T.C.
GAZİ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI

ÖRNEKLEME YÖNTEMLERİNDE YARDIMCI DEĞİŞKENLERİN KULLANIMI VE REGRESYON TAHMİN EDİCİSİ YÖNTEMİ İLE BİR UYGULAMA

YÜKSEK LİSANS TEZİ

Hazırlayan
Cenker Burak METİN

Tez Danışmanı
Prof. Dr. Özkan ÜNVER

ANKARA-2010

ONAY

Cenker Burak METİN tarafından hazırlanan "Örnekleme Yöntemlerinde Yardımcı Değişkenlerin Kullanımı ve Regresyon Tahmin Edicisi ile Bir Uygulama" başlıklı bu çalışma, 31.03.2010 tarihinde yapılan savunma sınavı sonucunda oybirliği ile başarılı bulunarak jürimiz tarafından Ekonometri Anabilim dalında Yüksek Lisans tezi olarak kabul edilmiştir.

Prof. Dr. Özkan ÜNVER (Başkan)

Prof. Dr. Bedriye SARAÇOĞLU

Prof. Dr. Hamza GAMGAM

ÖNSÖZ

Bu çalışmada, örnekleme araştırmaları sonucu verilen parametre tahminlerinin, daha iyi ve daha etkin sonuçlar verebilmesi için yardımcı değişken kullanımı incelenmiştir. Bu amaçla; 2006 Yıllık Sanayi ve Hizmet İstatistikleri Araştırmasında tamsayımla elde edilen İmalat Sanayi'nde 20+ çalışanı olan girişimler için ortalama ciro parametresi, 2006 araştırmasından belli sayıda birim örnek çekilerek, bu birimlere ilişkin 2005 Yıllık Sanayi ve Hizmet İstatistikleri Araştırmasından temin edilen yardımcı değişkenler kullanılarak tahmin edilmiştir. Yardımcı değişken kullanılarak yapılan tahminler; gerçek parametre değeriyle, klasik tahminle ve birbirleriyle karşılaştırılmıştır.

Bu konuyu çalışmamda yardımcı olan danışmanım Sayın Prof.Dr. Özkan ÜNVER'e, görüş ve önerilerini esirgemeyen TÜİK İşyerleri ve Tarım Araştırmaları Takım Sorumlusu Sayın Nilay EROL'a, uygulama aşamasında veriler konusunda yardımcı olan TÜİK Yıllık Sanayi ve Hizmet İstatistikleri Takımına, yardım ve desteklerinden ötürü TÜİK Örnekleme ve Kalite Teknikleri Grubundaki mesai arkadaşlarıma, huzurlu bir çalışma ortamı sağlayan sevgili aileme en içten teşekkür ve şükranlarımı sunarım.

Sabrını ve desteğini esirgemeyerek, tezin düzenlenmesinde bana yardımcı olan sevgili nişanlım Gözde KARANFİLCİ'ye ayrıca teşekkür ederim.

İÇİNDEKİLER

SİMGELER VE KISALTMALAR

SİMGELER

N :	Yığın büyüklüğü
n :	Örnek büyüklüğü
$\bar{Y}$:	İlgilenilen değişken için yığın ortalaması
$\bar{y}$:	İlgilenilen değişken için örnek ortalaması
S_y^2 :	İlgilenilen değişken için yığın varyansı
s_y^2 :	İlgilenilen değişken için örnek varyansı
π_i :	i. birim için kapsama olasılığı, örnekte bulunma olasılığı
$\hat{\tau}_\pi$:	İlgilenilen değişken için Horvitz-Thompson toplam tahmini
$\hat{\bar{Y}}$:	Horvitz-Thompson tahmin edicisi
f :	Örnekleme oranı
$V(.)$:	İlgili tahmin edici için varyans
$\bar{X}$:	Yardımcı değişken için yığın ortalaması
$\bar{x}$:	Yardımcı değişken için örnek ortalaması
S_x^2 :	Yardımcı değişken için yığın varyansı
s_x^2 :	Yardımcı değişken için örnek varyansı
ρ_{yx} :	Yığın korelasyon katsayısı
C_y :	İlgilenilen değişken için değişim katsayısı
C_x :	Yardımcı değişken için değişim katsayısı
ν :	Düzeltme terimi
β :	Yığın regresyon katsayısı
b :	Basit rasgele örneklemede örnek regresyon katsayısı
κ :	Kaur regresyon tahmin edicisi için belirlenen sabit
N_h:	h. tabaka için yığın büyüklüğü
n_h:	h. tabaka için örnek büyüklüğü

W_h :	h. tabakanın ağırlığı
$\overline{Y}_h$:	h. tabakanın ilgilenilen değişken için yığın ortalaması
$\overline{y}_h$:	h. tabakanın ilgilenilen değişken için örnek ortalaması
$S^2_{y_h}$:	h. tabakanın ilgilenilen değişken için yığın varyansı
C^2_{yh} :	h. tabakanın ilgilenilen değişken için değişim katsayısı
$\overline{X}_h$:	h. tabakanın yardımcı değişken için yığın ortalaması
$\overline{x}_h$:	h. tabakanın yardımcı değişken için örnek ortalaması
$S^2_{x_h}$:	h. tabakanın yardımcı değişken için yığın varyansı
C^2_{xh} :	h. tabakanın yardımcı değişken için değişim katsayısı
$G_{2h}(x)$:	h. tabakanın yardımcı değişken için basıklık katsayısı
β_c :	Bileşik regresyon tahmininde yığın regresyon katsayısı
b_c :	Tabakalı rasgele örneklemede örnek regresyon katsayısı
β_h :	h. tabaka için yığın regresyon katsayısı
b_h :	h. tabaka için örnek regresyon katsayısı
R^2 :	Belirlilik katsayısı
d_i :	i. birimin tasarım ağırlığı
b_d :	GREG için regresyon katsayılar vektörü
w_i :	i. birimin ayarlama ağırlığı
λ_j :	j. değişken için Lagrange çarpanı
$\hat{\overline{Y}}_{ORAN}$:	$\overline{Y}$ için klasik oran tahmin edicisi
$\hat{\overline{Y}}_{BEALE}$:	$\overline{Y}$ için Beale oran tahmin edicisi
$\hat{\overline{Y}}_{TIN}$:	$\overline{Y}$ için Tin oran tahmin edicisi
$\hat{\overline{Y}}_{SINTAIL}$:	$\overline{Y}$ için Singh-Tailor oran tahmin edicisi
$\hat{\overline{Y}}_{SRIV}$:	$\overline{Y}$ için Srivastava oran tahmin edicisi

$\hat{\bar{Y}}_{REG}$: $\bar{Y}$ için klasik doğrusal regresyon tahmin edicisi

$\hat{\bar{Y}}_{KAUR}$: $\bar{Y}$ için Kaur regresyon tahmin edicisi

$\hat{\bar{Y}}_{GREG}$: $\bar{Y}$ için Genelleştirilmiş regresyon tahmin edicisi

$\hat{\bar{Y}}_{TABAKA}$: $\bar{Y}$ için klasik tabakalı tahmin edicisi

$\hat{\bar{Y}}_{ORANAYRI}$: $\bar{Y}$ için ayrı oran tahmin edicisi

$\hat{\bar{Y}}_{ORANBİL}$: $\bar{Y}$ için bileşik oran tahmin edicisi

$\hat{\bar{Y}}_{REGAYRI}$: $\bar{Y}$ için ayrı regresyon tahmin edicisi

$\hat{\bar{Y}}_{REGBİL}$: $\bar{Y}$ için bileşik regresyon tahmin edicisi

$\hat{\bar{Y}}_{SDTAB}$: $\bar{Y}$ için Sisodia-Dwivedi tabakalı oran tahmin edicisi

$\hat{\bar{Y}}_{SKTAB}$: $\bar{Y}$ için Singh-Kakran tabakalı oran tahmin edicisi

$\hat{\bar{Y}}_{US1TAB}$: $\bar{Y}$ için Upadhyaya-Singh-1 tabakalı oran tahmin edicisi

$\hat{\bar{Y}}_{US2TAB}$: $\bar{Y}$ için Upadhyaya-Singh-2 tabakalı oran tahmin edicisi

KISALTMALAR

EFİS: Ekonomik Faaliyetlerin İstatistiki Sınıflaması
HKO: Hata kareler ortalaması
HTE: Horvitz-Thompson tahmin edicisi
GREG: Genelleştirilmiş Regresyon
ISI: Uluslararası İstatistik Kurumu
MC: Monte Carlo
İBBS2: İstatistiki Bölge Birimleri Sınıflaması (Düzey2)
TÜİK: Türkiye İstatistik Kurumu

TABLOLAR

GİRİŞ

Araştırmacıların kullanımına sunulan istatistik verileri çoğunlukla örnekleme yöntemlerinden elde edilmiş yığına ilişkin parametre tahmin değerleridir. Yığından çekilen örnek birimlerinden elde edilen verilerin yığına genişletilmesi ile parametre tahminlerine ulaşılmaktadır. Parametre değerlerinin yerine tahminlerinin kullanılabilmesi için bu tahminlerin tutarlı ve etkin olması beklenir. Diğer yandan örnek birimleri, yığındaki elemanların yalnızca bir bölümünü kapsadığından tahminler hatalara açıktır. Tahminin en iyi ve etkin düzeyde olması için sözü edilen hatalar minimize edilmeye çalışılır. Yardımcı değişkenlerin tahmin aşamasında kullanılarak tahminlerin etkinliğinin arttırılması; bu tezin konusunu oluşturmaktadır.

Klasik tahmin teorisinde ilgilenilen değişken için yığın parametresi tahmin edilirken, ilgili değişken için örnekleme araştırmasından gelen değerler kullanılır. Örneğin Y değişkeni için yığın ortalaması $\bar{Y}$'yı örnekleme dayalı olarak tahmin eden klasik ortalama tahmin edicisi $\bar{y}$ sadece örnekleme araştırmasındaki y_i değerlerini kullanmaktadır. Yapılan çalışmalar $\bar{Y}$ tahmin edilirken X gibi bir yardımcı değişkenin tahmin aşamasında kullanılabileceğini göstermiştir. X yardımcı değişkeninin kullanımı, tahminin güvenilir dışsal bir kaynakla kontrol edilerek yapılmasını sağlar, tahminde yapılan hata miktarını azaltır, daha iyi ve etkin tahminler üretir.

Tahminlerin yardımcı değişken kullanılarak yapılması ile ilgili çalışmalar Cochran (1964), Särndal, Svensson ve Wretman (1992) tarafından hazırlanan kitaplarda detaylı biçimde ele alınmıştır. Yardımcı değişken bilgisini kullanarak elde edilen ortalama tahmin edicilerinin en önemlileri, klasik oran tahmin edicisi ve klasik doğrusal regresyon tahmin edicisidir. Klasik oran tahmin edicisinin yanlı tahmin vermesi nedeni ile birçok istatistikçi yanlılığı azaltmak üzere alternatif oran tahmin edicileri önermişlerdir. Quenoille (1956), Beale (1962), Tin (1965), Srivastava (1965) basit rasgele örnekleme yönteminde klasik oran tahmin edicisine alternatif oran tahmin edicileri öneren ilk araştırmacılardır. Daha sonraki yıllarda Chakrabarty (1979), Ray ve Singh (1981), Sisodia ve Dwivedi (1981), Singh ve Kakran (1993), Upadhyaya ve Singh (1999), Singh ve Tailor (2003) basit rasgele örnekleme yönteminde alternatif oran tahmin edicileri önermişlerdir. Tabakalı rasgele örneklemede oran tahmin edicisi kullanımı ilk olarak Hansen, Hurwitz ve Gurney (1946) tarafından önerilmiştir. Kadılar ve Çıngı (2003), Sisodia ve Dwivedi (1981), Singh ve Kakran (1993), Upadhyaya ve Singh (1999) tarafından basit rasgele örnekleme için önerilen oran tahmin edicilerini tabakalı rasgele örneklemeye uyarlamışlardır.

Örnekleme araştırmalarında regresyonun kullanımı ile ilgili ilk referanslar Cochran (1942) ve Jessen (1942)'dir. Cochran (1942) örnekleme araştırmasında regresyon için doğrusal model teorisine dayanan temel teoriyi vermiştir. 1970 ve 1980'li yıllarda regresyon tahmin edicisinin yapısı üzerinde birçok araştırma yapılmıştır. Fuller (1973,1975) örnekleme araştırmasından elde edilen regresyon katsayıları vektörünün büyük örnek özelliklerini vermiştir. Royall ve Cumberland (1981) regresyon tahmin edicilerinin koşullu özelliklerini çalışmışlardır. Cassel, Särndal ve Wretman (1976) tahmin edici oluşturulmasında hem model hem de tasarım prensiplerini düşünerek, tasarım tutarlı tahmin ediciler için "genelleştirilmiş regresyon tahmin edicisi" terimini önermişlerdir.

Bu çalışmada yardımcı değişken bilgisi kullanılarak elde edilen temel oran tahmin edicileri ve regresyon tahmin edicileri ele alınacaktır. Birinci bölümde; tahmin edicilerin daha iyi anlaşılabilmesi için örnekleme teorisi hakkında genel bir bilgi verilerek, örneklemede kullanılan temel tanım ve kavramlar kısaca açıklanacaktır. İkinci bölümde, basit rasgele örnekleme yönteminde kullanılan bazı oran tahmin edicileri ve regresyon tahmin edicileri incelenecektir. Üçüncü bölümde, tabakalı rasgele örnekleme yönteminde kullanılan bazı oran tahmin edicileri ve regresyon tahmin edicileri incelenecektir. Dördüncü bölümde, genelleştirilmiş regresyon tahmin edicisinin teorisi anlatılacaktır.

Uygulama aşamasında Türkiye İstatistik Kurumu (TÜİK) tarafından gerçekleştirilen Yıllık Sanayi ve Hizmet İstatistikleri Araştırmasının 2005 ve 2006 yılı verileri kullanılacaktır. 2006 yılında İmalat Sanayi'nde 20+ çalışanı olan girişimlere ilişkin ortalama ciro parametresi; diğer bölümlerde anlatılan oran tahmin edicileri ve regresyon tahmin edicileri ile tahmin edilecektir.

BİRİNCİ BÖLÜM

ÖRNEKLEME NEDİR?

Toplumların sosyal, kültürel davranışlarını gözlemlemek ve ekonomik, bilimsel, teknolojik gelişmelerini incelemek için bilgiye ihtiyaç duyulmaktadır. Karar alıcı konumunda bulunan araştırmacılar, politikacılar, yöneticiler ve diğer kişi, kurum ve kuruluşlar ihtiyaç duydukları bu bilgilere ulaşmak için istatistik verilerinden yararlanırlar.

Bilginin elde edilmesinde kaynak olan istatistik verilerinin üretilmesi için başvurulan ilk yöntem tamsayım yöntemidir. Tamsayım yönteminde ilgilenilen karakteristiğe yönelik olarak yığını oluşturan tüm birimlerden gerekli bilgiler toplanır. Tamsayım yönteminin yüksek maliyetli olması, zaman alıcı olması vb. gibi çeşitli nedenlerden dolayı tamsayım yöntemiyle bilgi derlemek mümkün olmamaktadır. Bu nedenle araştırmacıların büyük çoğunluğu yığını temsil edebilecek, yığının özelliklerini yansıtabilecek bir alt gruptan faydalanarak bilgi üretir.

İşte, yığını temsil edebilecek nitelikte bir miktar birimin oluşturduğu bu alt gruba örnek, yığından örnek seçme işlemine de örnekleme adı verilir (Çıngı, 1994:2). Resmi kayıtların olmadığı veya ihtiyacı karşılayamadığı koşullarda istatistiki bilginin temel kaynağını örnekleme araştırmaları oluşturur.

Çalışmanın bu bölümünde örneklemenin avantaj ve dezavantajları ile örnekleme teorisindeki temel tanım ve kavramlar açıklanacaktır.

1.1 ÖRNEKLEMENİN TARİHÇESİ, AVANTAJ VE DEZAVANTAJLARI

İlk araştırmalar Antik Roma ve Mısır dönemine kadar dayanır. Askere alma ve vergilendirme amacı için yapılan bu araştırmalar tamsayım yöntemiyle yapılmaktaydı. 18. ve 19. yüzyılda araştırmalar sosyal problemler üzerinde odaklanmaya başlamıştır. 1895 yılında Uluslararası İstatistik Kurumu (ISI) toplantısında Anders Kaier sosyal araştırmalar için tamsayım yerine temsili örnekleme yönteminin kullanımını önermiştir. Bu toplantıda araştırmalarda olasılık modellerinin kullanımı da tartışılmıştır. ISI tarafından 1925 yılında yapılan toplantıda tamsayım araştırmalarına alternatif olarak örnek araştırmalarının kullanımı kabul edilmiştir.

Neyman (1934) makalesinde yaptığı çalışma ile olasılıklı yöntemlere dayanan örnekleme yöntemlerinin kullanımını güçlendirmiştir. Deming (1944) örnekleme yöntemleri ve yanlılık başta olmak üzere örnekleme araştırmaları hakkında çalışmalar yapmıştır. Neyman başta olmak üzere Stephan (1948), Yates (1949), Deming (1950, 1960), Cochran (1964, 1977) ve Kish (1965) makale ve kitaplarıyla örnekleme teorisinin tanıtılmasına ve geliştirilmesine katkıda bulunmuşlardır.

