

# Kısmi rastgeleleştirilmiş yanıt modellerinde yeni bir regresyon tahmin edicisi

Nilgün ÖZGÜL\*, Hülya ÇINGİ

Hacettepe Üniversitesi, İstatistik Bölümü, Beytepe, 06800, Ankara, Türkiye

Geliş Tarihi (Received Date): 29.01.2020

Kabul Tarihi (Accepted Date): 14.05.2020

## Öz

Bu çalışmada, Kısmi Rastgeleleştirilmiş Yanıt Modellerinde (KRYM), hassas değişkenin kitle ortalamasının tahmini için hassas olmayan iki yardımcı değişken kullanarak yeni bir regresyon tahmin edicisi geliştirilmiştir. Önerilen tahmin edici, simülasyon çalışması yapılarak RYM'de var olan diğer tahmin edicilerle karşılaştırılmış ve önerilen tahmin edicinin var olan diğer tahmin edicilere göre daha etkin olduğu bulunmuştur.

**Anahtar kelimeler:** Hassas değişken, yardımcı değişken, regresyon tahmin edicisi, rastgeleleştirilmiş yanıt modelleri, etkinlik.

## A new regression estimator in partial randomized response models

### Abstract

In this study, a new regression estimator has been developed with using two non-sensitive auxiliary variables for the population mean estimation of the sensitive study variable in Partial Randomized Response Models (PRRM's). The proposed estimator is compared with existing estimators based on auxiliary variable in RRM with a simulation study. The proposed estimator is found to be more efficient than other existing estimators.

**Keywords:** Sensitive variable, auxiliary variable, regression estimator, randomized response models, efficiency.

\* Nilgün ÖZGÜL, nozgul@hacettepe.edu.tr, <https://orcid.org/0000-0003-0331-9044>

Hülya ÇINGİ, hcingi@hacettepe.edu.tr, <https://orcid.org/0000-0002-7329-9037>

## 1.Giriş

Rastgeleleştirilmiş Yanıt Modelleri (RYM'ler) anketlerde hassas sorular sorulduğunda cevaplayıcıların gizliliğini koruyarak güvenilir veriler elde etmek ve yanıt vermeme oranlarını azaltmak için kullanılır. Anketlerde cevaplayıcılar genellikle doğru cevap vermezler veya bu tür hassas sorular sorulduğunda doğrudan cevap vermeyi reddederler. Araştırmacılar, son zamanlarda geçerli ve güvenilir bilgiler elde etmek için RYM'leri kullanmışlardır. RYM'ler bir kitledeki bazı hassas davranışların ortalama değerini tahmin etmek için kullanılır. Bu hassas davranışlar uyuşturucu tüketimi, alkol tüketimi, yıllık gelir, hanehalkı geliri, bir kadının kürtaj sayısı vb. olabilir. Ortalama tahmini için geliştirilen yanıt modellerinde, bir davranış ya da olayın ortalama değeri tahmin edilmek istenir. Ortalama tahmini için geliştirilen yanıt modelleri toplamsal ya da çarpımsal modeller olarak iki ayrı başlıkta incelenir. Ortalama tahmini için yanıt modellerinde bir kart destesi kullanılır. Bu kart destesindeki sayılar belirli dağılımlardan (Normal, Uniform, Kikare, Gamma, Weibull v.b.) türetilir. Cevaplayıcı rastgele olarak bu desteden bir kart seçer. Toplamsal model için, cevaplayıcı kartta yazan değer ile kendi gerçek cevabını toplayarak görüşmeciye yanıtını verir. Çarpımsal model için ise, kartta yazan değer ile kendi gerçek cevabını çarparak görüşmeciye yanıtını verir. Burada önemli olan nokta, kişinin cevabının şansa dayalı olduğunun bilinmesidir. Yani kişinin cevapları tamamen rastgele düzenin sonucuna bağlıdır ve sonuç olarak cevaplayıcı, cevabının gizli kalacağından emindir. RYM'ler, cevaplayıcıların katılım durumlarına göre de gruplandırılabilir. Eğer tüm cevaplayıcılardan yanıtlarını rastgele düzenden faydalanarak verilmesi isteniyorsa, model "tam rastgeleleştirilmiş model" olarak tanımlanır. Eğer, cevaplayıcılardan sadece bir kesimi rastgele düzenden faydalanarak yanıtlarını veriyorsa, model "kısmi rastgeleleştirilmiş model" olarak tanımlanır [1]. Warner [2] ilk kez hassas nicel değişkenlerin kitle ortalamasını tahmin etmek için toplamsal rasgeleleştirilmiş yanıt yöntemini geliştirmiştir. Warner [2] öncülüğünde, hassas nicel değişkenlerin ortalamasını tahmin etmek için literatürde çok sayıda RYM geliştirilmiştir. Greenberg [3] ve Thornton ve Gupta [4], Warner [1]'in yaklaşımını genişleterek hassas nicel değişkenlerin ortalamasının tahmini için yeni toplamsal yöntemler geliştirmiştir. Gjestvang ve Singh [5], Hussain ve Shabbir [6] ve Gupta vd. [7] Kısmi RYM'ler önermiştir. Örnekleme teorisinde, yardımcı bilgi kullanıldığında, özellikle ilgilenilen değişken ile yardımcı değişken arasındaki ilişki yüksek olduğunda, Hata kareler Ortalaması (HKO) denkleminde önemli bir azalma olduğu bilinmektedir [8]. Son yıllarda, RYM'lerde hassas değişkenin ortalama tahmini için yardımcı bilgiler kullanılmıştır. Sousa vd. [9], Diana ve Perri [10], Gupta vd. [11] hassas nicel değişkenin hesaplanmasında yardımcı değişken kullanılarak sırasıyla regresyon, oran, iç içe regresyon-oran tahmin edicilerini önermişlerdir.

