the Treatment of Non-response in Sample Surveys. Journal of Official Statistics, 1, 141-154.

Bethlehem J.G., 1999. Cross-sectional Research. In: H.J. Ader and G.J. Mellenberg, Research Methodology in the Social, Behaviuoural and Life Science. Sage publications, London, 110-142.

Bethlehem, J.G., 2002. Weighting Nonresponse Adjustments Based on Auxiliary Information. Survey Nonresponse, eds. R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little. New York, NY: Wiley.

Bethlehem J.G., Cobben F., Schouten B., 2011. Handbook of Nonresponse in Household Surveys. John Wiley and Sons., England.

Cassel, C.M., Särndal, C.E., Wretman, J.H., 1983. Some Uses of Statistical Models in Connection with the Nonresponse Problem. In: Madow, W.G., and Olkin, I. (eds): Incomplete Data in Sample Surveys, Vol.3, Proceedings of the Symposium, Academic Press, New York.

Groves, R. M., 1989. Survey Errors and Survey Costs, Wiley Series in Probability and Mathematical Statistics. New York, NY: Wiley.

Groves, R. M., 2006. Nonresponse Rates and Nonresponse Bias in Household Surveys. Public Opinion Quarterly, 70 (5), 646–75.

Groves, R. M.,Emilia Peytcheva., 2008. The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-analysis. Public Opinion Quarterly, 72 (2), 167–89.

Lindström, H., Wretman, J., Formsan, G.,Cassel, C., 1979. Standart Methods for Non-response Treatment in Statistical Estimation. National Central Bureau of Statistics, Sweden.

Little, R.J.A., Rubin, D.B., 2002. Statistical Analysis with Missind Data, Second Edition, Wiley Interscience, New York.

Lynn P., Beerten R.,Laiho J., Martin J., 2002. Towards Standardization of Survey Outcome Categories and Response Rate Calculations. Research in Official Statistics, 1, 63-86.

Peytchev, A., 2013. Consequencies of Survey Nonresponse. The Annals of the American Acedemy of Political and Social Science, 645, 88 (DOI: 10.1177/0002716212461748).

Särndal C.E., Lundström, S., 2006. Estimation in Surveys with Nonresponse. John Wiley and sons., Canada.

Schouten, B., 2007. A Selection Strategy for Weighting Variables Under a Not-missing-at-random Assumption, Journal of Official Statistics, 23-1, 51-68.