Örnekleme yöntemine dayanarak yapılan araştırmaların tamsayım ile yapılan araştırmalara göre bazı avantajları ve dezavantajları bulunmaktadır.

<u>Avantajları</u>

- Tamsayıma göre daha kısa zaman, daha az işgücü ve maliyet gerektirmektedir.
- Bilgiye daha çabuk ulaşmak mümkündür.
- Daha az birimden veri sağlandığından daha ayrıntılı bilgi alma olanağı vardır.

- Vasıflı eleman kullanımı tamsayıma göre olanaklı olduğundan anketör, veri girişi vb. gibi örnek dışı hata kaynaklarının kontrolü mümkündür.

Dezavantajları

- Örnekleme araştırmaları belirli zaman aralıklarındaki küçük değişimleri yakalayamaz.
- Küçük alt gruplara ilişkin istatistik istenilmesi durumunda örnek yeterli temsili sağlayamaz. Bu durumda tamsayım yapmak gereklidir.

1.2 ÖRNEKLEMEDE TEMEL TANIM VE KAVRAMLAR

Bu bölümde örneklemeyi ve örneğe dayalı olarak yapılan araştırma sonuçlarını daha iyi anlamak ve yorumlayabilmek için gerekli tanım ve kavramlar hakkında kısaca bilgi verilecektir.

Yığın ve örnek: Üzerinde araştırma yapılan herhangi bir canlılar veya cansızlar topluluğuna ana kütle, kitle, yığın veya popülasyon denilmektedir. Tezin bundan sonraki kısmında bu kavram için yığın terimi kullanılacaktır. Yığını temsil edebilecek nitelikte bir miktar birimin oluşturduğu alt gruba örnek denir (Çıngı, 1994:2). İçerisinden örneklerin seçildiği yığına "örneklenen yığın", hakkında bilgi toplanmak istenen yığına ise "hedef yığın" adı verilir (Serper ve Aytaç, 2000:4). Örneklenen yığın ve hedef yığın araştırmalarda aynı veya farklı olabilirler. Sonlu sayıda ya da sayılabilir çoklukta birim içeren yığına sonlu yığın denir. Sonsuz sayıda birim içeren yığınlara sonsuz yığın denir. Teorik olarak yığın büyüklüğü çok fazla olan sonlu yığınlar, sayılarının belirlenmesi gerek zaman ve gerek ekonomik bakımdan mümkün olmadığı için sonsuz yığın olarak kabul edilebilir.

Birim: Yığının sınırlandırılabilen, ayırt edilebilen ve çerçevede ayrı ayrı gösterilebilen her bir parçasıdır. Çerçevedeki bu birimler tamsayımlar için istatistik birimi, örnekleme uygulamaları için örnekleme birimi olarak adlandırılır. Gözlem birimi ise hakkında ayrı bilgi toplanan yığının küçük bir parçasıdır. Örnekleme birimi ile gözlem birimi genellikle aynı olmakla beraber farklı da olabilirler. Örneğin hanehalkı örnekleme birimi iken hanehalkı fertleri gözlem birimi olarak alınır (Esin, 1975:22-25). Araştırma sonunda elde edilen verilerin analiz edildiği ve sonuçların verildiği birime de analiz birimi denir.

Çerçeve: Örnekleme yönteminin uygulanabilmesi için örneklerin seçileceği araştırma yapılan yığının fiziksel biçimde belirtilmesi gerekmektedir. En basit anlatımı ile çerçeve örnekleme birimlerini tamamen kapsayarak gösterimini sağlayan yapıdır (Verma,1998:5-2). Yardımcı değişken kullanımında çerçevelerin önemi büyüktür.

Rassal değişken ve karakteristik: Örnek uzayının her bir noktasını gerçek bir sayıya bağlayan fonksiyona rassal değişken denir (Çıngı, 1994:5). Yığın ile ilgili herhangi bir nicelik veya ilişki o yığının bir karakteristiğini gösterir (Hansen, Hurwitz, Madow, 1964:2). Araştırmalarda bir rassal değişken ile ilgili nicelik bir karakteristiği verebildiği gibi birden çok rassal değişkenin bir araya gelmesi ile oluşan nicelik de bir karakteristiği verebilir. Örneğin bir araştırmada tarımsal işletmelere ait keçi sayısı ve koyun sayısı ayrı ayrı rassal değişken olup toplam koyun ve keçi sayısı da ayrı ayrı karakteristiklerdir. Diğer yandan keçi ve koyun sayısının toplamları da küçükbaş hayvan sayısını verir ve bu da iki rassal değişken yardımıyla tanımlanan bir karakteristiktir.

Yığın değerleri ve istatistikler: Yığının bütün N elemanı için bazı karakteristiklerin değerlerini gösteren sayısal gösterime yığın değeri adı verilir. İstatistik veya örnek değeri ise n elemanlı bir örnekten hesaplanan tahmin değeridir (Kish,1965:9).

Tahmin edici ve tahmin: "Sonlu yığından çekilen örnekten yararlanılarak, yığın parametrelerinin özelliklerini tahmin etmek amacıyla tanımlanan matematiksel eşitliğe tahmin edici denir." (Çıngı, 1994:4). Herhangi bir yığında mevcut olan bir değere ait ölçüme, örnek verisi kullanılarak ulaşılması sürecine ise tahmin denir. Bu ölçüm bilinmeyen değerin bir göstergesidir. Tahmin sonuçları nokta tahmini olarak adlandırılan tek bir değer olarak ya da bilinen değerler aralığı olarak açıklanabilir.

Örnek dağılımı, örnekleme oranı ve genişletme katsayısı: Bir tahminin örnek dağılımı her birinin P gerçekleşme olasılığı ile o tahminin olası bütün değerlerinin teorik dağılımıdır. Bu olası değerler ve olasılıklar örnekleme tasarımına bağlıdır (Kish,1965:11). Örnek büyüklüğü n'nin yığın büyüklüğü N'ye oranına (n/N=f) örnekleme oranı veya kapsama olasılığı, örnekleme oranının tersi olan ve örneklerin yığına genişletilmesinde kullanılan (N/n=d) oranına ise genişletme katsayısı veya tasarım ağırlığı denir.

Varyans, standart sapma, standart hata: Yığındaki her bir birimin yığın ortalamasından olan uzaklığının ortalama ölçüsüne varyans ($S^2 = \dfrac{\sum(Y_i - \overline{Y})^2}{N-1}$) denir. Varyansın karekök değerine de standart sapma adı verilir. n genişliğindeki mümkün olan tüm örnekler üzerinden ortalama alındığında, bu ortalama değerlerin gerçek yığın ortalama değerinden ayrılışlarının kareleri örnek ortalamalarının varyansını verir ($V(\hat{\theta}) = E(\hat{\theta} - E(\hat{\theta}))^2$). Bu değerin karekökü ise n çaplı ortalamaların standart hatası olarak tanımlanır.

Yansızlık: Bir parametreye ilişkin tahmin edicinin beklenen değeri $E(\hat{\theta})$ yığın değeri θ'ya eşit olabilir ya da olmayabilir. Bu iki değer arasındaki farka ($E(\hat{\theta}) - \theta$) örnek yanlılığı denir. $E(\hat{\theta}) = \theta$ olması durumunda örnek tasarımının yansız olduğu söylenir (Kish, 1965:11). İyi bir tahmin edicinin yansız olması beklenir.

Tutarlılık: İyi bir tahmin edicinin ikinci özelliği de tutarlılıktır. Örnek hacmi büyüdükçe tahminler tahmin edilen yığın değerine yaklaşıyorsa bu tahmin ediciye tutarlı tahmin edici denir. Matematiksel olarak ifade etmek gerekirse; $\hat{\theta}$'nın θ değerini verme olasılığı n büyüdükçe 1'e yaklaşır, $n \to \infty$ ise $P(\hat{\theta} \to \theta) \to 1$ (Yamane, 1967:41).

Duyarlılık ve değişim katsayısı: Bir tahmin edicinin varyansının tersi alınarak bulunan değere duyarlılık ölçüsü ($1/V(\hat{\theta})$) denilmektedir. Varyansı küçük olan tahmin ediciler, yüksek duyarlılığa sahip olurlar. İki farklı tahmini birbiri ile karşılaştırırken ölçü birimleri ve değişken değerlerinin farklı düzeylerinden ötürü duyarlılığı varyans ile ölçmek doğru olmaz. Bunun yerine tahminin standart hatasının kendisine bölümü ile elde edilen değişim katsayısı ($sh(\hat{\theta})/\hat{\theta}$) kullanılır (Çıngı, 1994:10).

Hata kareler ortalaması, doğruluk: Bir tahmin edici ile tahmin ettiği yığın değeri arasındaki farkın karesinin beklenen değerine hata kareler ortalaması (HKO=$E(\hat{\theta}-\theta)^2$) denir. Hata kareler ortalamasının gösterimi varyans ve yanlılık ile de yapılabilir.

$$\text{HKO} = V(\hat{\theta}) + (Yanlılık)^2$$

Yanlılığın olmadığı yerde varyans hata kareler ortalamasına eşit olurken, yanlılığın olduğu durumlarda varyans yerine HKO kullanılır. Bir tahmin edicinin hata kareler ortalamasının küçük olması istenir. Tahmin edicinin hata kareler ortalaması küçüldükçe doğruluğu artar.

Etkinlik: Eğer belirlenen koşullarda birim maliyet için bir örnek tasarımı diğerinden daha güvenilir sonuçlar veriyorsa o örnek tasarımının diğerinden daha etkin olduğu söylenir (Hansen, Hurwitz, Madow, 1964:34). Yansız tahmin edicilerin etkinlikleri varyans, yanlı tahmin edicilerin etkinlikleri ise hata kareler ortalamaları ile karşılaştırılır. Varyansı veya hata kareler ortalaması küçük olan tahmin ediciler karşılaştırıldıkları tahmin edicilerden daha etkindir.

1.3. ÖRNEKLEMEDE YARDIMCI DEĞİŞKENLERİN KULLANIMI

Örnek araştırmalarında ilgilenilen değişken Y'nin dışında bir veya birden fazla yardımcı değişken X de yığın birimleriyle ilişkilendirilebilir. Yardımcı bilgi örnekleme tasarımında ve tahmin aşamasında kullanılabilir. Yardımcı bilgi tasarım aşamasında genellikle tabakalama için veya büyüklüğe orantılı olasılıksal örnekleme yöntemlerinde kullanılır (Mussa,1999). Yardımcı değişkenler ilgilenilen değişkeninin tahmini aşamasında ise, daha iyi sonuçlar veren tahmin ediciler elde etmek için kullanılır.

Yardımcı değişken değerleri, idari kayıtlardan veya diğer kaynaklardan elde edilen değerlerin örnek çerçevesindeki birimlerle eşleştirilmesi sonucunda elde edilebilir. Tahmin aşamasında yardımcı değişkenler, sadece örnekte bulunma olasılığı ile değil açık biçimde tahmin edicinin formülüne de girer. Yardımcı değişken kullanımının arkasındaki temel varsayım onların ilgilenilen değişken ile ilişkili olması ve böylece ilgilenilen değişken hakkında bilgi taşımasıdır (Särndal, Swensson, Wretman, 1992:220). Bu özellik sayesinde belli koşullar altında yardımcı değişken kullanılarak yapılan tahminler, klasik tahminlerden daha güvenilir ve daha etkin sonuçlar verir.

İKİNCİ BÖLÜM

BASİT RASGELE ÖRNEKLEME YÖNTEMİNDE TAHMİN

Basit rasgele örnekleme; n farklı birimin N büyüklüğündeki yığından seçildiği ve tüm olası n birimli kombinasyonların eşit seçilme olasılığının olduğu örnekleme tasarımıdır. Basit rasgele örnekleme yönteminde yığının i'nci biriminin örnekte bulunma olasılığı π_i=n/N olduğundan yığındaki tüm birimlerin örnekte bulunma olasılığı eşittir (Thompson,1992:11).

N büyüklüğündeki bir yığın için Y ilgilendiğimiz değişken, $X=(X_1,X_2,X_3,.....,X_p)$ ise p tane değişken içeren yardımcı değişkenler vektörü olsun. İlgilendiğimiz Y değişkeninin ortalama parametresini tahmin etmek için basit rasgele örnekleme yöntemiyle n büyüklüğünde çekilen örnek birimlerinin ölçümleri de y ve $x=(x_1,x_2,x_3,.....,x_p)$ olmak üzere; bu bölümde yığın ortalaması $\overline{Y}$'yı tahmin etmek için kullanılan bazı tahmin ediciler incelenecektir.

2.1 HORVİTZ-THOMPSON TAHMİN EDİCİSİ

Horvitz ve Thompson (1952) yerine koyularak veya yerine koyulmadan yapılan örnek tasarımlarının herhangi biri için, örnekte bulunan i'nci birimin π_i olasılık değeri biliniyorsa yığın toplamı τ'nin yansız tahmini için

$$\hat{\tau}_\pi = \sum_{i=1}^{n} \frac{y_i}{\pi_i} \tag{2.1}$$

tahmin edicisini önermişlerdir (Thompson,1992:49). Horvitz-Thompson tahmin edicisi (HTE) yansız olması, basit kullanımı ve istatistiksel sunumunun kolaylığı nedeniyle geleneksel olarak basit rasgele örnekleme tasarımında kullanılmaktadır. Yığın ortalamasını tahmin etmek için HTE Denklem (2.2)'deki gibi düzenlenebilir.

$$\hat{\bar{Y}} = \frac{1}{N}\sum_{i=1}^{n}\frac{y_i}{\pi_i} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y} \tag{2.2}$$

HTE'de yapılan düzenleme sonucunda basit rasgele örneklemede yığın ortalamasının HTE'sinin örnek ortalaması olduğu görülmektedir. $\bar{y}$'nın varyansı ve varyansın yansız tahmin edicisi sırasıyla Denklem (2.3) ve (2.4)'deki gibidir (Hedayat ve Sinha,1991:71).

$$V(\bar{y}) = \frac{(1-f)}{n} S_y^2 \tag{2.3}$$

$$\hat{V}(\bar{y}) = \frac{(1-f)}{n} s_y^2 \tag{2.4}$$

2.2 ORAN TAHMİN EDİCİLERİ

X yardımcı değişkenler vektörümüzün sadece bir tane yardımcı değişken içerdiğini varsayalım. X değişkeninin yığın değerleri X=(X_1, X_2, $X_{3,\ldots\ldots}$, X_N), yığın toplamı veya yığın ortalamasını biliyorsak; $\frac{\bar{y}}{\bar{x}}$ oranı yığın bazında çok büyük değişkenlik göstermiyorsa oran tahmini kullanılabilir. Bu oranın fazla değişkenlik göstermemesi için örnekteki x_i ve y_i değerlerinin orijinden geçen bir doğru oluşturması gereklidir. Özetle $\frac{\bar{y}}{\bar{x}} = \hat{R}$ gibi bir eşitlik sağlanabilmelidir.