Bu çalışmada yardımcı değişken bilgisinin kullanıldığı toplamsal RYM'ler ele alınmıştır ve hassas olmayan iki yardımcı değişken kullanarak toplamsal KRYM'de hassas değişkeninin kitle ortalaması için yeni bir regresyon tahmin edicisi önerilmiştir.

Makalenin geri kalan kısmı şu şekilde düzenlenmiştir. Bölüm 2'de, RYM'lerde hassas değişkenin ortalaması için literatürde var olan yardımcı değişkene dayalı çeşitli tahmin ediciler verilmiştir. 3. bölümde önerilen regresyon tahmin edici ve varyans eşitliği verilmiştir. 4. bölümde, önerilen tahmin edici ile RYM'lerde var olan diğer tahmin ediciler bir simülasyon çalışması ile karşılaştırılmıştır. Bölüm 5 ile makale sonuçlandırılmıştır.

## 2. RYM'lerde yardımcı değişken bilgisinin kullanıldığı çeşitli tahmin ediciler

Bu bölümde RYM litaretüründe var olan yardımcı değişken bilgisinin kullanıldığı tahmin ediciler verilmiştir.

Y, doğrudan gözlemlenemeyen hassas olan ilgilenilen değişken olsun. X, Y ile pozitif ilişkisi olan hassas olmayan bir yardımcı değişken olsun. Sonlu bir kitleden rastgele n büyüklüğünde bir örneklem seçilsin. i. birim için ( $i = 1, 2, \dots, N$ )  $y_i$  ve  $x_i$  sırasıyla ilgilenilen değişken Y ve yardımcı değişken X'in değerlerini göstermektedir.  $z_i$  ise cevaplayıcılardan rastgele düzen yardımıyla alınan rastgeleleştirilmiş yanıt değişkeninin değerlerini göstermektedir.

Sousa vd. [9], tam rastgeleleştirilmiş toplamsal modelde, hassas olmayan bir yardımcı değişken kullanarak hassas değişkenin ortalaması için bir oransal tahmin edici önermiştir.

Sousa vd. [9]'nin önerdiği oransal tahmin edici,

$$\hat{\mu}_{SR} = \bar{z} \left( \frac{\mu_x}{\bar{x}} \right) \quad (1)$$

şeklindedir. Burada  $\bar{z}$ , hassas değişken için oluşturulan rastgeleleştirilmiş yanıtların örneklem ortalamasıdır ( $Z=Y+W$ ). W değişkeni, araştırmacı tarafından belirlenen dağılımdan türetilir.  $E(\bar{x}) = \mu_x = \frac{1}{N} \sum_{i=1}^N x_i$  hassas olmayan yardımcı değişkenin kitle

ortalaması,  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  hassas olmayan yardımcı değişkenin örneklem ortalamasıdır.