2.2.1 Klasik Oran Tahmin Edicisi

Klasik oran tahmininde amaç örneklemeden elde edilen y_i değerlerinin x_i değerleri ile olan ilişkisinden yararlanılarak $\overline{Y}$'nın tahmininin varyansını düşürmektir. y_i ilgilenilen değişken ile x_i yardımcı değişkeni arasındaki ilişki pozitif olduğu zaman $\overline{Y}$'nın klasik oransal tahmini Denklem (2.5)'teki eşitlik ile yapılır.

$$\hat{\overline{Y}}_{ORAN} = \frac{\overline{y}}{\overline{x}}\overline{X} = \hat{R}\overline{X} \tag{2.5}$$

Klasik oran tahmininin yanı ve hata kareler ortalamasının (HKO) hesaplanması için $\hat{R}$'nin incelenmesi gereklidir. $\frac{\overline{Y}}{\overline{X}} = R$ olmak üzere $\hat{R}$'nin yanı;

$$\text{Yan}(\hat{R}) = E(\hat{R} - R) = E\left(\frac{\overline{y}}{\overline{x}} - R\right) \tag{2.6}$$

eşitliği ile gösterilir. Denklem (2.6)'da Taylor Serisi açılımı kullanılarak $\hat{R}$'nin yanı Denklem (2.7)'deki gibi bulunur. Burada ρ_{yx} yığın korelasyon katsayısını göstermektedir.

$$\text{Yan}(\hat{R}) \cong \frac{1-f}{n}\frac{1}{\overline{X}^2}(RS_x^2 - \rho_{yx}S_yS_x) \tag{2.7}$$

$\hat{R}$'nin yanı kullanılarak klasik oran tahmininin yanı Denklem (2.8) ve (2.9)'da gösterildiği gibi hesaplanabilir.

$$\text{Yan}(\hat{\bar{Y}}_{ORAN})=\text{Yan}(\hat{R})\bar{X} \tag{2.8}$$

$$\text{Yan}(\hat{\bar{Y}}_{ORAN})=\frac{1-f}{n}\frac{1}{\bar{X}}(RS_x^2-\rho_{yx}S_yS_x) \tag{2.9}$$

$\text{RS}_x^2\text{-}\rho_{yx}\text{S}_y\text{S}_x = 0$ olduğu durumda klasik oran tahmin edicisi yansız olur. Değişkenler arasındaki ilişkinin miktarı arttıkça, yani ρ_{yx} korelasyon katsayısı 1'e yaklaştıkça, yan azalır (Çıngı,1994). $\hat{\text{R}}$'nin hata kareler ortalaması Taylor doğrusallaştırması yardımıyla Denklem (2.10)'daki gibi bulunur (Särndal, Swensson, Wretman, 1992:179).

$$\hat{R}=R+\frac{1}{\bar{X}}\frac{1}{n}\sum_{i=1}^{n}(y_i-Rx_i)=R+\frac{\bar{y}-R\bar{x}}{\bar{X}}$$

$$HKO(\hat{R})=\frac{1}{\bar{X}^2}\frac{1-f}{n}\frac{1}{N-1}\sum_{i=1}^{N}(y_i-Rx_i)^2$$

$$=\frac{1}{\bar{X}^2}\frac{1-f}{n}(S_y^2+R^2S_x^2-2RS_yS_x) \tag{2.10}$$

Klasik oransal tahmin edicinin hata kareler ortalaması ise Denklem (2.11)'deki gibi elde edilebilir.

$$\begin{aligned} HKO(\hat{\bar{Y}}_{ORAN}) &= \bar{X}^2 HKO(\hat{R}) \\ HKO(\hat{\bar{Y}}_{ORAN}) &= \frac{1-f}{n}(S_y^2+R^2S_x^2-2R\rho_{yx}S_yS_x) \end{aligned} \tag{2.11}$$

$C_y=\frac{\text{S}_y}{\bar{\text{Y}}}$ ve $C_x=\frac{\text{S}_x}{\bar{\text{X}}}$ değişim katsayıları olmak üzere klasik oransal tahminin hata kareler ortalaması Denklem (2.12)'deki gibi de bulunabilir.

$$HKO(\hat{\bar{Y}}_{ORAN}) = \frac{1-f}{n}\bar{Y}^2(C_y^2 + C_x^2 - 2\rho_{yx}C_yC_x) \quad (2.12)$$

ρ_{yx} korelasyon katsayısı 1 olduğunda $\text{HKO}(\hat{\bar{Y}}_{ORAN})$ en küçük değerini alır. Yani klasik oransal tahmin edici, X_i ve Y_i arasında pozitif tam bir ilişki olduğunda minimum hata kareler ortalamasına sahip olur.

Klasik oran tahmini her zaman HTE tahmininden daha iyi sonuçlar vermeyebilir. Klasik oran tahmininin HTE tahmininden daha etkin bir sonuç verebilmesi için $\text{HKO}(\hat{\bar{Y}}_{ORAN}) < \text{Var}(\bar{y})$ eşitsizliğinin sağlanması gerekir. Denklem (2.3) ve Denklem (2.11)'de eşitlikleri yerine koyarak karşılaştırma yaparsak;

$$\begin{aligned}
HKO(\hat{\bar{Y}}_{ORAN}) &< Var(\bar{y}) \\
\frac{1-f}{n}(S_y^2 + R^2S_x^2 - 2R\rho_{yx}S_yS_x) &< \frac{(1-f)}{n}S_y^2 \\
R^2S_x^2 &< 2R\rho_{yx}S_yS_x \\
\frac{\bar{Y}}{\bar{X}}\frac{S_x}{S_y} &< 2\rho_{yx} \\
\frac{C_x}{2C_y} &< \rho_{yx}
\end{aligned} \quad (2.13)$$

koşulunun sağlandığı X_i ve Y_i değişkenlerinde yığın ortalaması için klasik oran tahmini HTE'den daha etkin tahmin üretir. Klasik oran tahmin edicisinin yanı genellikle çok küçüktür. Ancak küçük örneklerde yanlılık gözardı edilemeyecek kadar önemli olabilir. Diğer taraftan örnek hacmi 20 ve daha büyük olan örneklerde yanlılığın getirdiği sonuçlar ihmal edilebilir. Bazı araştırmacılar örnek tasarımını veya tahmin ediciyi değiştirerek klasik oran tahmininin yanlılığını ortadan kaldırmaya veya azaltmaya çalışmışlardır (Särndal, Swensson, Wretman, 1992:251).

2.2.2 Beale'in Oran Tahmin Edicisi

Klasik oran tahmin edicisine alternatif olarak çeşitli oran tahmin edicileri türetilmiştir. Bu tahmin ediciler daha düşük yanlılık yaratarak kullanıcılara avantaj sağlar. Beale (1962) tarafından önerilen oransal tahmin edici Denklem (2.14)'deki gibidir (Tin, 1965).

$$\hat{\bar{Y}}_{BEALE} = \left(\frac{\bar{y}}{\bar{x}}\right)\left(\frac{1+\nu C_{yx}}{1+\nu C_x^2}\right)\bar{X} \tag{2.14}$$

$\nu = \dfrac{1\text{-f}}{\text{n}}$ olmak üzere Beale tahmin edicisinin yanı ve hata kareler ortalaması ise sırasıyla Denklem (2.15) ve Denklem (2.16)'da verilmiştir.

$$\text{Yan}(\hat{\bar{Y}}_{BEALE}) = \nu\bar{Y}\left(\frac{1+\nu C_{yx}}{1+\nu C_x^2}\right)(C_x^2 - \rho_{yx}C_yC_x) \tag{2.15}$$

$$HKO(\hat{\bar{Y}}_{BEALE}) = \nu\bar{Y}^2\left(\frac{1+\nu C_{yx}}{1+\nu C_x^2}\right)^2(C_y^2 + C_x^2 - 2\rho_{yx}C_yC_x) \tag{2.16}$$

Beale oran tahmin edicisinin klasik oran tahmin edicisinden daha etkin tahmin üretebilmesi için bu iki tahmin edicinin hata kareler ortalamaları $\text{HKO}(\hat{\bar{Y}}_{\text{BEALE}}) < \text{HKO}(\hat{\bar{Y}}_{\text{ORAN}})$ olacak şekilde karşılaştırılırsa,

$$HKO(\hat{\bar{Y}}_{BEALE}) \;<\; HKO(\hat{\bar{Y}}_{ORAN})$$

$$\nu\bar{Y}^2\left(\frac{1+\nu C_{yx}}{1+\nu C_x^2}\right)^2(C_y^2+C_x^2-2\rho_{yx}C_yC_x)<\nu\bar{Y}^2(C_y^2+C_x^2-2\rho_{yx}C_yC_x)$$

$$\left(\frac{1+\nu C_{yx}}{1+\nu C_x^2}\right)^2<1 \tag{2.17}$$

$$-1<\left(\frac{1+\nu C_{yx}}{1+\nu C_x^2}\right)<1$$

koşulunun sağlanması gerektiği görülmektedir.

2.2.3 Tin'in Oran Tahmin Edicisi

Tin (1965), Beale tarafından önerilen tahmin edicinin ardından değiştirilmiş bir tahmin edici geliştirmiştir. Tin'in önerdiği tahmin edici Denklem (2.18)'de gösterilmiştir.

$$\hat{\bar{Y}}_{TIN}=\left(\frac{\bar{y}}{\bar{x}}\right)[1+\nu(C_{yx}-C_x^2)]\bar{X} \tag{2.18}$$

Tin tahmin edicisini yanlılık, etkinlik ve normalliğe yakınsama açısından klasik oran tahmin edicisi ve Beale oran tahmin edicisi ile karşılaştırmıştır. Tin, X ve Y'nin iki değişkenli normal dağıldığı durumlar gibi belli koşullarda $\hat{\bar{Y}}_{TIN}$ tahmin edicisinin diğerlerinden daha etkin olduğu sonucuna varmıştır (Mussa,1999). $\hat{\bar{Y}}_{TIN}$ tahmin edicisinin yanı ve hate kareler ortalaması sırasıyla Denklem (2.19) ve Denklem (2.20)'de verilmiştir.

$$\text{Yan}(\hat{\bar{Y}}_{TIN})=\nu\bar{Y}[1+\nu(C_{yx}-C_x^2)](C_x^2-\rho_{yx}C_yC_x) \tag{2.19}$$

$$HKO(\hat{\bar{Y}}_{TIN}) = \nu\bar{Y}^2[1+\nu(C_{yx}-C_x^2)]^2(C_y^2+C_x^2-2\rho_{yx}C_yC_x) \tag{2.20}$$

Tin oran tahmin edicisinin klasik oran tahmin edicisinden daha etkin tahmin üretebilmesi için bu iki tahmin edicinin hata kareler ortalamaları $\text{HKO}(\hat{\bar{Y}}_{TIN}) < \text{HKO}(\hat{\bar{Y}}_{\text{ORAN}})$ olacak şekilde karşılaştırılırsa,

$$\begin{gathered} HKO(\hat{\bar{Y}}_{TIN}) < HKO(\hat{\bar{Y}}_{ORAN}) \\ \nu\bar{Y}^2[1+\nu(C_{yx}-C_x^2)]^2(C_y^2+C_x^2-2\rho_{yx}C_yC_x) < \nu\bar{Y}^2(C_y^2+C_x^2-2\rho_{yx}C_yC_x) \\ [1+\nu(C_{yx}-C_x^2)]^2 < 1 \\ -1 < [1+\nu(C_{yx}-C_x^2)] < 1 \end{gathered} \tag{2.21}$$

koşulunun sağlanması gerektiği görülmektedir.

2.2.4 Singh ve Tailor'un Oran Tahmin Edicisi

Singh ve Tailor (2003), X_i ve Y_i değişkenlerinin aralarındaki korrelasyonun pozitif olması durumunda kullanılabilecek yeni bir oran tahmin edicisi türetmişlerdir.

$$\hat{\bar{Y}}_{SINTAIL} = \bar{y}\frac{\bar{X}+\rho_{yx}}{\bar{x}+\rho_{yx}} \tag{2.22}$$

$\dfrac{\bar{X}}{\bar{X}+\rho_{yx}} = \omega$ ve $\rho_{yx}\dfrac{C_y}{C_x} = \theta$ olmak üzere $\hat{\bar{Y}}_{\text{SINTAIL}}$ tahmin edicisinin yanı ve hate kareler ortalaması ise sırasıyla Denklem (2.23) ve Denklem (2.24)'deki gibi olmaktadır (Kadılar ve Çıngı, 2006).

$$\text{Yan}(\hat{\bar{Y}}_{SINTAIL}) = \nu\bar{Y}\left(\omega^2 C_x^2 - \omega\rho_{yx}C_yC_x\right) \tag{2.23}$$

$$HKO(\hat{\bar{Y}}_{SINTAIL}) = \nu\bar{Y}^2\left[C_y^2 + C_x^2\omega(\omega - 2\theta)\right] \tag{2.24}$$

Singh Tailor oran tahmininin klasik oran tahmininden daha etkin bir sonuç verebilmesi için $\text{HKO}(\hat{\bar{\text{Y}}}_{\text{SINTAIL}}) < \text{HKO}(\hat{\bar{\text{Y}}}_{\text{ORAN}})$ eşitsizliğinin sağlanması gerekir. Denklem (2.12) ve Denklem (2.24)'de eşitlikleri yerine koyarak karşılaştırma yaparsak;

$$\begin{gathered}
HKO(\hat{\bar{Y}}_{SINTAIL}) < HKO(\hat{\bar{Y}}_{ORAN}) \\
\nu\bar{Y}^2\left[C_y^2 + C_x^2\omega(\omega - 2\theta)\right] < \nu\bar{Y}^2(C_y^2 + C_x^2 - 2\rho_{yx}C_yC_x) \\
C_y^2 + C_x^2\omega(\omega - 2\rho_{yx}\frac{C_y}{C_x}) < C_y^2 + C_x^2 - 2\rho_{yx}C_yC_x \\
C_x^2\omega^2 - C_x^2 < \omega 2\rho_{yx}C_yC_x - 2\rho_{yx}C_yC_x \\
C_x^2(\omega - 1)(\omega + 1) < (\omega - 1)2\rho_{yx}C_yC_x
\end{gathered} \tag{2.25}$$

$(\omega - 1) < 0$ olduğu durumda korelasyon pozitif olacağından eşitsizliğin iki tarafı -1 ile çarpılarak $C_x^2(\omega-1)(\omega+1) > (\omega-1)2\rho_{yx}C_yC_x$ elde edilir.

$$\rho_{yx} < \frac{\text{C}_\text{x}}{2\text{C}_\text{y}}(1+\omega) \tag{2.26}$$

Sonuçta Denklem (2.26) koşulunun sağlandığı X_i ve Y_i değişkenlerinde yığın ortalaması için Singh Tailor oran tahmini klasik oran tahmininden daha etkin tahmin üretir.

2.2.5 Srivastava İki Değişkenli Oran Tahmin Edicisi

X yardımcı değişkenler vektörümüzün iki yardımcı değişkenden oluştuğu X=(X_1, X_2), X_1 yardımcı değişkeni ile Y arasında pozitif korelasyon X_2 yardımcı değişkeni ile Y arasında negatif korelasyon olduğu durumda ve yardımcı değişkenlere ilişkin yığın bilgilerinin bilindiği durumda Srivastava (1965) Denklem (2.27)'deki iki değişkenli oran tahmin edicisini önermiştir.

$$\hat{\bar{Y}}_{SRIV} = \bar{y}\left(d_1 \frac{\bar{X}_1}{\bar{x}_1} + d_2 \frac{\bar{x}_2}{\bar{X}_2} \right) \tag{2.27}$$

$d_1 + d_2 = 1$ olacak şekilde seçilen ağırlıklardır (Upadhyaya ve Singh, 2003).

$\hat{\bar{Y}}_{SRIV}$ tahmin edicisinin yanı ve hate kareler ortalaması ise sırasıyla Denklem (2.28) ve Denklem (2.29)'daki gibi olmaktadır (Karakülah, 2006).

$$\text{Yan}(\hat{\bar{Y}}_{SRIV}) = \nu\bar{Y}\left(d_1 C_{x_1}^2 - d_1 \rho_{yx_1} C_y C_{x_1} + d_2 \rho_{yx_2} C_y C_{x_2} \right) \tag{2.28}$$

$$HKO(\hat{\bar{Y}}_{SRIV}) = \nu\bar{Y}^2 \left[C_y^2 + d_1^2 C_{x_1}^2 + d_2^2 C_{x_2}^2 - 2d_1 \rho_{yx_1} C_y C_{x_1} + 2d_2 \rho_{yx_2} C_y C_{x_2} - 2d_1 d_2 \rho_{x_1x_2} C_{x_1} C_{x_2} \right] \tag{2.29}$$

Srivastava iki değişkenli oran tahmininin klasik oran tahmininden daha etkin bir sonuç verebilmesi için $HKO(\hat{\bar{Y}}_{SRIV}) < HKO(\hat{\bar{Y}}_{ORAN})$ eşitsizliğinin sağlanması gerekir. Bu eşitsizliğin sağlanması için gerekli sadeleştirilmiş bir koşul elde edilememiştir.

2.3 REGRESYON TAHMİN EDİCİLERİ

X yardımcı değişkenler vektörümüzdeki değişkenlerin Y ilgilenilen değişken ile ilişkisi koordinat düzleminde bir doğru oluşturuyorsa ancak bu doğru orijinden geçmiyorsa oran tahmin edicisi yerine regresyon tahmin edicisi kullanılmalıdır.

Açıklayıcı değişkenin bağımsız olduğu bilinen doğrusal regresyon durumunda yığın ortalamasının standart regresyon tahmin edicisi sonlu yığın örneklemesi açısından yanlı olarak değerlendirilir; fakat model tabanlı teoriye göre standart regresyon tahmin edicisi yansızdır (Cochran, 1977).

Regresyon tahmin edicileri doğrusal tahmin ediciler sınıfına girmektedir. Doğrusal tahmin edicilerin örnek araştırmalarında özel bir avantajı vardır. Bir kez hesaplanan ağırlıkların kullanımı diğer analiz değişkenleri için de kullanılabilir (Fuller, 2002).

2.3.1 Klasik Doğrusal Regresyon Tahmin Edicisi

Y bağımlı değişken X bağımsız değişken olmak üzere basit yığın doğrusal regresyon denklemi, beklenen değer ve varyansı Denklem (2.30)'daki gibi olmaktadır.

$$\begin{aligned} Y_i &= \alpha + \beta X_i + \varepsilon_i \\ E(Y_i) &= \alpha + \beta X_i \\ V(Y_i) &= \sigma^2 \end{aligned} \tag{2.30}$$

Yığın regresyon denkleminde ε_i hata terimidir. Örnekleme yöntemiyle elde edilen y ve x değişkenleri için örnek doğrusal regresyon denklemi;

$$y_i = \text{a} + bx_i + e_i \quad \text{ve} \quad b = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{s_{xy}}{s_x^{\ 2}}$$

olmak üzere klasik doğrusal regresyon tahmin edicisi Denklem (2.31)'deki gibi elde edilir. Örnek regresyon denklemindeki e_i değeri artık değerlerini göstermektedir.

$$\hat{\overline{Y}}_{REG} = \overline{y} + b\left(\overline{X} - \overline{x}\right) \tag{2.31}$$

b yığın regresyon denklemindeki β parametresinin tahminidir ve farklı örneklerde değişik sonuçlar vermektedir. b değerinin değişkenlik göstermesi istenmiyorsa periyodik araştırmalar için daha önceki çalışmalardan elde edilen β^* parametresi veya pilot çalışma sonucu elde edilen b^* değeri sabit olarak alınıp b'nin yerine kullanılabilir. Bu durumda ortalamaya ilişkin doğrusal tahmin fark tahmin edicisi olarak adlandırılır.

Klasik doğrusal regresyon tahmin edicisinde de amaç oran tahmin edicilerinde olduğu gibi HTE tahminini geliştirmektir. Denklem (2.31)'de açık bir biçimde görüldüğü gibi klasik doğrusal regresyon tahmin edicisi HTE ve düzeltme teriminden oluşmaktadır. Klasik doğrusal regresyon tahmin edicisi iyi çalıştığında düzeltme terimi sıklıkla HTE tahmininin hatasıyla negatif olarak ilişkilidir. HTE'nin büyük hatalar verdiği örneklerde, örnek hacmi yeterince büyük ve doğrusal ilişki kuvvetliyse düzeltme terimi de yaklaşık olarak aynı büyüklükte ancak ters işaretlidir. Böylece klasik doğrusal regresyon tahmin edicisi ile elde edilen tahminin hatası HTE tahmininin hatasından daha küçüktür.

Klasik doğrusal regresyon tahmin edicisinde; regresyon modeli uygun b değerini bulmak için bir araçtır. Klasik doğrusal regresyon tahmin edicisinin etkinliği uyum iyiliğine bağlıdır. Diğer yandan yansızlık, varyans formülünün geçerliliği gibi temel özellikler regresyon modelinin varsayımlarının gerçekleşmesinden bağımsızdır. Bu yüzden klasik doğrusal regresyon tahmin edicisi yöntemi model-bağımlı değil model-yardımcı bir yöntemdir (Särndal, Swensson, Wretman, 1992:226-227).

Klasik doğrusal regresyon tahmin edicisi yansız değildir. Ancak yeterince büyük örnek hacmine sahip örneklerde yaklaşık olarak yansızdır. Klasik doğrusal regresyon tahmin edicisinin varyansı ise yaklaşık olarak Denklem (2.32)'deki gibi hesaplanır.

$$V(\hat{\bar{Y}}_{REG}) \cong \nu S_y^{\ 2}(1-\rho_{yx}^{\ 2}) \qquad (2.32)$$

Basit rasgele örnekleme altında klasik regresyon tahmin edicisi genelde HTE ve oran tahmin edicilerinden daha iyi performans gösterir. HTE ve klasik regresyon tahmin edicisinin varyanslarını karşılaştırırsak;

$$\begin{gathered} V(\hat{\bar{Y}}_{REG}) < V(\bar{y}) \\ \frac{1-f}{n} S_y^{\ 2}(1-\rho_{yx}^{\ 2}) < \frac{(1-f)}{n} S_y^2 \\ (1-\rho_{yx}^{\ 2}) < 1 \end{gathered} \qquad (2.33)$$

koşulu sağlandığında klasik regresyon tahmin edicisi HTE'den daha etkin sonuç vermektedir. Sonuç olarak $\rho_{yx} \neq 0$ olduğu sürece $\hat{\bar{Y}}_{REG}$ HTE tahminini geliştirir ve bu gelişme $\rho_{yx} > 0.8$ olduğunda çok fazladır.

Klasik oran tahmin edicisi ile klasik regresyon tahmin edicisinin varyanslarını karşılaştırırsak; $V(\hat{Y}_{REG}) \leq HKO(\hat{Y}_{ORAN})$ eşitsizliği

$$\beta = \frac{\sum_{i=1}^{N}(Y_i - \overline{Y})(X_i - \overline{X})}{\sum_{i=1}^{N}(X_i - \overline{X})^2} = \frac{\overline{Y}}{\overline{X}} \qquad (2.34)$$

olduğu durumda gerçekleşir. Böylece klasik regresyon tahmin edicisi $\beta \neq \overline{Y} / \overline{X}$ olduğu sürece klasik oran tahmin edicisinden daha iyidir.

$\hat{Y}_{REG}$ tahmini daha etkin sonuç vermesine rağmen $\hat{Y}_{ORAN}$ bazı araştırmalarda $\hat{Y}_{REG}$ tahminine tercih edilmektedir. Bunun birçok nedeni vardır. $\hat{Y}_{ORAN}$'ın yapısı daha basittir ve Y değişkenine ilişkin yığın parametreleri dışında yığın ile ilgili $\overline{Y} / \overline{X}$ gibi oran bilgileri de elde edilmek isteniyorsa $\hat{Y}_{ORAN}$'ın kullanımı daha avantajlıdır.

Çok küçük örnek hacimlerinde $\hat{Y}_{ORAN}$'ın varyansı $\hat{Y}_{REG}$'in varyansından daha küçük olabilir. Küçük yığınlar üzerinde 12 veya daha küçük örnek hacimli basit rasgele örnekleme yöntemi ile yapılan ampirik çalışmalar, $\hat{Y}_{ORAN}$'ın hata kareler ortalamasının $\hat{Y}_{REG}$'in hata kareler ortalamasından daha küçük olduğunu ortaya koymuştur. Bu çalışmalarda hata kareler ortalamasının büyük olmasının nedeninin bazen yanlılıktan değil varyanstan kaynaklı olduğu görülmüştür. Bu yüzden çok küçük örnek hacimleri de klasik oran tahmin edicisinin klasik doğrusal regresyon tahmin edicisine tercih edilmesinin nedenlerindendir. Ancak gerekli koşullar sağlandığında klasik doğrusal regresyon tahmin edicisi $\hat{Y}_{REG}$'in kullanılamaması doğruluk kaybına neden olur (Särndal, Swensson, Wretman, 1992:274).

2.3.2 Kaur'un Regresyon Tahmin Edicisi

Kaur (1985) tarafından önerilen regresyon tipi tahmin edici, yığın ortalamasını tahmin etmekte klasik doğrusal regresyon tahmin edicisinden ve birçok araştırmacı tarafından türetilen parametrik tahmin edicilerden daha etkindir. Kaur'un regresyon tahmin edicisinin hata kareler ortalaması klasik doğrusal regresyon tahmin edicisinin hata kareler ortalamasının alt sınırındadır (Menendez,Reyes,1998).

Kaur (1985)'un yığın ortalamasını tahmin etmek için önerdiği regresyon tipi tahmin edici

$$\hat{\bar{Y}}_{KAUR} = \kappa\bar{y} + b\left(\bar{X} - \bar{x}\right) \tag{2.35}$$

biçimindedir. Burada b örnek regresyon katsayısı κ ise uygun olarak seçilen bir sabit sayıdır. Kaur yaklaşık yanlılık ve hata kareler ortalamasını da sırasıyla Denklem (2.36) ve (2.37)'deki gibi elde etmiştir.

$$\text{Yan}(\hat{\bar{Y}}_{KAUR}) = \bar{Y}\left(\kappa - 1\right) \tag{2.36}$$

$$HKO(\hat{\bar{Y}}_{KAUR}) = \nu\left[\kappa^2\bar{Y}^2C_y^2 + \beta^2\bar{X}^2C_x^2 - 2\kappa\beta\rho_{yx}\bar{X}\bar{Y}C_xC_y\right] \tag{2.37}$$

Birinci dereceden yaklaşık olarak, $\hat{\bar{Y}}_{REG}$ klasik doğrusal regresyon tahmin edicisinin hata kareler ortalaması,

$$HKO(\hat{\bar{Y}}_{REG}) = \nu\left[\bar{Y}^2C_y^2 + \beta^2\bar{X}^2C_x^2 - 2\beta\rho_{yx}\bar{X}\bar{Y}C_xC_y\right] \tag{2.38}$$

biçimindedir (Sukhatme v.d.,1970). Kaur'un önerdiği regresyon tipi tahmin edici ile klasik doğrusal regresyon tahmin edicisi karşılaştırıldığında;

$$\begin{gathered} HKO(\hat{\bar{Y}}_{KAUR}) < HKO(\hat{\bar{Y}}_{REG}) \\ \nu\left[\kappa^2\bar{Y}^2C_y^2+\beta^2\bar{X}^2C_x^2-2\kappa\beta\rho_{yx}\bar{X}\bar{Y}C_xC_y\right] < \nu\left[\bar{Y}^2C_y^2+\beta^2\bar{X}^2C_x^2-2\beta\rho_{yx}\bar{X}\bar{Y}C_xC_y\right] \\ (\kappa^2-1) < 2\rho_{yx}{}^2(\kappa-1) \\ 2\rho_{yx}{}^2-1<\kappa<1 \end{gathered} \tag{2.39}$$

eşitsizliğini sağlayacak κ değeri için Kaur'un tahmin edicisi daha etkindir. $\text{HKO}(\hat{\bar{Y}}_{\text{KAUR}})$ en küçük değerini, $\kappa=\rho_{yx}{}^2$ olduğu durumda alır.

$$HKO(\hat{\bar{Y}}_{KAUR}) = HKO(\hat{\bar{Y}}_{REG}).\kappa \tag{2.40}$$

Menendez ve Reyes (1998), Kaur tahmin edicisinde Taylor doğrusallaştırma tekniğini kullanarak hata kareler ortalamasını yeniden hesaplamışlardır. Hesaplamalarında Kaur tahmin edicisinin hata kareler ortalamasını yeniden Denklem (2.41)'de gösterildiği gibi tanımlamışlardır.

$$HKO(\hat{\bar{Y}}^*_{KAUR}) = \nu\left[\kappa^2\bar{Y}^2C_y^2+\beta^2\bar{X}^2C_x^2-2\kappa\beta\rho_{yx}\bar{X}\bar{Y}C_xC_y\right]+\bar{Y}^2\left(\kappa-1\right)^2 \tag{2.41}$$

Denklem (2.38) ve Denklem (2.41)'den $HKO(\hat{\bar{Y}}_{KAUR}) < HKO(\hat{\bar{Y}}_{REG})$ koşulunun sağlanması için Denklem (2.39) sağlanmalıdır; ancak bu durumda

$$\kappa = \frac{1+\nu\rho_{yx}{}^2C_y{}^2}{1+\nu C_y{}^2} \tag{2.42}$$

şeklinde hesaplanırsa Kaur tahmin edicisi en küçük değerini alır ve klasik doğrusal regresyon tahmin edicisinden daha etkin sonuç verir (Menendez,Reyes,1998).

ÜÇÜNCÜ BÖLÜM

TABAKALI ÖRNEKLEME YÖNTEMİNDE TAHMİN

Geniş anlamıyla tabakalı örnekleme şu aşamalardan oluşmaktadır (Kish,1965:75).

a. Bütün yığındaki örnekleme birimleri ayrı alt yığınlara bölünür, bu alt yığınlara tabaka adı verilmektedir.
b. Her tabaka içerisinde tabakayı oluşturan örnekleme birimlerinden ayrı ayrı örnekler seçilir.
c. Her tabakada elde edilen örnekten, tabakaya ilişkin ortalama (veya başka bir istatistik) hesaplanır. Bu tabakalardan elde edilen ortalamalar uygun bir biçimde ağırlıklandırılarak tüm yığın için bileşik tahmin elde edilir.
d. Varyanslar da her tabaka için hesaplanır ve aynı şekilde ağırlıklandırılarak yığın için bileşik tahminin parçalarını oluştururlar.

Basit rasgele örnekleme yerine tabakalı rasgele örnekleme yönteminin tercih edilmesinin nedenleri üç madde ile özetlenebilir (Scheaffer, Mendenhall, Ott, 1990:98-99).

a. Aynı örnek hacmi için tabakalı rasgele örneklemenin tahmininin hata sınırı basit rasgele örneklemeninkinden daha küçük olabilir. Bu sonuç özellikle tabaka içi homojenlik sağlandığında doğrudur.
b. Yığın elemanları kullanışlı gruplanarak tabakalanırsa, araştırmada gözlem başı maliyet düşürülebilir.
c. Yığın parametrelerinin tahminleri yığının alt gruplarında istenebilir. Bu alt gruplar tabakalar sayesinde belirlenebilir.

N büyüklüğündeki bir yığın için h=(1,2,…….,H) tabaka sayısı, N_h tabaka büyüklükleri, n_h tabakadan seçilen örnek sayısı, Y ilgilendiğimiz değişken, $X=(X_1,X_2,X_3,.....,X_p)$ ise p tane değişken içeren yardımcı değişkenler vektörü olmak üzere; bu bölümde tabakalı örnekleme yöntemi kullanılarak yığın ortalamasını tahmin eden bazı tahmin ediciler incelenecektir.

3.1 KLASİK ORTALAMA TAHMİNİ

Tabakalı rasgele örnekleme yönteminde klasik tahmin, tabakalardaki ortalama tahminlerinin ağırlıklı etkilerini yığın tahminine aktarmasıyla oluşur. Bu nedenle ağırlıklandırılmış ortalama ifadesi de kullanılır.

$$\hat{\bar{Y}}_{TABAKA} = \sum_{h=1}^{H} W_h \bar{y}_h \tag{3.1}$$

Tabakalı rasgele örneklemede klasik tahmin Denklem (3.1)'deki gibi yapılır. Burada $W_h = N_h / N$ olmak üzere $\sum_{h=1}^{H} W_h = 1$ olmalıdır. Tabakalı rasgele örneklemede klasik ortalama tahminini Horvitz-Thompson tahmin edicisi kullanarak da elde edebiliriz. Denklem (2.2) tabakalı rasgele örnekleme için

$$\hat{\bar{Y}}_{TABAKA} = \frac{1}{N} \sum_{i=1}^{n} \frac{y_i}{\pi_i} \tag{3.2}$$

biçiminde ifade edilebilir. Tabakalı rasgele örneklemede basit rasgele örneklemeden farklı olarak π_i kapsanma olasılığı sabit olmayıp tabakalara göre değişiklik göstermektedir. Denklem (3.2)'de gösterilen HTE kullanılarak da klasik ortalama tahmini elde edebiliriz.

Klasik ortalama tahmini yansız bir tahmin edici olup varyansı ise Denklem (3.3)'de gibi hesaplanır.

$$V(\hat{\bar{Y}}_{TABAKA}) = \frac{1}{N^2}\sum_{h=1}^{H} N_h^2 V(\bar{y}_h) \tag{3.3}$$

$$V(\hat{\bar{Y}}_{TABAKA}) = \frac{1}{N^2}\sum_{h=1}^{H} N_h^2 \nu_h S_{y_h}^2$$

Burada $S_{y_h}^2$, h tabakasında Y değişkeni için hesaplanan varyansı ifade etmektedir.

3.2 ORAN TAHMİN EDİCİLERİ

Basit rasgele örneklemede olduğu gibi tabakalı rasgele örneklemede de tek bir X yardımcı değişkenini kullanarak oran tahmin edicileri elde etmek mümkündür. Tabakalı rasgele örneklemede basit rasgele örneklemeden farklı olarak $\frac{\bar{y}}{\bar{x}}$ oranı kullanılarak bileşik oran tahmini veya $\frac{\bar{y}_h}{\bar{x}_h}$ oranlarının kullanımı tek tek tabakalarda yapılarak ayrı oran tahmini elde edilebilir.

3.2.1 Bileşik Oran Tahmin Edicisi

X ve Y değişkeni arasındaki ilişki pozitif olduğunda Hansen, Hurwitz ve Gurney (1946) tarafından önerilen bileşik oransal tahmin edici,

$$\hat{\bar{Y}}_{ORANBİL} = \left(\frac{\hat{\bar{Y}}_{TABAKA}}{\hat{\bar{X}}_{TABAKA}}\right)\bar{X} \tag{3.4}$$

şeklindedir. Burada $\hat{\bar{X}}_{TABAKA}$ tabakalı rasgele örneklemede X değişkeni için klasik ortalama tahmin edicisidir. Bileşik oransal tahmin edicinin yanı ve hata kareler ortalaması ise sırasıyla Denklem (3.5) ve Denklem (3.6)'da verilmiştir.

$$Yan(\hat{\bar{Y}}_{ORANBİL}) \cong \frac{1}{\bar{X}} \sum_{h=1}^{H} W_h^2 v_h (RS_{xh}^2 - S_{xyh}) \tag{3.5}$$

$$HKO(\hat{\bar{Y}}_{ORANBİL}) \cong \sum_{h=1}^{H} W_h^2 v_h (S_{yh}^2 + R^2 S_{xh}^2 - 2RS_{xyh}) \tag{3.6}$$

R basit rasgele örneklemede olduğu gibi $\frac{\bar{Y}}{\bar{X}}$ yığın ortalamaları oranını göstermektedir. v_h ise h. tabaka için düzeltme terimini ifade etmektedir.