Sousa vd. [9] tahmin edicisi yanlı bir tahmin olduğu için bu tahmin için hata kareler ortalaması eşitliği verilmiştir.

$\hat{\mu}_{SR}$  tahmin edicisinin hata kareler ortalaması,

$$HKO(\hat{\mu}_{SR}) \cong \lambda \mu_z^2 [C_z^2 + C_x^2 - 2\rho_{xz} C_x C_z] \quad (2)$$

şeklindedir. Burada,  $\lambda = \frac{1}{n} - \frac{1}{N}$ ,  $C_z = \frac{S_z}{\mu_z}$  ve  $C_x = \frac{S_x}{\mu_x}$  sırasıyla Z ve X değişkenlerinin

kitle değişim katsayılarıdır.  $E(\bar{z}) = \mu_z = \frac{1}{N} \sum_{i=1}^N z_i$  Z değişkeninin kitle ortalaması,  $\rho_{xz}$ ,

Z ve X değişkenleri arasındaki kitle korelasyon katsayısıdır.

Diana ve Perri [10] kısmi rastgeleleştirilmiş toplamsal modelde, hassas olmayan bir yardımcı değişken kullanarak hassas değişkenin ortalaması için bir regresyon tahmin edicisi önermiştir. Diana ve Perri [10] modelinde, hassas değişken Y için, P oranında cevaplayıcıdan doğrudan gerçek yanıtlar alınırken, geriye kalan cevaplayıcılardan(1-P) rastgeleleştirilmiş toplamsal yanıtlar alınır. Hassas olmayan yardımcı değişken için, tüm cevaplayıcılardan doğrudan gerçek yanıtlar alınır. Buradaki P oranını araştırmacının kendisi belirler.

Bu bilgilere göre, Diana ve Perri [10] regresyon tahmin edicisi,

$$\hat{\mu}_{DP} = \bar{z} + b(\mu_x - \bar{x}) \quad (3)$$

şeklindedir. Burada,  $\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i$  hassas değişken için oluşturulan rastgeleştirilmiş değişkenin örneklem ortalamasıdır  $\{Z=PY+(1-P)(Y+W)\}$ . W değişkeni, araştırmacı tarafından belirlenen dağılımdan türetilir.  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  hassas olmayan yardımcı değişken için örneklem ortalaması, b ise X ile Z değişkeni arasındaki örneklem regresyon katsayısıdır.

$\hat{\mu}_{DP}$  tahmin edicisinin varyansı,

$$Var(\hat{\mu}_{DP}) = \frac{1}{n} (S_z^2 - 2bS_{zx} + b^2 S_x^2) \quad (4)$$

şeklindedir. Burada,  $S_z^2 = \frac{\sum_{i=1}^N (z_i - \mu_z)^2}{N-1}$  ve  $S_x^2 = \frac{\sum_{i=1}^N (x_i - \mu_x)^2}{N-1}$  sırasıyla Z ve X

değişkenlerinin kitle varyansları,  $S_{zx} = \frac{\sum_{i=1}^N (z_i - \mu_z)(x_i - \mu_x)}{N-1}$  Z ve X değişkenleri arasındaki

kitle kovaryansı,  $b = b_0 = \frac{S_{zx}}{S_x^2}$  Z ve X değişkenleri arasındaki kitle regresyon katsayısını

göstermektedir.  $E(\bar{z}) = \mu_z = \frac{1}{N} \sum_{i=1}^N z_i$  ve  $E(\bar{x}) = \mu_x = \frac{1}{N} \sum_{i=1}^N x_i$  sırasıyla Z ve X değişkenlerinin

kitle ortalamalarıdır.  $\hat{\mu}_{DP}$  tahmin edicisinin minimum varyansı,

$$Var(\hat{\mu}_{DP})_{min} = \frac{S_z^2}{n} (1 - \rho_{zx}^2) \quad (5)$$

şeklindedir.  $\rho_{zx} = \frac{S_{zx}}{S_z S_x}$  Z ve X değişkenleri arasındaki kitle korelasyon katsayısıdır.