Bileşik oran tahmin edicisininin klasik ortalama tahmininden daha etkin sonuç verebilmesi için $HKO(\hat{\bar{Y}}_{ORANBİL}) < V(\hat{\bar{Y}}_{TABAKA})$ eşitsizliğinin sağlanması gereklidir. Bu eşitsizliği sağlayan koşul Denklem (3.7)'deki gibi elde edilebilir.

$$\begin{gathered} HKO(\hat{\bar{Y}}_{ORANBİL}) < V(\hat{\bar{Y}}_{TABAKA}) \\ \sum_{h=1}^{H} W_h^2 v_h (S_{yh}^2 + R^2 S_{xh}^2 - 2RS_{xyh}) < \sum_{h=1}^{H} W_h^2 v_h S_{yh}^2 \\ \sum_{h=1}^{H} W_h^2 v_h \ (R^2 S_{xh}^2 - 2RS_{xyh}) < 1 \end{gathered} \tag{3.7}$$

3.2.2 Ayrı Oran Tahmin Edicisi

Ayrı oran tahmini X ve Y değişkenleri arasında pozitif korelasyon varken oran tahminlerinin her tabakada ayrı ayrı yapılıp bu tahminlerin ağırlıklandırılmasıyla elde edilir.

$$\hat{\bar{Y}}_{ORANAYRI} = \sum_{h=1}^{H} W_h \frac{\bar{y}_h}{\bar{x}_h} \bar{X}_h \tag{3.8}$$

Ayrı oran tahmin edicisi bir tip grup oran modelidir. Grup oran modeli, bilinen pozitif bir yardımcı değişken X elde edilebilir olduğunda ve y_i/x_i oranı aynı gruplardaki birimler için yaklaşık olarak sabit bir sayı verdiğinde kullanılabilir. Bu oran farklı gruplardaki birimler için farklılık gösterebilir. Eğer gruplar örnekleme tasarımı hazırlanmadan önce belirlenebilirse tabaka olarak işlev görebilirler. Aksi takdirde gruplar sonradan tabakalanır ki bu grupların örnekleme tasarımı sonrasında belirlenmesidir. Bu durumda ayrı oran tahmin edicisi sonradan tabakalama tahmin edicisi olarak da adlandırılır (Särndal, Swensson, Wretman, 1992:270).

Ayrı oran tahmin edicisinin yanı ve hata kareler ortalaması ise sırasıyla Denklem (3.9) ve Denklem (3.10)'daki biçimde hesaplanabilir (Koyuncu,2007) .

$$\text{Yan}(\hat{\bar{Y}}_{ORANAYRI}) \cong \sum_{h=1}^{H} W_h \bar{Y}_h v_h (C_{xh}^2 - C_{xyh}) \tag{3.9}$$

$$HKO(\hat{\bar{Y}}_{ORANAYRI}) \cong \sum_{h=1}^{H} W_h^2 v_h \bar{Y}_h^2 (C_{yh}^2 + C_{xh}^2 - 2C_{xyh}) \tag{3.10}$$

Ayrı oran tahmin edicisininin klasik ortalama tahmininden daha etkin sonuç verebilmesi için $HKO(\hat{\bar{Y}}_{ORANAYRI}) < V(\hat{\bar{Y}}_{TABAKA})$ eşitsizliğinin sağlanması gereklidir. Bu eşitsizliği sağlayan sadeleştirilmiş bir koşul elde edilememiştir.

3.2.3 Sisodia-Dwivedi Oran Tahmin Edicisi

Sisodia ve Dwivedi (1981) basit rasgele örnekleme yönteminde, yığın ortalamasını tahmin etmek için yardımcı değişken X'in değişim katsayısı C_x'i kullanan değiştirilmiş bir oran tahmin edicisi önermişlerdir.

$$\hat{\bar{Y}}_{SD} = \bar{y}\,\frac{\bar{X} + C_x}{\bar{x} + C_x} \tag{3.11}$$

Kadılar ve Çıngı (2003), bu tahmin ediciyi tabakalı rasgele örneklemede yığın ortalamasını tahmin etmek için Denklem (3.12)'deki gibi tanımlamışlardır.

$$\hat{\bar{Y}}_{SDTAB} = \hat{\bar{Y}}_{TABAKA}\,\frac{\sum_{h=1}^{H} W_h(\bar{X}_h + C_{xh})}{\sum_{h=1}^{H} W_h(\bar{x}_h + C_{xh})} \tag{3.12}$$

$x_{SD} = \sum_{h=1}^{H} W_h(\bar{x}_h + C_{xh})$ ve $X_{SD} = \sum_{h=1}^{H} W_h(\bar{X}_h + C_{xh})$ biçiminde tanımlanırsa Sisodia-Dwivedi oran tahmin edicisini Denklem (3.13)'deki gibi belirtebiliriz.

$$\hat{\bar{Y}}_{SDTAB} = \frac{\hat{\bar{Y}}_{TABAKA}}{x_{SD}} X_{SD} = \hat{R}_{SD} X_{SD} \tag{3.13}$$

Sisodia-Dwivedi oran tahmin edicisi ile bileşik oran tahmin edicisi karşılaştırıldığında tek farkın $\hat{R}_{SD}$ terimi olduğu görülmektedir. Bu yüzden

Sisodia-Dwivedi oran tahmin edicisinin yanı ve hata kareler ortalaması da bileşik oran tahmin edicisindeki gibi ifade edilebilir.

$$Yan(\hat{\bar{Y}}_{SDTAB}) \cong \frac{1}{X_{SD}} \sum_{h=1}^{H} W_h^2 \nu_h (R_{SD} S_{xh}^2 - S_{xyh}) \qquad (3.14)$$

$$HKO(\hat{\bar{Y}}_{SDTAB}) \cong \sum_{h=1}^{H} W_h^2 \nu_h (S_{yh}^2 + R_{SD}^2 S_{xh}^2 - 2R_{SD} S_{xyh}) \qquad (3.15)$$

Denklem (3.14) ve (3.15)'te yer alan $R_{SD} = \dfrac{\hat{\bar{Y}}_{TABAKA}}{X_{SD}}$ biçiminde hesaplanır. Sisodia-Dwivedi oran tahmin edicisi bileşik oran tahmin edicisi ile karşılaştırılırsa;

$$\begin{gathered} HKO(\hat{\bar{Y}}_{SDTAB}) < HKO(\hat{\bar{Y}}_{ORANBİL}) \\ \sum_{h=1}^{H} W_h^2 \nu_h (S_{yh}^2 + R_{SD}^2 S_{xh}^2 - 2R_{SD} S_{xyh}) < \sum_{h=1}^{H} W_h^2 \nu_h (S_{yh}^2 + R^2 S_{xh}^2 - 2R S_{xyh}) \\ R_{SD}^2 \sum_{h=1}^{H} W_h^2 \nu_h S_{xh}^2 - 2R_{SD} \sum_{h=1}^{H} W_h^2 \nu_h S_{xyh} < R^2 \sum_{h=1}^{H} W_h^2 \nu_h S_{xh}^2 - 2R \sum_{h=1}^{H} W_h^2 \nu_h S_{xyh} \end{gathered} \qquad (3.16)$$

$L = \sum_{h=1}^{H} W_h^2 \nu_h S_{xyh}$ ve $M = \sum_{h=1}^{H} W_h^2 \nu_h S_{xh}^2$ olmak üzere Denklem (3.16)

$$\begin{aligned} R_{SD}^2 M - 2R_{SD} L &< R^2 M - 2RL \\ 2RL - 2R_{SD} L &< R^2 M - R_{SD}^2 M \\ 2L(R - R_{SD}) &< M(R^2 - R_{SD}^2) \\ 2L(R - R_{SD}) &< M(R - R_{SD})(R + R_{SD}) \end{aligned} \qquad (3.17)$$

biçiminde ifade edilebilir. Bu durumda $(R - R_{SD})(R + R_{SD}) > 0$ ise

$$M > \frac{2L}{(R + R_{SD})}$$

$(R - R_{SD})(R + R_{SD}) < 0$ ise

$M < \dfrac{2L}{(R + R_{SD})}$ koşulları sağlandığında Sisodia-Dwivedi oran tahmin edicisi bileşik oran tahmin edicisinden daha etkin sonuç verir.

3.2.4 Singh-Kakran Oran Tahmin Edicisi

Sisodia-Dwivedi tahmin edicisinden yola çıkarak Singh ve Kakran (1993) yardımcı değişken X'in yığın basıklık katsayısını kullanarak ilgilenilen değişken Y'nin yığın ortalaması için oran-tipi bir tahmin edici türetmişlerdir (Upadhyaya ve Singh, 1999).

$$\hat{\bar{Y}}_{SK} = \bar{y}\,\frac{\bar{X} + G_2(x)}{\bar{x} + G_2(x)} \tag{3.18}$$

Burada $G_2(x)$ yığın basıklık katsayısını temsil etmektedir. Kadılar ve Çıngı (2003) bu tahmin ediciyi tabakalı rasgele örneklemede yığın ortalamasını tahmin etmek için Denklem (3.19)'daki gibi tanımlamışlardır.

$$\hat{\bar{Y}}_{SKTAB} = \hat{\bar{Y}}_{TABAKA}\,\frac{\sum_{h=1}^{H} W_h(\bar{X}_h + G_{2h}(x))}{\sum_{h=1}^{H} W_h(\bar{x}_h + G_{2h}(x))} \tag{3.19}$$

$x_{SK} = \sum_{h=1}^{H} W_h(\overline{x}_h + G_{2h}(x))$ ve $X_{SK} = \sum_{h=1}^{H} W_h(\overline{X}_h + G_{2h}(x))$ biçiminde tanımlanırsa Singh-Kakran oran tahmin edicisini Denklem (3.20)'deki gibi belirtebiliriz.

$$\hat{\overline{Y}}_{SKTAB} = \frac{\hat{\overline{Y}}_{TABAKA}}{x_{SK}} X_{SK} = \hat{R}_{SK} X_{SK} \tag{3.20}$$

Singh-Kakran oran tahmin edicisi ile bileşik oran tahmin edicisi karşılaştırıldığında tek fark $\hat{R}_{SK}$ terimidir. Singh-Kakran oran tahmin edicisinin yanı ve hata kareler ortalaması da bileşik oran tahmin edicisindeki gibi ifade edilebilir.

$$Yan(\hat{\overline{Y}}_{SKTAB}) \cong \frac{1}{X_{SK}} \sum_{h=1}^{H} W_h^2 \nu_h (R_{SK} S_{xh}^2 - S_{xyh}) \tag{3.21}$$

$$HKO(\hat{\overline{Y}}_{SKTAB}) \cong \sum_{h=1}^{H} W_h^2 \nu_h (S_{yh}^2 + R_{SK}^2 S_{xh}^2 - 2R_{SK} S_{xyh}) \tag{3.22}$$

$R_{SK} = \dfrac{\hat{\overline{Y}}_{TABAKA}}{\sum_{h=1}^{H} W_h(\overline{X}_h + G_{2h}(x))}$ biçiminde hesaplanır. Singh-Kakran oran tahmin edicisi bileşik oran tahmin edicisi ile karşılaştırılırsa;

$$\begin{gathered} HKO(\hat{\overline{Y}}_{SKTAB}) < HKO(\hat{\overline{Y}}_{ORANBİL}) \\ \sum_{h=1}^{H} W_h^2 \nu_h (S_{yh}^2 + R_{SK}^2 S_{xh}^2 - 2R_{SK} S_{xyh}) < \sum_{h=1}^{H} W_h^2 \nu_h (S_{yh}^2 + R^2 S_{xh}^2 - 2R S_{xyh}) \\ R_{SK}^2 \sum_{h=1}^{H} W_h^2 \nu_h S_{xh}^2 - 2R_{SK} \sum_{h=1}^{H} W_h^2 \nu_h S_{xyh} < R^2 \sum_{h=1}^{H} W_h^2 \nu_h S_{xh}^2 - 2R \sum_{h=1}^{H} W_h^2 \nu_h S_{xyh} \end{gathered} \tag{3.23}$$

$L = \sum_{h=1}^{H} W_h^2 \nu_h S_{xyh}$ ve $M = \sum_{h=1}^{H} W_h^2 \nu_h S_{xh}^2$ olmak üzere Denklem (3.23)

$$\begin{aligned} R_{SK}^2 M - 2R_{SK} L &< R^2 M - 2RL \\ 2RL - 2R_{SK} L &< R^2 M - R_{SK}^2 M \\ 2L(R - R_{SK}) &< M(R^2 - R_{SK}^2) \\ 2L(R - R_{SK}) &< M(R - R_{SK})(R + R_{SK}) \end{aligned} \tag{3.24}$$

biçiminde ifade edilebilir. Bu durumda $(R - R_{SK})(R + R_{SK}) > 0$ ise

$$M > \frac{2L}{(R + R_{SK})}$$

$(R - R_{SK})(R + R_{SK}) < 0$ ise

$M < \dfrac{2L}{(R + R_{SK})}$ koşulları sağlandığında Singh-Kakran oran tahmin edicisi bileşik oran tahmin edicisinden daha etkin sonuç verir.

3.2.5 Upadhyaya-Singh Oran Tahmin Edicileri

Upadhyaya ve Singh (1999) yardımcı değişken X'in hem yığın basıklık katsayısını hem de değişim katsayısını kullanarak ilgilenilen değişken Y'nin yığın ortalaması için oran-tipi bir tahmin edici türetmişlerdir.

$$\hat{\bar{Y}}_{US1} = \bar{y} \frac{\bar{X} G_2(x) + C_x}{\bar{x} G_2(x) + C_x} \tag{3.25}$$

Kadılar ve Çıngı (2003) bu tahmin ediciyi tabakalı rasgele örneklemede yığın ortalamasını tahmin etmek için Denklem (3.26)'daki gibi tanımlamışlardır.

$$\hat{\bar{Y}}_{US1TAB} = \hat{\bar{Y}}_{TABAKA} \frac{\sum_{h=1}^{H} W_h (\bar{X}_h G_{2h}(x) + C_{xh})}{\sum_{h=1}^{H} W_h (\bar{x}_h G_{2h}(x) + C_{xh})} \tag{3.26}$$

$$x_{US1} = \sum_{h=1}^{H} W_h (\bar{x}_h G_{2h}(x) + C_{xh}) \quad \text{ve} \quad X_{US1} = \sum_{h=1}^{H} W_h (\bar{X}_h G_{2h}(x) + C_{xh})$$

biçiminde tanımlanırsa $\hat{\bar{Y}}_{US1TAB}$ oran tahmin edicisini Denklem (3.27)'deki gibi belirtebiliriz.

$$\hat{\bar{Y}}_{US1TAB} = \frac{\hat{\bar{Y}}_{TABAKA}}{x_{US1}} X_{US1} = \hat{R}_{US1} X_{US1} \tag{3.27}$$

$\hat{\bar{Y}}_{US1TAB}$ oran tahmin edicisinin yanı ve hata kareler ortalaması da bileşik oran tahmin edicisindeki gibi ifade edilebilir.

$$Yan(\hat{\bar{Y}}_{US1TAB}) \cong \frac{1}{X_{US1}} \sum_{h=1}^{H} W_h^2 v_h (R_{US1} S_{xh}^2 - S_{xyh}) \tag{3.28}$$

$$HKO(\hat{\bar{Y}}_{US1TAB}) \cong \sum_{h=1}^{H} W_h^2 v_h (S_{yh}^2 + R_{US1}^2 S_{xh}^2 - 2R_{US1} S_{xyh}) \tag{3.29}$$

$R_{US1} = \dfrac{\hat{\bar{Y}}_{TABAKA}}{\sum_{h=1}^{H} W_h (\bar{X}_h G_{2h}(x) + C_{xh})}$ biçiminde hesaplanır. Upadhyaya-Singh oran tahmin edicisi bileşik oran tahmin edicisi ile karşılaştırılırsa;

$$HKO(\hat{\bar{Y}}_{US1TAB}) < HKO(\hat{\bar{Y}}_{ORANBİL})$$

$$\sum_{h=1}^{H} W_h^2 \nu_h (S_{yh}^2 + R_{US1}^2 S_{xh}^2 - 2R_{US1} S_{xyh}) < \sum_{h=1}^{H} W_h^2 \nu_h (S_{yh}^2 + R^2 S_{xh}^2 - 2RS_{xyh}) \tag{3.30}$$

$$R_{US1}^2 \sum_{h=1}^{H} W_h^2 \nu_h S_{xh}^2 - 2R_{US1} \sum_{h=1}^{H} W_h^2 \nu_h S_{xyh} < R^2 \sum_{h=1}^{H} W_h^2 \nu_h S_{xh}^2 - 2R \sum_{h=1}^{H} W_h^2 \nu_h S_{xyh}$$

$L = \sum_{h=1}^{H} W_h^2 \nu_h S_{xyh}$ ve $M = \sum_{h=1}^{H} W_h^2 \nu_h S_{xh}^2$ olmak üzere Denklem (3.30)

$$\begin{aligned} R_{US1}^2 M - 2R_{US1} L &< R^2 M - 2RL \\ 2RL - 2R_{US1} L &< R^2 M - R_{US1}^2 M \\ 2L(R - R_{US1}) &< M(R^2 - R_{US1}^2) \\ 2L(R - R_{US1}) &< M(R - R_{US1})(R + R_{US1}) \end{aligned} \tag{3.31}$$

biçiminde ifade edilebilir. Bu durumda $(R - R_{US1})(R + R_{US1}) > 0$ ise

$$M > \frac{2L}{(R + R_{US1})}$$

$(R - R_{US1})(R + R_{US1}) < 0$ ise

$M < \dfrac{2L}{(R + R_{US1})}$ koşulları sağlandığında Upadhyaya-Singh oran tahmin edicisi bileşik oran tahmin edicisinden daha etkin sonuç verir.

Upadhyaya ve Singh (1999) basıklık katsayısı ve değişim katsayısının yerlerini değiştirerek Y'nin yığın ortalamasını tahmin eden başka bir tahmin edici daha önermişlerdir.

$$\hat{\bar{Y}}_{US2} = \bar{y} \frac{\bar{X} C_x + G_2(x)}{\bar{x} C_x + G_2(x)} \tag{3.32}$$

Kadılar ve Çıngı (2003) bu tahmin ediciyi tabakalı rasgele örneklemede yığın ortalamasını tahmin etmek için Denklem (3.33)'daki gibi tanımlamışlardır.

$$\hat{\bar{Y}}_{US2TAB} = \hat{\bar{Y}}_{TABAKA} \frac{\sum_{h=1}^{H} W_h(\bar{X}_h C_{xh} + G_{2h}(x))}{\sum_{h=1}^{H} W_h(\bar{x}_h C_{xh} + G_{2h}(x))} \tag{3.33}$$

$$x_{US2} = \sum_{h=1}^{H} W_h(\bar{x}_h C_{xh} + G_{2h}(x)) \text{ ve } X_{US1} = \sum_{h=1}^{H} W_h(\bar{X}_h C_{xh} + G_{2h}(x))$$

biçiminde tanımlanırsa $\hat{\bar{Y}}_{US2TAB}$ oran tahmin edicisini Denklem (3.34)'deki gibi belirtebiliriz.

$$\hat{\bar{Y}}_{US2TAB} = \frac{\hat{\bar{Y}}_{TABAKA}}{x_{US2}} X_{US2} = \hat{R}_{US2} X_{US2} \tag{3.34}$$

$\hat{\bar{Y}}_{US2TAB}$ oran tahmin edicisinin yanı ve hata kareler ortalaması R_{US1} yerine $R_{US2} = \dfrac{\hat{\bar{Y}}_{TABAKA}}{\sum_{h=1}^{H} W_h(\bar{X}_h C_{xh} + G_{2h}(x))}$ değeri gelmek üzere $\hat{\bar{Y}}_{US1TAB}$ oran tahmin edicisininkilerle aynıdır. $\hat{\bar{Y}}_{US2TAB}$ oran tahmin edicisinin bileşik oran tahmin edicisi ile karşılaştırılmasında da $HKO(\hat{\bar{Y}}_{US2TAB}) < HKO(\hat{\bar{Y}}_{ORANBİL})$ eşitsizliğinin sağlanması için R_{US1} yerine R_{US2} kullanılmak üzere aynı koşul geçerlidir.

3.3 REGRESYON TAHMİN EDİCİLERİ

Basit rasgele örneklemede olduğu gibi tabakalı rasgele örneklemede de X yardımcı değişkeni ile Y ilgilenilen değişken arasındaki ilişkiyi gösteren doğru orijinden geçmiyorsa oran tahmin edicisi yerine regresyon tahmin edicisi kullanılır. Değişkenler arasındaki ilişki yığın ortalamaları ve yığın ortalama tahminleri bazında incelenerek regresyon tahmini yapılırsa bileşik regresyon tahmini elde edilir. Tabaka bazında ortalamalar ve ortalama tahminleri arasındaki ilişki incelenerek buna ilişkin regresyon tahmini yapılırsa ayrı regresyon tahmini elde edilir.

3.3.1 Bileşik Regresyon Tahmin Edicisi

Tabakalı rasgele örnekleme yönteminde bileşik regresyon tahmin edicisi Denklem (3.35)'de verilmiştir.

$$\hat{\bar{Y}}_{REGBİL} = \hat{\bar{Y}}_{TABAKA} + b_c(\bar{X} - \hat{\bar{X}}_{TABAKA}) \tag{3.35}$$

Burada $b_c = \dfrac{\sum_{h=1}^{H} W_h^2 v_h s_{yxh}}{\sum_{h=1}^{H} W_h^2 v_h s_{xh}^2}$ olmak üzere β_c yığın regresyon parametresinin tahminidir.

$$\beta_c = \frac{\sum_{h=1}^{H} W_h^2 v_h S_{yxh}}{\sum_{h=1}^{H} W_h^2 v_h S_{xh}^2}, \mu_{rsh} = \frac{1}{N_h - 1}\sum_{j=1}^{N_h} (y_{hj} - \bar{Y}_h)^r (x_{hj} - \bar{X}_h)^s ,$$

$$\alpha_h = \frac{N_h^2 W_h v_h}{(N_h - 1)(N_h - 2)\sum_{h=1}^{H} W_h^2 v_h S_{xh}^2}, \alpha_h^* = \frac{N_h^2 W_h v_h}{(N_h - 1)(N_h - 2)\sum_{h=1}^{H} W_h^2 v_h S_{xyh}^2}$$

olmak üzere bileşik regresyon tahmin edicisinin yanı

$$Yan(\hat{\bar{Y}}_{REGBİL}) \cong -\beta_c \sum_{h=1}^{H} W_h^2 \nu_h (\alpha_h^* \mu_{12h} - \alpha_h \mu_{03h}) \tag{3.36}$$

eşitliği ile hesaplanır (Sukhatme ve Sukhatme, 1984). Bileşik regresyon tahmin edicisi $b_c = \beta_c$ gibi bir sabit alındığında yansız olur. Bileşik regresyon tahmin edicisinin hata kareler ortalaması ise

$$HKO(\hat{\bar{Y}}_{REGBİL}) \cong \sum_{h=1}^{H} W_h^2 \nu_h S_{yh}^2 (1 - \rho_c^2) \tag{3.37}$$

eşitliği ile hesaplanır. Burada

$$\rho_c = \frac{\sum_{h=1}^{H} W_h^2 \nu_h S_{yxh}}{\left(\sqrt{\sum_{h=1}^{H} W_h^2 \nu_h S_{yh}^2}\right)\left(\sqrt{\sum_{h=1}^{H} W_h^2 \nu_h S_{xh}^2}\right)} \tag{3.38}$$

tabakalı örneklemede tüm tabakalar üzerinden korelasyon katsayısı olmaktadır (Koyuncu,2007).

Bileşik regresyon tahmin edicisi ile tabakalı rasgele örnekleme için klasik tahmin ediciyi karşılaştırdığımızda $\rho_c \neq 0$ koşulunu sağlayan tüm X ve Y değişkenleri için bileşik regresyon tahmin edicisi tabakalı rasgele örnekleme için klasik tahmin ediciden daha etkindir.

$$HKO(\hat{\bar{Y}}_{REGBİL}) < V(\hat{\bar{Y}}_{TABAKA})$$
$$\sum_{h=1}^{H} W_h^2 v_h S_{yh}^2 (1-\rho_c^2) < \sum_{h=1}^{H} W_h^2 v_h S_{yh}^2 \tag{3.39}$$
$$(1-\rho_c^2) < 1$$

3.3.2 Ayrı Regresyon Tahmin Edicisi

Basit doğrusal regresyon modeli, ayrı ayrı her tabakaya uygulanarak elde edilen tahminler ağırlıklandırılırsa, bu durumda tahmin ayrı regresyon tahmini olarak adlandırılır.

$$\hat{\bar{Y}}_{REGAYRI} = \sum_{h=1}^{H} W_h \left[\bar{y}_h + b_h (\bar{X}_h - \bar{x}_h) \right] \tag{3.40}$$

$b_h = \dfrac{s_{xyh}}{s_{xh}^2}$ olmak üzere β_h yığın regresyon parametresinin tahminidir.

$\lambda_{rsh} = \dfrac{\mu_{rsh}}{\mu_{20h}^{r/2}\mu_{02h}^{s/2}}$ olmak üzere ayrı regresyon tahmin edicisinin yanı,

$$Yan(\hat{\bar{Y}}_{REGAYRI}) \cong \sum_{h=1}^{H} W_h \beta_h \bar{X}_h v_h C_{xh} \left(\lambda_{03h} - \frac{\lambda_{12h}}{\rho_{xyh}} \right) \tag{3.41}$$

eşitliği ile bulunur. Ayrı regresyon tahmin edicisi $b_h = \beta_h$ gibi bir sabit alındığında yansız olur. Ayrı regresyon tahmin edicisinin hata kareler ortalaması

$$HKO(\hat{\bar{Y}}_{REGAYRI}) \cong \sum_{h=1}^{H} W_h^2 v_h \left[S_{yh}^2 + \beta_h^2 S_{xh}^2 - 2\beta_h S_{xyh} \right] \qquad (3.42)$$

eşitliği ile hesaplanır.

Ayrı regresyon tahmin edicisininin klasik ortalama tahmininden daha etkin sonuç verebilmesi için $HKO(\hat{\bar{Y}}_{REGAYRI}) < V(\hat{\bar{Y}}_{TABAKA})$ eşitsizliğinin sağlanması gereklidir. Bu eşitsizliği sağlayan sadeleştirilmiş bir koşul elde edilememiştir.

DÖRDÜNCÜ BÖLÜM

GENELLEŞTİRİLMİŞ REGRESYON (GREG) TAHMİN EDİCİSİ

İkinci ve üçüncü bölümde N büyüklüğündeki bir yığın için Y ilgilendiğimiz değişken, $X=(X_1,X_2,X_3,.....,X_p)$ ise p tane değişken içeren yardımcı değişkenler vektörü olarak tanımlanmış; Srivastava İki Değişkenli Oran Tahmin Edicisi dışında X vektöründen sadece bir yardımcı değişken kullanan tahmin ediciler açıklanmıştı. Bu bölümde çoklu regresyon denkleminden yararlanılarak oluşturulan “Genelleştirilmiş Regresyon (GREG) Tahmin Edicisi” anlatılacaktır.

Öncelikle basit rasgele örnekleme yöntemi için tek değişkenli klasik doğrusal regresyon tahmin edicisini hatırlarsak $\hat{\bar{Y}}_{REG} = \bar{y} + b\left(\bar{X} - \bar{x}\right)$’dir. Eğer regresyon modelimiz $X=(X_1,X_2)$ şeklinde iki yardımcı değişkenden oluşursa bu durumda regresyon tahmin edicisi,

$$\hat{\bar{Y}}_{REG} = \bar{y} + b_1\left(\bar{X}_1 - \bar{x}_1\right) + b_2\left(\bar{X}_2 - \bar{x}_2\right) \tag{4.1}$$

biçiminde olacaktır. Denklem (4.1)’de b_1, b_2 yığın regresyon katsayıları β_1, β_2’nin tahminleridir. $x_i = \left(1, x_{1i}, x_{2i}\right)'$ olmak üzere $(b_1, b_2)' = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\sum_{i=1}^{n} x_i y_i\right)$ eşitliği ile b_1, b_2 bulunur. Yardımcı değişken sayısının $X=(X_1,X_2,X_3,.....,X_p)$ p tane olması durumunda iki yardımcı değişkenli regresyon denklemini genelleştirirsek; $X_i = \left(1, X_{1i}, X_{2i}, ..., X_{pi}\right)'$ olmak üzere yığın regresyon katsayıları $\beta = (\beta_0, \beta_1, \beta_2, ..., \beta_p)'$ Denklem (4.2)’deki gibi bulunur.

$$\beta = (X'X)^{-1}X'Y = \left(\sum_{i=1}^{N} X_i X_i'\right)^{-1}\left(\sum_{i=1}^{N} X_i Y_i\right) \tag{4.2}$$

$x_i = \left(1, x_{1i}, x_{2i}, \ldots\ldots, x_{pi}\right)'$ olmak üzere $\beta = (\beta_1, \beta_2, \ldots\ldots, \beta_p)'$ katsayılarını tahmin eden örnek regresyon katsayıları ise

$$b = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\sum_{i=1}^{n} x_i y_i\right) \tag{4.3}$$

eşitliği ile bulunur. b tahmin edicisi β'nın asimptotik olarak tasarım yansız tahmin edicisidir. Bunun anlamı büyük örnekler için yanlılığın yok olacak olmasıdır. Basit rasgele örnekleme için genelleştirilmiş regresyon tahmin edicisi,

$$\hat{\bar{Y}}_{GREG} = \bar{y} + \left(\bar{X} - \bar{x}\right)'b \tag{4.4}$$

eşitliği ile tanımlanır. Burada $\bar{X} = (\bar{X}_1, \bar{X}_2, \ldots\ldots, \bar{X}_p)$, $\bar{x} = (\bar{x}_1, \bar{x}_2, \ldots\ldots\ldots, \bar{x}_p)$ sırasıyla yığın ortalamaları ve örnek ortalamaları vektörlerini göstermektedir. Genelleştirilmiş regresyon tahmin edicisi ilgilenilen değişkenin yığın ortalamasının asimptotik olarak tasarım yansız tahmin edicisidir (Bethlehem, Cobben, Schouten, 2007:98).

Genel regresyon tahmin edicisi ile elde edilen hassasiyet iki faktöre bağlıdır.

a. Tahmin ediciyi elde etmek için oluşturulan modele
b. Örnekleme tasarımına

Oluşturulan model tahmin edicinin formülünü belirler; dolayısıyla varyansı etkiler. Örnekleme tasarımı tahmin edicinin dağılımını belirler; dolayısıyla varyansı etkiler. Sonuçta gerçekleşen varyans tasarım etkisi ve model uyumu etkisine bağlı olarak oluşur.

Yaklaşık varyans ε_i'lerin bir fonksiyonudur. Küçük artıklar küçük varyansa neden olur. Basit rasgele örnekleme için varyans Denklem (4.5)'teki gibi yazılabilir.

$$V(\hat{\hat{Y}}_{GREG}) = \nu S_y^2 (1 - R^2) \tag{4.5}$$

Burada R^2 belirlilik katsayısı olup Denklem (4.6)'daki gibi de ifade edilebilir.

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \varepsilon_i^2}{\sum_{i=1}^{N} (Y_i - \bar{Y})^2} = 1 - \frac{\sum_{i=1}^{N} \varepsilon_i^2}{(N-1) S_y^2} \tag{4.6}$$

Regresyon teorisinden bilindiği gibi modele yeni bir değişken eklendiğinde R^2 artar veya değişmez. Basit rasgele örnekleme durumunda modele bir yardımcı değişken eklenerek R^2'de düzgün bir gelişme hissedilirse, bu yardımcı değişkeni regresyon tahmininde kullanmak için güçlü bir neden olur. Eğer modele katılan yardımcı değişken; ilgilenilen değişkenin tahmininin etkinliğinde önemli bir kazanç sağlamıyorsa basit model, daha karmaşık olana tercih edilir (Särndal, Swensson, Wretman, 1992:276).

Basit rasgele örnekleme için GREG tahmin edicisi ile klasik doğrusal regresyon tahmin edicisinin etkinliklerini karşılaştıracak olursak,

$$V(\hat{\bar{Y}}_{GREG}) < V(\hat{\bar{Y}}_{REG})$$
$$\nu S_y^2(1-R^2) < \nu S_y{}^2(1-\rho_{yx}{}^2) \tag{4.7}$$
$$R^2 > \rho_{yx}{}^2$$

eşitsizliğinin sağlandığı durumlarda GREG tahmin edicisi klasik doğrusal regresyondan daha etkindir.

4.1 TASARIM AĞIRLIKLARI İLE GREG TAHMİN EDİCİSİ HESAPLAMA

Denklem (4.4)'de hesaplanan GREG tahmin edicisi, genelleştirilmiş regresyon tahmininin basit rasgele örneklemede birden fazla yardımcı değişken olması durumundaki özel halidir. Basit rasgele örnekleme yöntemi dışındaki diğer örnekleme yöntemlerinde de GREG tahmin edicisini elde etmek mümkündür. Bunun için araştırma birimlerinin kapsanma olasılığı π_i değerleri kullanılarak Horvitz-Thompson tahmin edicisini elde etmemiz ve $\frac{1}{\pi_i} = d_i$ tasarım ağırlıklarını kullanmamız yeterlidir. Tasarım ağırlıkları ve Horvitz-Thompson tahmin edicisi kullanılarak ilgilenilen değişken Y'nin ortalaması için GREG tahmini

$$\hat{\bar{Y}}_{GREG} = \frac{1}{N}\left[\hat{Y}_{HT} + (\sum_{i=1}^{N} X_i - \sum_{i=1}^{n} d_i x_i)' b_d\right] \tag{4.8}$$

eşitliği ile hesaplanır. Burada $\hat{Y}_{HT} = \hat{\tau}_\pi$ Y değişkeni için Horvitz-Thompson toplam tahmini, b_d ise regresyon katsayıları vektörüdür.

$$b_d = (\sum_{i=1}^{n} d_i c_i x_i (x_i)')^{-1} (\sum_{i=1}^{n} d_i c_i x_i y_i) \tag{4.9}$$

c_i istatistikçiler tarafından belirlenen ağırlıklardır. Standart olarak tüm i'ler için $c_i = 1$ olarak seçilir.

Örnekleme yöntemi ne olursa olsun, örneğe çıkan birimler için tasarım ağırlıkları biliniyorsa GREG tahmin edicisi hesaplanabilir. GREG tahmin edicisi asimptotik olarak tasarım yansız tahmin edicidir.