Gupta vd. [11] tam rastgeleştirilmiş toplamsal modelde, hassas olmayan bir yardımcı değişken kullanarak hassas değişkenin ortalaması için iki farklı iç içe regresyon-oransal tahmin edicisi önermiştir. Gupta vd. [11] tahmin edicileri yanlı tahminler olduğu için bu tahminler için hata kareler ortalaması eşitlikleri verilmiştir.

Gupta vd. [11]'nin önerdiği ilk tahmin edici;

$$\hat{\mu}_{RR} = \left[ \bar{z} + \hat{\beta}_{zx}(\mu_x - \bar{x}) \right] \left( \frac{\mu_x}{\bar{x}} \right) \quad (6)$$

şeklindedir. Burada,  $\hat{\beta}_{zx}$ , Z ve X değişkenleri arasındaki örneklem regresyon katsayısını göstermektedir.

$\hat{\mu}_{RR}$  edicisinin hata kareler ortalaması,

$$HKO(\hat{\mu}_{RR}) \cong \lambda \mu_z^2 [C_z^2 (1 - \rho_{zx}^2) + C_x^2] \quad (7)$$

şeklindedir.

Gupta vd. [11]'nin önerdiği ikinci tahmin edici;

$$\hat{\mu}_{GRR} = [b_1 \bar{z} + b_2 (\mu_x - \bar{x})] \left( \frac{\mu_x}{\bar{x}} \right) \quad (8)$$

şeklindedir. Burada,  $b_1$  ve  $b_2$  sabit katsayılarıdır.

$\hat{\mu}_{GRR}$  tahmin edicisinin hata kareler ortalaması,

$$HKO(\hat{\mu}_{GRR}) \cong (b_1 - 1)^2 \mu_z^2 + \lambda \left[ b_1^2 \mu_z^2 \{C_z^2 + 3C_x^2 - 4\rho_{zx} C_z C_x\} + b_2^2 \mu_x^2 C_x^2 - 2b_1 \mu_z^2 \{C_x^2 - \rho_{zx} C_z C_x\} - 2b_2 \mu_z \mu_x C_x^2 - 2b_1 b_2 \mu_z \mu_x \{\rho_{zx} C_z C_x - 2C_x^2\} \right]. \quad (9)$$

Eşitlik (9)'un  $b_1$  ve  $b_2$  'ye göre türevi alınırsa, hata kareler ortalamasını minimum yapan optimal değerler

$$b_{1(opt)} = \frac{1 - \lambda C_x^2}{1 - \lambda \{C_x^2 - C_z^2 (1 - \rho_{zx}^2)\}} \quad \text{ve} \quad b_{2(opt)} = \frac{\mu_z}{\mu_x} \left\{ 1 + b_{1(opt)} \left( \frac{\rho_{zx} C_z}{C_x} - 2 \right) \right\} \quad (10)$$

şeklinde elde edilir.

Optimal  $b_1$  ve  $b_2$  değerleri Eşitlik (9)'da yerine konulursa,  $\hat{\mu}_{GRR}$  tahmin edicisinin minimum HKO eşitliği,

$$HKO(\hat{\mu}_{GRR})_{\min} \cong \frac{\lambda \mu_z^2 C_z^2 (1 - \rho_{zx}^2) (1 - \lambda C_x^2)}{\lambda C_z^2 (1 - \rho_{zx}^2) + (1 - \lambda C_x^2)} \quad (11)$$

olarak elde edilir.

### 3. Toplamsal kısmi rastgeleştirilmiş yanıt modelinde önerilen regresyon tahmin edicisi

Bu çalışmada, toplamsal KRYM'de hassas nicel değişkenin kitle ortalaması tahmini için yeni bir regresyon tahmin edicisi geliştirilmiştir. Geliştirilen regresyon tahmin edicisinde, hassas nicel değişken tahmini için iki hassas olmayan yardımcı değişken kullanılmıştır (X ve M). Önerilen yöntemde, cevaplar şu şekilde alınmaktadır. Hassas değişken Y için, P oranında cevaplayıcıdan doğrudan gerçek yanıtlar alınırken, geriye kalan cevaplayıcılardan (1-P) rastgeleştirilmiş toplamsal yanıtlar alınır. Hassas olmayan yardımcı değişkenler X ve M için, tüm cevaplayıcılardan doğrudan gerçek yanıtlar alınır.

N birimlik kitleden yerine konmadan basit rastgele örnekleme ile n büyüklüğünde örneklem seçilir. Örneklemeden alınan cevaplar  $\{(z_1, x_1, m_1), (z_2, x_2, m_2), \dots, (z_n, x_n, m_n)\}$

şeklinde olacaktır.

Buna göre, bu çalışmada önerilen regresyon tahmin edicisi,

$$\hat{\mu}_{PR} = \bar{z} + b_1(\mu_x - \bar{x}) + b_2(\mu_m - \bar{m}) \quad (12)$$

şeklinde verilir. Burada,  $\bar{z}$  rastgeleleştirilmiş yanıt değişkeninin örneklem ortalamasıdır  $\{Z=PY+(1-P)(Y+W)\}$ .  $W$  değişkeni, araştırmacı tarafından belirlenen dağılımdan türetilir.  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  ve  $\bar{m} = \frac{1}{n} \sum_{i=1}^n m_i$  sırasıyla birinci ve ikinci yardımcı değişkenin

örneklem ortalamalarını gösterir.  $E(\bar{m}) = \mu_m = \frac{1}{N} \sum_{i=1}^N m_i$  ikinci yardımcı değişkenin kitle ortalamasıdır.  $b_1$  ve  $b_2$  sabit katsayılarıdır.

Önerilen tahmin edicinin varyansını bulmak için öncelikle aşağıdaki notasyonlar tanımlanır.

$$e_0 = \frac{(\bar{z} - \mu_z)}{\mu_z}, \quad e_1 = \frac{(\bar{x} - \mu_x)}{\mu_x}, \quad e_2 = \frac{(\bar{m} - \mu_m)}{\mu_m}.$$

Burada,

$$E(e_0) = E(e_1) = E(e_2) = 0; E(e_0^2) = \lambda C_z^2, E(e_1^2) = \lambda C_x^2, E(e_2^2) = \lambda C_m^2,$$

$$E(e_0 e_1) = \lambda \rho_{zx} C_z C_x, E(e_0 e_2) = \lambda \rho_{zm} C_z C_m, E(e_1 e_2) = \lambda \rho_{xm} C_x C_m,$$

$$C_m^2 = \frac{S_m^2}{\mu_m^2}, S_m^2 = \frac{\sum_{i=1}^N (m_i - \mu_m)^2}{N-1}, \rho_{zx} = \frac{S_{zx}}{S_z S_x}, \rho_{zm} = \frac{S_{zm}}{S_z S_m}, \rho_{xm} = \frac{S_{xm}}{S_x S_m}.$$

Bu notasyonları kullanarak

$$[\hat{\mu}_{PR} - \mu_z] = [\bar{z} + b_1(\mu_u - \bar{u}) + b_2(\mu_v - \bar{v}) - \mu_z] \quad (13)$$

yazılır. Eşitlik (13)'te her iki tarafın beklenen değeri alınır, önerilen tahmin edicinin yan eşitliği elde edilir. Buna göre  $\hat{\mu}_{PR}$  tahmin edicisinin yan eşitliği

$$Yan(\hat{\mu}_{PR}) = E(\hat{\mu}_{PR} - \mu_z) = E(\bar{z}) + b_1 \{ \mu_u - E(\bar{u}) \} + b_2 \{ \mu_v - E(\bar{v}) \} - \mu_z = 0 \quad (14)$$

olarak elde edilir. Eşitlik (14)'ten, önerilen tahmin edicinin yansız olduğu söylenebilmektedir.

Eşitlik (13)'te her iki tarafın karesi ve beklenen değeri alındığında önerilen tahmin edicinin varyans eşitliği elde edilir. Buna göre  $\hat{\mu}_{PR}$  tahmin edicisinin varyans eşitliği,

$$\begin{aligned} \text{Var}(\hat{\mu}_{PR}) &= E(\mu_z^2 e_0^2 + b_1^2 \mu_x^2 e_1^2 + b_2^2 \mu_m^2 e_2^2 - 2b_1 \mu_z \mu_x e_0 e_1 - 2b_2 \mu_z \mu_m e_0 e_2 + 2b_1 b_2 \mu_x \mu_m e_1 e_2) \\ &= \lambda(\mu_z^2 C_z^2 + b_1^2 S_x^2 + b_2^2 S_m^2 - 2b_1 S_{zx} - 2b_2 S_{zm} + 2b_1 b_2 S_{xm}), \end{aligned} \quad (15)$$

şeklinde elde edilir. Burada,  $S_{xm} = \frac{\sum_{i=1}^N (x_i - \mu_x)(m_i - \mu_m)}{N-1}$  X ve M değişkenleri arasındaki kitle kovaryans eşitliğidir.