Basit rasgele örnekleme yöntemi dışındaki örnekleme tasarımlarında varyans R^2'nin basit bir fonksiyonu değildir. Model uyumluluğunun etkisi ile örnekleme tasarımının etkisi arasında etkileşim vardır (Särndal, Swensson, Wretman, 1992:276).

4.2 AYARLAMA TAHMİN EDİCİSİ İLE GREG TAHMİN EDİCİSİ HESAPLAMA

Örnekleme araştırmalarında tahminin etkinliğini arttırmak için kullanılan yöntemlerden biri de ayarlama tahminidir. Ayarlama örnekleme araştırmalarında yaygın olarak kullanılan bir yöntem olmaya başlamıştır. Ayarlama etkin tahminler üretmek için yardımcı bilgiyi kullanır. Ayarlama için bir veya birden fazla yardımcı değişkene ilişkin yığın toplamlarını bilmemiz gerekmektedir. Ayarlama tahmin edicisinin etkinliği yardımcı değişkenlerin ilgilenilen değişken Y'nin değişkenliğini ne kadar açıklayabildiğine bağlıdır.

Ayarlama tahmin edicilerini oluşturmak için geleneksel olarak "uzaklık minimizasyon yaklaşımı" kullanılır. Önerilen uzaklık ölçüleri yaklaşık olarak aynı tahmin edicileri üretir. Uzaklık minimizasyonu ayarlama için tek olası başlangıç noktası değildir. Alternatif olarak fonksiyonel biçim yaklaşımı da kullanılabilir. Ayarlanmış ağırlıklar iki parametreye bağlı olarak basit matematiksel bir biçimde verilebilir ve bu fonksiyonel ayarlama tahmin ediciler sınıfını tanımlar. Bu sınıf GREG tahmin edicileri sınıfını da kapsamaktadır (Estevao, Särndal, 2000).

İlgilenilen değişken Y için ayarlanmış toplam tahmini doğrusal ağırlıklandırma yöntemi ile Denklem (4.10)'deki gibi elde edilir.

$$\hat{Y}_{AY} = \sum_{i=1}^{n} w_i y_i \tag{4.10}$$

Burada w_i araştırmadaki uygun yardımcı bilgi tarafından belirlenen ayarlanmış ağırlıklardır. Yardımcı bilgi vektörü yığın için $X_i=(X_{1i},X_{2i},X_{3i},.....,X_{pi})'$ örnek için $x_i=(x_{1i},x_{2i},x_{3i},.....,x_{pi})'$ olmak üzere yığın vektör toplamı $X^T = \sum_{i=1}^{N} X_i$ biliniyorsa, Denklem (4.11)'deki ayarlanma eşitliği sağlanıyorsa ağırlıklandırma ayarlanmıştır.

$$\sum_{i=1}^{n} w_i x_i = \sum_{i=1}^{N} X_i \tag{4.11}$$

Eğer bu eşitlik tüm yardımcı değişkenler için sağlanıyorsa; yani tüm yardımcı değişkenler için tahmin yığın toplamlarını veriyorsa ayarlanmış ağırlıklar tutarlıdır.

w_i ağırlıkları hesaplanırken mümkün olduğunca d_i tasarım ağırlıklarına yakın olması sağlanmalıdır. Öncelikle tasarım ağırlıkları ile gerekli ayarlanmış ağırlıkların arasındaki uzaklık tanımlanmalıdır. Daha sonra bu uzaklık, ayarlanmış ağırlıkları elde etmek için ayarlama koşulu Denklem (4.11)'yi sağlayacak biçimde minimize edilir.

Genel olarak, w_i ve d_i arasındaki uzaklık $G(w_i, d_i)$ ile gösterilir. Örnekleme tasarım ağırlıkları ile ayarlanmış ağırlıklar arasındaki toplam uzaklık;

$$U = \sum_{i=1}^{n} G(w_i, d_i) \tag{4.12}$$

şeklinde ifade edilebilir. Toplam uzaklığı minimum yapmak için Lagrange çarpanı kullanılır.

$$L = \sum_{i=1}^{n} G(w_i, d_i) + \sum_{j=1}^{p} \lambda_j \left[X_j^T - \sum_{j=1}^{n} w_i x_{ji} \right] \tag{4.13}$$

λ_j ‘ler Lagrange çarpanlarını göstermektedir. Amaç ayarlanmış ağırlıkları hesaplamak olduğundan, L denklemini minimum yapacak şekilde w_i değerlerini, U için ki-kare uzaklık ölçümünü yazarak elde edebiliriz.

Uzaklık ölçümü olarak araştırmacılar tarafından yaygın olarak kullanılan ki-kare uzaklık ölçümü kullanılmıştır. Yeniden ağırlıklandırmada ki-kare uzaklık ölçümü yerine diğer uzaklık ölçümleri de kullanılabilir. Ancak Deville ve Särndal (1992) diğer uzaklık ölçümlerinin ki-kare ölçümüyle hemen hemen aynı sonuçları verdiğini ve ki-kare uzaklık ölçümüne asimptotik olarak denk olduklarını göstermişlerdir. Ki-kare uzaklık ölçümü;

$$\sum_{i=1}^{n} G(w_i, d_i) = \frac{1}{2} \sum_{i=1}^{n} \frac{(w_i - d_i)^2}{d_i} \tag{4.14}$$

şeklindedir. Ki-kare uzaklık ölçümü Denklem (4.13)’te yerine konulursa Denklem (4.15)’daki eşitliğe ulaşılır.

$$L = \frac{1}{2}\sum_{i=1}^{n}\frac{(w_i - d_i)^2}{d_i} + \sum_{j=1}^{p}\lambda_j\left[X_j^T - \sum_{j=1}^{n} w_i x_{ji}\right] \quad (4.15)$$

Optimum w_i Denklem (4.15)'de L'nin w_i'ye göre türevinin alınıp sıfıra eşitlenmesiyle Denklem (4.16)'deki eşitlik elde edilir.

$$\frac{\partial L}{\partial w_i} = \left[\frac{w_i}{d_i} - 1\right] - \sum_{j=1}^{p}\lambda_j x_{ji} = 0 \quad (4.16)$$

$\frac{\partial L}{\partial w_i}$ eşitliğinde $\sum_{j=1}^{p}\lambda_j x_{ji}$ yerine $x_i^{'}\lambda$ yazılabilir. Buradan, Denklem (4.16)'da w_i yalnız bırakılırsa,

$$w_i = d_i(1 + x_i^{'}\lambda) \quad (4.17)$$

sonucuna ulaşılır. Ayarlanmış ağırlıkları elde etmek için gerekli olan λ

$$\lambda = \left[\sum_{i=1}^{n} d_i x_i x_i^{'}\right]^{-1}(X^T - \hat{X}^T) \quad (4.18)$$

eşitliği ile elde edilir. Denklem (4.17) ayarlama tahmin edicisinde yerine konulursa,

$$\hat{Y}_{AY} = \sum_{i=1}^{n} d_i y_i + (X^T - \hat{X}^T)' b_d \quad (4.19)$$

eşitliğine ulaşılır. Böylece ayarlama tahmin edicisi kullanılarak GREG tahmini elde edilmiş olur (Alkaya, Esin, 2005).

Ayarlama için elde edilen w_i ağırlıkları burada GREG tahminini elde etmede kullanıldığından GREG ağırlıkları olarak da adlandırılabilir.

GREG ağırlıkları Y bağımlı değişkeni olan bir regresyon modelinden gelmesine rağmen ağırlıklar Y değişkenine bağlı değildir. GREG ağırlıkları araştırma verisinin tüm analizlerinde kullanılabilen genel ağırlıklardır (Kalton, Cervantes, 2003). Bu sayede bir tür ayarlama ağırlığı olan GREG ağırlıklarını kullanarak ilgilenilen değişken Y dışında diğer değişkenlere ilişkin tahminlerin etkinliklerini de arttırabiliriz. Bu yönüyle GREG tahmin edicisi diğer tahmin edicilere avantaj sağlamaktadır.

BEŞİNCİ BÖLÜM

UYGULAMA

Uygulama aşamasında TÜİK tarafından yapılan Yıllık Sanayi ve Hizmet İstatistikleri Araştırmasının 2005 ve 2006 yılına ait veriler kullanılmıştır. Yıllık Sanayi ve Hizmet İstatistikleri Araştırmasında işyerlerinin ekonomik faaliyetleri, Avrupa Topluluğunda Ekonomik Faaliyetlerin İstatistiki Sınıflamasına (EFİS- Rev.1.1) göre sınıflandırılmıştır. Araştırmanın kapsamına giren iktisadi faaliyetler şunlardır:

- Madencilik ve Taşocakçılığı
- İmalat Sanayi
- Elektrik, Gaz ve Su
- Yapı (inşaat) ve Bayındırlık İşleri
- Toptan ve Perakende Ticaret; Motorlu Araç, Motosiklet, Kişisel ve Ev Eşyalarının Onarımı
- Otel, Lokanta ve Kahvehane
- Ulaştırma, Depolama ve Haberleşme
- Gayrimenkul Kiralama ve İş Faaliyetleri
- Eğitim
- Sağlık İşleri ve Sosyal Hizmetler
- Diğer Sosyal, Toplumsal ve Kişisel Hizmet Faaliyetleri

Yıllık Sanayi ve Hizmet İstatistikleri Araştırma tasarımı oluşturulurken İş Kayıtları adres çerçevesi kullanılarak örnekleme birimi girişim olarak belirlenmiştir. Madencilik ve Taşocaklığı, Elektrik, Gaz ve Su, Demiryolu Taşımacılığı, Boru Hattı Taşımacılığı, Havayolu Taşımacılığı iktisadi faaliyetlerdeki girişim bilgileri ile diğer iktisadi faaliyetlerde 20+ çalışanı olan girişimlere ilişkin bilgiler tamsayımla, kalan girişimlere ilişkin bilgiler ise örnekleme yöntemiyle elde edilmiştir.

Uygulamada ele alınacak iktisadi faaliyet imalat sanayidir. 2005 ve 2006 yılında yapılan Yıllık Sanayi ve Hizmet İstatistikleri Araştırmalarında İmalat Sanayi'nde 20+ çalışanı olan tüm girişimlere ait bilgiler tamsayım yöntemi ile elde edilmiştir. Uygulamada 2006 yılı Yıllık Sanayi ve Hizmet İstatistikleri Araştırmasında İmalat Sanayi'nde 20+ çalışanı olan girişimler için tamsayım ile elde edilen ortalama ciro parametresi ($\bar{Y}$), basit rasgele örnekleme ve tabakalı rasgele örnekleme yöntemleri için önceki bölümlerde anlatılan tahmin edicilerle tahmin edilmiştir. 2006 yılında İmalat Sanayi'nde 20+ çalışanı olan girişimler için $\bar{Y}$ parametresi tahmin edilirken aynı araştırmanın 2005 yılındaki değişkenleri yardımcı değişken olarak kullanılmıştır. Uygulama aşamasında SAS Enterprise Guide 3.0 ve R Gui 2.9.0 programları kullanılmıştır.

Yardımcı değişken kullanımında bazı tahmin ediciler için girişim düzeyinde eşleştirmeler gerekmektedir. Bazı girişimlerin faaliyetine son vermesi, bazılarının faaliyet değiştirmesi, yeni girişimlerin açılması vb. gibi nedenlerden ötürü 2005 ve 2006 yılına ilişkin girişimlerin hepsinin birebir eşleştirilmesi mümkün olamamıştır. 2005 ve 2006 yıllarında İmalat Sanayi'nde 20+ çalışanı olan girişim sayılarının eşleştirme bilgisi Tablo 5.1'de verilmiştir.

Tablo 5.1 2005 ve 2006 İmalat Sanayi 20+ Çalışanı Olan Girişimlerin Eşleşme Sayıları

	2005 yılı 20+ çalışanı olan girişim sayısı
2006 ile eşleşen	17006
2006 ile eşleşmeyen	899
Toplam	17995
	2006 yılı 20+ çalışanı olan girişim sayısı
2005 ile eşleşen	17405
2005 ile eşleşmeyen	2518
Toplam	19923

Tablo 5.1’de görüldüğü gibi 2006 yılında İmalat Sanayi’nde 20+ çalışanı olan girişim sayısı 19923’tür. Buna karşın $\bar{Y}$’yı tahmin etmek için girişim bazında eşleşmenin sağlanabildiği, değişkenleri yardımcı değişken olarak kullanılabilecek 2005 yılı girişim sayısı 17405’tir. 2005 yılına ait bilgisi olmayan 2518 girişime ilişkin değişkenlere ise, bulundukları EFİS 4’lü alt kısımların ortalama değerleri atanmıştır. EFİS 4’lü alt kısımlar İmalat Sanayi iktisadi faaliyet kolunun “Kağıt Hamuru İmalatı” vb. gibi alt kısımlarını belirtmektedir. Örneğin; 2005 yılında sektörde olmayan ve 2006 yılında sektöre girerek “Kağıt Hamuru İmalatı” yapan bir girişimin 2005 yılına ilişkin değişken değerleri olmayacaktır. Bu durumda, bu girişimin 2005 yılı değişken değerleri 2005 yılında “Kağıt Hamuru İmalatı” yapan tüm girişimlerin değişken değerlerinin ortalaması ile doldurulmuştur. Bu işleme literatürde “ortalama imputasyonu” adı verilmektedir.

5.1 $\bar{Y}$ PARAMETRESİNİN TAHMİNLERİNDE KULLANILACAK YARDIMCI DEĞİŞKENLERİN SEÇİMİ

$\bar{Y}$ parametresini tahmin ederken kullanılacak oran tahmin edicisi, regresyon tahmin edicisi vb. gibi tahmin edicileri kullanabilmek için öncelikle kullanılacak yardımcı değişkenler belirlenmiştir. 2006 yılında İmalat Sanayi sektöründe 20+ çalışanı olan girişimlerin ortalama cirosunu ($\bar{Y}$) tahmin etmek için 2005 yılı değişkenlerinden yardımcı değişken olarak kullanılabilir olan değişkenler şunlardır:

Tablo 5.2 İlgilenilen Değişken ve Kullanılabilir Yardımcı Değişkenler

	İmalat Sanayi 20+ Çalışanı Olan Girişimler İçin Değişkenler
Y	2006 Yılı Cirosu
X_1	2005 Yılı Cirosu
X_2	2005 Yılı Stok Değişimi
X_3	2005 Yılı Toplam Çalışan Sayısı
X_4	2005 Yılı Toplam Gelir
X_5	2005 Yılı Toplam Gider
X_6	2005 Yılı Üretim Değeri
X_7	2005 Yılı Personel Gideri

Eşleşen 17405 girişim için $X=(X_1,X_2,X_3,.....,X_7)$ yardımcı değişkenler vektöründeki değişkenlerin Y ilgilenilen değişken ile korelasyonları incelendiğinde;

Tablo 5.3 Y ile X Yardımcı Değişkenleri Arasındaki Korelasyonlar

	Y değişkeni ile X yardımcı değişkenleri arasındaki korelasyon (ρ_{yx_i})						
	X_1	**X_2**	**X_3**	**X_4**	**X_5**	**X_6**	**X_7**
Y	0.9909	0.4979	0.4957	0.9900	0.9910	0.9897	0.4957

7 değişkenin de önemli düzeyde Y değişkeni ile ilişkili olduğu görülmektedir. (X_1,X_4,X_5,X_6) değişkenleri öncelikle tercih edilmek üzere; bu 7 değişken de tek yardımcı değişken kullanan tahmin ediciler de kullanılabilir. Diğer yandan birden fazla yardımcı değişken kullanan GREG tahmin edicisi için kurulacak regresyon modelinde çoklu bağlantı sorununun önüne geçmek için yardımcı değişkenler arasındaki korelasyonlar da incelenmiştir.

Tablo 5.4 X_1 ile Diğer X Yardımcı Değişkenleri Arasındaki Korelasyonlar

	X_1 değişkeni ile diğer X yardımcı değişkenleri arasındaki korelasyon					
	X_2	**X_3**	**X_4**	**X_5**	**X_6**	**X_7**
X_1	0.4701	0.5371	**0.9977**	**0.9999**	**0.9954**	0.6620

Tablo 5.4’te görülen yüksek korelasyon ilişkisinden ötürü X_1 değişkeni modeldeyken (X_4,X_5,X_6) değişkenlerinin modelde olmasına gerek yoktur. Kalan (X_2,X_3,X_7) değişkenlerini incelersek;

Tablo 5.5 X_2 ile Kalan X Yardımcı Değişkenleri Arasındaki Korelasyonlar

	X_2 değişkeni ile kalan X yardımcı değişkenler arasındaki korelasyon		
	X_1	X_3	X_7
X_2	0.4701	0.3972	0.5048

Tablo 5.6 X_3 ile Kalan X Yardımcı Değişkenleri Arasındaki Korelasyonlar

	X_3 değişkeni ile kalan X yardımcı değişkenler arasındaki korelasyon		
	X_1	X_2	X_7
X_3	0.5371	0.3972	**0.8974**

Tablo 5.6’da X_3 ile X_7 arasında da yüksek korelasyon olduğu görülmektedir. Tüm değişkenlere ilişkin korelasyonlar incelendiğinde sonuç olarak Y ilgilenilen değişkeninin GREG tahmininde kullanılabilecek yardımcı değişkenlerin (X_1,X_2,X_3) oldukları belirlenmiştir. (X_4,X_5,X_6,X_7) değişkenleri oran tahmin edicilerinde kullanabilir olmasına rağmen uygulamada tahmin edicilerin etkinlikleri birbirleriyle karşılaştırılacağı için bu değişkenler kullanılarak tahmin üretilmemiştir. Sonuç olarak 2006 yılında İmalat Sanayi 20+ çalışanı olan girişimlerin ortalama cirosu ($\bar{Y}$) tahmin edilirken, aynı girişimlerin 2005 yılı cirosu (X_1), 2005 yılı stok değişimi (X_2) ve 2005 yılı çalışan sayısı (X_3) yardımcı değişken olarak kullanılmıştır.