Eşitlik (15)'in  $b_1$  ve  $b_2$  sabitlerine göre türevi alınıp sıfıra eşitlenerek  $b_1$  ve  $b_2$  sabitlerinin optimal değerleri elde edilir. Buna göre, şu eşitlikler elde edilir:

$$\begin{aligned} \frac{\partial \text{Var}(\hat{\mu}_{PR})}{\partial b_1} &= \frac{\partial \lambda(\mu_z^2 C_z^2 + b_1^2 S_x^2 + b_2^2 S_m^2 - 2b_1 S_{zx} - 2b_2 S_{zm} + 2b_1 b_2 S_{xm})}{\partial b_1} = 0 \\ &= 2b_1 S_x^2 - 2S_{zx} + 2b_2 S_{xm} = 0 \end{aligned} \quad (16)$$

$$\begin{aligned} \frac{\partial \text{Var}(\hat{\mu}_{PR})}{\partial b_2} &= \frac{\partial \lambda(\mu_z^2 C_z^2 + b_1^2 S_x^2 + b_2^2 S_m^2 - 2b_1 S_{zx} - 2b_2 S_{zm} + 2b_1 b_2 S_{xm})}{\partial b_2} = 0 \\ &= 2b_2 S_m^2 - 2S_{zm} + 2b_1 S_{xm} = 0 \end{aligned} \quad (17)$$

Eşitlik (16) ve (17) eşanlı olarak çözümlerse,  $\hat{\mu}_{PR}$  tahmin edicisinin varyansını minimum yapan optimal  $b_1$  ve  $b_2$  değerleri

$$b_{1(opt)} = \frac{(\rho_{zx} - \rho_{zm}\rho_{xm})C_z/C_x}{1 - \rho_{xm}^2} \quad \text{ve} \quad b_{2(opt)} = \frac{\rho_{zm} - \rho_{zx}\rho_{xm}(C_z/C_m)}{1 - \rho_{xm}^2} \quad (18)$$

şeklinde elde edilir.

$b_1$  ve  $b_2$  optimal değerleri Eşitlik(15)'te yerine konulursa,  $\hat{\mu}_{PR}$  tahmin edicisinin minimum varyans eşitliği,

$$\text{Var}_{\min}(\hat{\mu}_{PR}) = \lambda S_z^2 (1 - R_{zx,m}^2), \quad (19)$$

olarak elde edilir. Burada,  $R_{zx,m}^2 = \frac{\rho_{zx}^2 + \rho_{zm}^2 - 2\rho_{zx}\rho_{zm}\rho_{xm}}{1 - \rho_{xm}^2}$ .

#### 4.Simülasyon çalışması

Simülasyon çalışmasında hassas değişkenin ortalama tahmini için bu çalışmada önerilen tahmin edici ile literatürde var olan tahmin ediciler karşılaştırılmıştır. Çalışmada önerilen tahmin edici,  $\hat{\mu}_{PR}$ ; Diana ve Perri [10] tahmin edicisi,  $\hat{\mu}_{DP}$ , Sousa vd. [9] tahmin edicisi,  $\hat{\mu}_{SR}$ , Gupta vd. [11] tahmin edicileri  $\hat{\mu}_{RR}$  ve  $\hat{\mu}_{GRR}$  ile karşılaştırılmıştır. Çalışma için R programında N=1000 büyüklüğünde çok değişkenli normal dağılımdan faydalanarak 3 farklı kitle türetilmiştir. Y, X ve M değişkenleri R programında çoklu normal dağılımından varyans-kovaryans matrisi oluşturularak türetilmiştir. Y, X ve M değişkenlerinin ortalamaları (5, 5, 5) olarak ve varyansları (9, 4, 4) olarak belirlenmiştir. Yardımcı değişken bilgisinin kullanıldığı tahmin edicilerde değişkenler arasındaki ilişki önemli olduğu için kitleler değişkenler arasındaki ilişkilerin derecesine göre türetilmiştir. Birinci kitle için ilişki katsayıları  $\rho_{yx} = 0.58$ ,  $\rho_{ym} = 0.52$ ,  $\rho_{xm} = 0.30$  (orta derece ilişki),

ikinci kitle için ilişki katsayıları  $\rho_{yx} = 0.87$ ,  $\rho_{ym} = 0.70$ ,  $\rho_{xm} = 0.50$  'dir ( yüksek derece ilişki). Z rastgeleştirilmiş yanıt değişkeninde kullanılacak olan W değişkeni  $N(0;0.30^2)$  dağılımından türetilmiştir. Simülasyon çalışmasında dağılımlar türetilirken Diana ve Perri [10] ve Singh v.d. [11]'nin simülasyon çalışmasında kullandıkları dağılımlardan faydalanılmıştır. Çalışmada 1000 tekrar yapılarak  $n=50, 100, 200, 300$  örneklem büyüklükleri için hesaplamalar yapılmıştır. Çalışmada  $P=0.20$  alınmıştır. Yani örnekleme cevaplayıcıların %20'sinden doğrudan cevaplar alınırken, %80'inden rastgeleştirilmiş cevaplar alınmıştır. Çalışmada kullanılan deneysel hata kareler ortalaması formülü:

$$HKO(\hat{\mu}) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\mu} - \mu_y)^2 \quad (20)$$

şeklindedir. Simülasyon sonuçları Tablo 1'de verilmiştir. Tahmin edicilerin karşılaştırılmasında tahmin edicilerin teorik ve deneysel hata kareler ortalaması kullanılmıştır.

Tablo 1'de, hassas değişken ve hassas olmayan yardımcı değişkenler arasındaki ilişki derecesine ve örneklem büyüklüklerine göre tahmin edicilerin teorik ve deneysel HKO değerleri verilmektedir. Tablo1'e göre, hem ilişki derecesi hem de örneklem büyüklüğünden bağımsız olarak, önerilen regresyon tahmin edicisi, Sousa vd.[8], Diana ve Perri [9] ve Gupta vd. [10] tahmin edicilerinden her zaman daha etkilidir. Önerilen tahmin edicinin hatasının her iki kitle için de diğer tahmin edicilerin hatalarından daha küçük olduğu görülmektedir. Hassas değişken ile yardımcı değişkenler arasındaki ilişki yüksek düzeyde olduğunda literatürde önerilen tahmin edicilerin hata kareler ortalamaları ile bu çalışmada önerilen tahmin edicinin hata kareler ortalamaları arasındaki farklılık oldukça artmaktadır. Bu çalışmada önerilen tahmin edicinin yüksek ilişkili değişkenler kullanıldığında literatürdeki diğer tahmin edicilere göre oldukça duyarlı olduğu görülmektedir. Ayrıca, değişkenler arasındaki ilişkinin artmasıyla tahmin edicilerin hata kareler ortalamasında azalma olduğu gözlenmiştir. Aynı şekilde örneklem büyüklüğü arttıkça tahmin edicilerin HKO'ları da düşmektedir ve bu beklenen bir sonuçtur.

## 5. Sonuç

Bu çalışmada, kısmi rastgeleştirilmiş yanıt modellerinde hassas değişkenin kitle ortalaması tahmini için hassas olmayan iki yardımcı değişkene dayanan yeni bir regresyon tahmin edicisi geliştirilmiştir. Önerilen tahmin edici her durumda mevcut diğer tahmin edicilerden daha etkilidir. Hassas değişken ile yüksek ilişkili hassas olmayan yardımcı değişkenler elde edilebildiğinde, önerilen tahmin edici güvenilir olarak değerlendirilebilir ve hassas konularda çalışan araştırmacılar için uygulanabilir bir yöntem olabilir. Türkiye'de hassas konularla ilgili çalışmalar yapılmakta olup, rastgeleştirilmiş yanıt modelleri uygulama alanı bulamamıştır. Bu yüzden, ileriki çalışmalarda rastgeleştirilmiş yanıt modellerinin bir uygulamasının yapılması planlanmaktadır.



Tablo 1. İlişki derecesi ve örneklem büyüklüklerine göre, tahmin edicilerin teorik ve deneysel HKO değerleri.

n	Tahmin Ediciler	Kitle 1		Kitle 2	
		HKO		HKO	
		Teorik	Deneysel	Teorik	Deneysel
50	$\hat{\mu}_{DP}$	0.1136	0.1099	0.0333	0.0334
	$\hat{\mu}_{SR}$	0.1152	0.1117	0.0434	0.0435
	$\hat{\mu}_{RR}$	0.1926	0.1963	0.1085	0.1091
	$\hat{\mu}_{GRR}$	0.1130	0.1106	0.0333	0.0334
	$\hat{\mu}_{PR}$	<b>0.0898</b>	<b>0.0868</b>	<b>0.0051</b>	<b>0.0059</b>
100	$\hat{\mu}_{DP}$	0.0537	0.0477	0.0158	0.0161
	$\hat{\mu}_{SR}$	0.0545	0.0489	0.0205	0.0210
	$\hat{\mu}_{RR}$	0.0912	0.0874	0.0514	0.0535
	$\hat{\mu}_{GRR}$	0.0536	0.0475	0.0157	0.0160
	$\hat{\mu}_{PR}$	<b>0.0425</b>	<b>0.0378</b>	<b>0.0024</b>	<b>0.0026</b>
200	$\hat{\mu}_{DP}$	0.0239	0.0232	0.0070	0.0068
	$\hat{\mu}_{SR}$	0.0242	0.0236	0.0091	0.0085
	$\hat{\mu}_{RR}$	0.0405	0.0419	0.0228	0.0243
	$\hat{\mu}_{GRR}$	0.0238	0.0232	0.0070	0.0068
	$\hat{\mu}_{PR}$	<b>0.0189</b>	<b>0.0186</b>	<b>0.0010</b>	<b>0.0010</b>
300	$\hat{\mu}_{DP}$	0.0139	0.0136	0.0040	0.0042
	$\hat{\mu}_{SR}$	0.0141	0.0137	0.0053	0.0054
	$\hat{\mu}_{RR}$	0.0236	0.0235	0.0133	0.0144
	$\hat{\mu}_{GRR}$	0.0139	0.0136	0.0040	0.0042
	$\hat{\mu}_{PR}$	<b>0.0110</b>	<b>0.0111</b>	<b>0.0006</b>	<b>0.0006</b>

**Kaynaklar**

- [1] Özgül, N., Oran ve Ortalama Tahmin Edicileri, Doktora Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, (2013).
- [2] Warner, S.L., The Linear Randomized Response Model, **Journal of the American Statistical Association**, 66, 336, 884-888, (1971).
- [3] Greenberg B.G., Kuebler R.R., Abernathy J.R., Horvitz D.G., Application of the randomized response technique in obtaining quantitative data, **Journal of American Statistical Association**, 66, 243–250, (1971).
- [4] Thornton, B. ve Gupta, S. N., Comparative Validation of a Partial (versus Full) Randomized Response Technique: Attempting to Control for Social Desirability Response Bias to Sensitive Questions, **Individual Differences Research**, 2, 214-224, (2004).
- [5] Gjesvang, C.R., Singh,S., An improved Randomized Response Model:Estimation of Mean, **Journal of Applied Statistics**, 36, 12, 1361-1367, (2009).
- [6] Hussain Z., Shabbir, J., Improved Estimation Procedures for the Mean of Sensitive Variable using Randomized Response Model, **Pakistan Journal of Statistics**, 25, 2, 205-220, (2009).
- [7] Gupta S., Shabbir J., Sousa R., Sehra S., Mean and sensitivity estimation in optional randomized response model, **Journal of Statistical Planning and Inference**, 140, 10, 2870-2874, (2010).
- [8] Cingi, H., **Örnekleme Kuramı**, Ankara: 3.Baskı, Bizim Büro Basımevi, (2009).
- [9] Sousa, R., Shabbir, J. Real, P. C., Gupta, S., Ratio estimation of the mean of a sensitive variable in the presence of auxiliary information, **Journal of Statistical Theory and Practice**, 4, 3, 495-507, (2010).
- [10] Diana G., ve Perri P.F., A class of Estimators for Quantitative Sensitive Data, **Statistical Papers**, 52, 633-650, (2011).
- [11] Gupta, S., Shabbir, J., Sousa, R., Real, P. C., 2012, Estimation of the mean of a sensitive variable in the presence of auxiliary information, **Communications in Statistics, Theory and Methods**, 41, 1–12, (2012).