5.2 BASİT RASGELE ÖRNEKLEME YÖNTEMİNDE $\bar{Y}$ TAHMİNİ

Yığın büyüklüğü N=19923 olmak üzere 2005-2006 İmalat Sanayi 20+ çalışanı olan girişimlerdeki ilgili değişkenler için yığın bilgileri Tablo 5.7’de verilmiştir. Y, X_1, X_2 değişkenlerine ilişkin değerler hesaplama ve gösterim kolaylığı için 100.000 ile bölünerek işlemler yapılmıştır.

Tablo 5.7 Uygulamada Kullanılan Değişkenlere İlişkin Yığın Bilgileri

N	19923	C_y	9.8730	ρ_{x_1y}	0.9892
$\bar{Y}$	178.9818	C_{x_1}	8.7062	ρ_{x_2y}	0.4945
$\bar{X}_1$	158.3825	C_{x_2}	13.6749	ρ_{x_3y}	0.4936
$\bar{X}_2$	3.2247	C_{x_3}	2.8568	$\rho_{x_1x_2}$	0.4701
$\bar{X}_3$	95.7900	C_{x_1y}	85.0243	$\rho_{x_1x_3}$	0.5369
S_y^2	3122593	C_{x_2y}	66.7626	$\rho_{x_2x_3}$	0.3956
$S_{x_1}^2$	1901384	C_{x_3y}	13.9217		
$S_{x_2}^2$	1944.621				
$S_{x_3}^2$	74884.4				

Basit rasgele örnekleme için örnek hacmi; Denklem (5.1)'deki eşitlikler kullanılarak hesaplanmıştır.

$$n_0 = \frac{t^2 S_y^2}{d^2} \qquad n = \frac{n_0}{1+\frac{n_0}{N}} \tag{5.1}$$

2006 ciro değişkeni için varyans çok büyük olduğundan bu formülde S_y^2 hesaplanırken aykırı değerler çıkarılmıştır. $1-\alpha=0.95$ güven düzeyinde hoş görülecek hata miktarı d=11 için örnek hacmi yaklaşık olarak n=2500 olarak hesaplanmıştır.

5.2.1 $\bar{Y}$ için Horvitz-Thompson Tahmini

Basit rasgele örneklemede $\bar{Y}$ için Horvitz-Thompson tahmini örneklem ortalamasına eşit olup tahmin edici yansızdır.

$$\hat{\bar{Y}} = \frac{1}{2500}\sum_{i=1}^{2500} y_i = \bar{y} = 184.6025 \tag{5.2}$$

$\hat{\bar{Y}}$ yansız tahmininin varyansı Denklem (5.3)'teki eşitlikle bulunur.

$$V(\bar{y}) = \frac{(1-f)}{n} S_y^2 = 1092.3040 \tag{5.3}$$

5.2.2 $\bar{\mathrm{Y}}$ için Klasik Oran Tahminleri

Belirlenen (X_1,X_2,X_3) yardımcı değişkenleri için 3 farklı klasik oran tahmini yapılmıştır. Elde edilen tahminler Denklem (5.4)'te verilmiştir.

$$\begin{aligned} \hat{\bar{Y}}_{ORAN_1} &= \frac{\bar{y}}{\bar{x}_1}\bar{X}_1 = 176.6560 \\ \hat{\bar{Y}}_{ORAN_2} &= \frac{\bar{y}}{\bar{x}_2}\bar{X}_2 = 119.7739 \\ \hat{\bar{Y}}_{ORAN_3} &= \frac{\bar{y}}{\bar{x}_3}\bar{X}_3 = 169.6846 \end{aligned} \tag{5.4}$$

Tahminlerin yanı ve hata kareler ortalaması ise sırasıyla Denklem (5.5) ve (5.6)'da verilmiştir.

$$\begin{aligned} \mathrm{Yan}(\hat{\bar{Y}}_{ORAN_1}) &= v\frac{1}{\bar{X}_1}(R_1 S_{x_1}^2 - \rho_{yx_1} S_y S_{x_1}) = -0.5777 \\ \mathrm{Yan}(\hat{\bar{Y}}_{ORAN_2}) &= v\frac{1}{\bar{X}_2}(R_2 S_{x_2}^2 - \rho_{yx_2} S_y S_{x_2}) = 7.5281 \\ \mathrm{Yan}(\hat{\bar{Y}}_{ORAN_3}) &= v\frac{1}{\bar{X}_3}(R_3 S_{x_3}^2 - \rho_{yx_3} S_y S_{x_3}) = -0.3607 \end{aligned} \tag{5.5}$$

$$HKO(\hat{\bar{Y}}_{ORAN_1}) = v\bar{Y}^2(C_y^2 + C_{x_1}^2 - 2\rho_{yx_1}C_yC_{x_1}) = 36.1373$$
$$HKO(\hat{\bar{Y}}_{ORAN_2}) = v\bar{Y}^2(C_y^2 + C_{x_2}^2 - 2\rho_{yx_2}C_yC_{x_2}) = 1691.5640 \quad (5.6)$$
$$HKO(\hat{\bar{Y}}_{ORAN_3}) = v\bar{Y}^2(C_y^2 + C_{x_3}^2 - 2\rho_{yx_3}C_yC_{x_3}) = 871.7477$$

$\hat{\bar{Y}}_{ORAN_2}$ tahmin edicisi ortalama tahmin edicisinden daha kötü sonuç vermektedir. Bunun nedeni oran tahmin edicisinin ortalama tahmininden daha etkin sonuç vermesi için gerekli koşul olan $\frac{C_{x_2}}{2C_y} = 0.6925 < \rho_{yx_2} = 0.4945$ koşulunun sağlanmamış olmasıdır. Diğer yandan $\hat{\bar{Y}}_{ORAN_1}$'in $\hat{\bar{Y}}_{ORAN_3}$'den daha etkin sonuç vermesinin nedeni ise Y değişkeni ile daha yüksek korelasyonu olmasıdır.

5.2.3 $\bar{Y}$ için Beale Oran Tahminleri

Belirlenen (X_1,X_2,X_3) yardımcı değişkenleri için 3 farklı Beale oran tahmini yapılmıştır. Elde edilen tahminler Denklem (5.7)'de verilmiştir.

$$\hat{\bar{Y}}_{BEALE_1} = \left(\frac{\bar{y}}{\bar{x}_1}\right)\left(\frac{1+vC_{yx_1}}{1+vC_{x_1}^2}\right)\bar{X}_1 = 177.2114$$
$$\hat{\bar{Y}}_{BEALE_2} = \left(\frac{\bar{y}}{\bar{x}_2}\right)\left(\frac{1+vC_{yx_2}}{1+vC_{x_2}^2}\right)\bar{X}_2 = 115.0454 \quad (5.7)$$
$$\hat{\bar{Y}}_{BEALE_3} = \left(\frac{\bar{y}}{\bar{x}_3}\right)\left(\frac{1+vC_{yx_3}}{1+vC_{x_3}^2}\right)\bar{X}_3 = 170.0255$$

Tahminlerin yanı ve hata kareler ortalaması ise sırasıyla Denklem (5.8) ve (5.9)'da verilmiştir.

$$\text{Yan}(\hat{\bar{Y}}_{BEALE_1}) = \nu \bar{Y}\left(\frac{1+\nu C_{yx_1}}{1+\nu C_{x_1}^2}\right)(C_{x_1}^2 - \rho_{yx_1} C_y C_{x_1}) = -0.5795$$

$$\text{Yan}(\hat{\bar{Y}}_{BEALE_2}) = \nu \bar{Y}\left(\frac{1+\nu C_{yx_2}}{1+\nu C_{x_2}^2}\right)(C_{x_2}^2 - \rho_{yx_2} C_y C_{x_2}) = 7.2309 \qquad (5.8)$$

$$\text{Yan}(\hat{\bar{Y}}_{BEALE_3}) = \nu \bar{Y}\left(\frac{1+\nu C_{yx_3}}{1+\nu C_{x_3}^2}\right)(C_{x_3}^2 - \rho_{yx_3} C_y C_{x_3}) = -0.3614$$

$$HKO(\hat{\bar{Y}}_{BEALE_1}) = \nu \bar{Y}^2\left(\frac{1+\nu C_{yx_1}}{1+\nu C_{x_1}^2}\right)^2(C_y^2 + C_{x_1}^2 - 2\rho_{yx_1} C_y C_{x_1}) = 36.3649$$

$$HKO(\hat{\bar{Y}}_{BEALE_2}) = \nu \bar{Y}^2\left(\frac{1+\nu C_{yx_2}}{1+\nu C_{x_2}^2}\right)^2(C_y^2 + C_{x_2}^2 - 2\rho_{yx_2} C_y C_{x_2}) = 1560.6400 \qquad (5.9)$$

$$HKO(\hat{\bar{Y}}_{BEALE_3}) = \nu \bar{Y}^2\left(\frac{1+\nu C_{yx_3}}{1+\nu C_{x_3}^2}\right)^2(C_y^2 + C_{x_3}^2 - 2\rho_{yx_3} C_y C_{x_3}) = 875.2545$$

Beale tahmin edicisinin klasik oran tahmin edicisinden daha etkin sonuç vermesi için gerekli olan $-1 < \left(\frac{1+\nu C_{yx}}{1+\nu C_x^2}\right) < 1$ koşulu sadece $\hat{\bar{Y}}_{BEALE_2}$ tahmin edicisinde sağlanmaktadır. Bu nedenle $\hat{\bar{Y}}_{BEALE_1}$ ve $\hat{\bar{Y}}_{BEALE_3}$ klasik tahmin edicilerinden daha etkin değildirler.

5.2.4 $\bar{Y}$ için Tin Oran Tahminleri

Belirlenen (X_1,X_2,X_3) yardımcı değişkenleri için 3 farklı Tin oran tahmini yapılmıştır. Elde edilen tahminler Denklem (5.10)'da verilmiştir.

$$\hat{\bar{Y}}_{TIN_1} = \left(\frac{\bar{y}}{\bar{x}_1}\right)[1+\nu(C_{yx_1} - C_{x_1}^2)]\bar{X}_1 = 177.2261$$

$$\hat{\bar{Y}}_{TIN_2} = \left(\frac{\bar{y}}{\bar{x}_2}\right)[1+\nu(C_{yx_2} - C_{x_2}^2)]\bar{X}_2 = 114.7361 \qquad (5.10)$$

$$\hat{\bar{Y}}_{TIN_3} = \left(\frac{\bar{y}}{\bar{x}_3}\right)[1+\nu(C_{yx_3} - C_{x_3}^2)]\bar{X}_3 = 170.0265$$

Tahminlerin yanı ve hata kareler ortalaması ise sırasıyla Denklem (5.11) ve (5.12)'de verilmiştir.

$$\text{Yan}(\hat{\bar{Y}}_{TIN_1}) = \nu\bar{Y}[1+\nu(C_{yx_1} - C_{x_1}^2)](C_{x_1}^2 - \rho_{yx_1}C_yC_{x_1}) = -0.5795$$

$$\text{Yan}(\hat{\bar{Y}}_{TIN_2}) = \nu\bar{Y}[1+\nu(C_{yx_2} - C_{x_2}^2)](C_{x_2}^2 - \rho_{yx_2}C_yC_{x_2}) = 7.2115 \qquad (5.11)$$

$$\text{Yan}(\hat{\bar{Y}}_{TIN_3}) = \nu\bar{Y}[1+\nu(C_{yx_3} - C_{x_3}^2)](C_{x_3}^2 - \rho_{yx_3}C_yC_{x_3}) = -0.3614$$

$$HKO(\hat{\bar{Y}}_{TIN_1}) = \nu\bar{Y}^2[1+\nu(C_{yx_1} - C_{x_1}^2)]^2(C_y^2 + C_{x_1}^2 - 2\rho_{yx_1}C_yC_{x_1}) = 36.3709$$

$$HKO(\hat{\bar{Y}}_{TIN_2}) = \nu\bar{Y}^2[1+\nu(C_{yx_2} - C_{x_2}^2)]^2(C_y^2 + C_{x_2}^2 - 2\rho_{yx_2}C_yC_{x_2}) = 1552.2600 \qquad (5.12)$$

$$HKO(\hat{\bar{Y}}_{TIN_3}) = \nu\bar{Y}^2[1+\nu(C_{yx_3} - C_{x_3}^2)]^2(C_y^2 + C_{x_3}^2 - 2\rho_{yx_3}C_yC_{x_3}) = 875.2646$$

Tin tahmin edicileri de $\hat{\bar{Y}}_{TIN_2}$ dışında klasik oran tahmin edicilerinden daha etkin sonuç vermemektedir. Tin tahmin edicisinin daha etkin olması için gerekli olan $-1 < [1+\nu(C_{yx} - C_x^2)] < 1$ koşulu sadece $\hat{\bar{Y}}_{TIN_2}$ tahmininde gerçekleşmektedir.

5.2.5 $\bar{Y}$ için Singh-Tailor Oran Tahminleri

Belirlenen (X_1,X_2,X_3) yardımcı değişkenleri için 3 farklı Singh-Tailor oran tahmini yapılmıştır. Elde edilen tahminler Denklem (5.13)'da verilmiştir.

$$\hat{\bar{Y}}_{SINTAIL_1} = \bar{y}\frac{\bar{X}_1+\rho_{yx_1}}{\bar{x}_1+\rho_{yx_1}} = 176.7032$$
$$\hat{\bar{Y}}_{SINTAIL_2} = \bar{y}\frac{\bar{X}_2+\rho_{yx_2}}{\bar{x}_2+\rho_{yx_2}} = 125.6402 \qquad (5.13)$$
$$\hat{\bar{Y}}_{SINTAIL_3} = \bar{y}\frac{\bar{X}_3+\rho_{yx_3}}{\bar{x}_3+\rho_{yx_3}} = 169.7549$$

Tahminlerin yanı ve hata kareler ortalaması ise sırasıyla Denklem (5.14) ve (5.15)'de verilmiştir.

$$\text{Yan}(\hat{\bar{Y}}_{SINTAIL_1}) = \nu\bar{Y}\left(\omega_1^2 C_{x_1}^2 - \omega_1\rho_{yx_1}C_y C_{x_1}\right) = -0.6034$$
$$\text{Yan}(\hat{\bar{Y}}_{SINTAIL_2}) = \nu\bar{Y}\left(\omega_2^2 C_{x_2}^2 - \omega_2\rho_{yx_2}C_y C_{x_2}\right) = 5.1775 \qquad (5.14)$$
$$\text{Yan}(\hat{\bar{Y}}_{SINTAIL_3}) = \nu\bar{Y}\left(\omega_3^2 C_{x_3}^2 - \omega_3\rho_{yx_3}C_y C_{x_3}\right) = -0.3614$$

$$HKO(\hat{\bar{Y}}_{SINTAIL_1}) = \nu\bar{Y}^2\left[C_y^2 + C_{x_1}^2\omega_1(\omega_1 - 2\theta_1)\right] = 37.4535$$
$$HKO(\hat{\bar{Y}}_{SINTAIL_2}) = \nu\bar{Y}^2\left[C_y^2 + C_{x_2}^2\omega_2(\omega_2 - 2\theta_2)\right] = 1370.3190 \qquad (5.15)$$
$$HKO(\hat{\bar{Y}}_{SINTAIL_3}) = \nu\bar{Y}^2\left[C_y^2 + C_{x_3}^2\omega_3(\omega_3 - 2\theta_3)\right] = 872.4120$$

Singh-Tailor tahmin edicileri de Beale ve Tin tahmin edicileri gibi sadece X_2'nin yardımcı değişken olduğu durumda klasik oran tahmin edicisinden daha etkin sonuç vermektedir. Bunun nedeni de $\rho_{yx} < \frac{C_x}{2C_y}(1+\omega)$ koşulunun sadece X_2 için sağlanmasıdır.

5.2.6 $\overline{Y}$ için Srivastava İki Değişkenli Oran Tahminleri

Srivastava iki değişkenli oran tahmin edicisinin uygulanabilmesi için gerekli koşul olan iki değişkenden birinin Y ile negatif korelasyonlu olması veri setimiz için geçerli olmadığından Srivastava iki değişkenli oran tahminleri yapılamamıştır.

5.2.7 $\overline{Y}$ için Klasik Doğrusal Regresyon Tahminleri

Belirlenen (X_1,X_2,X_3) yardımcı değişkenleri için 3 farklı klasik doğrusal regresyon tahmini yapılmıştır. Elde edilen tahminler Denklem (5.16)'da verilmiştir.

$$\begin{aligned}
\hat{\overline{Y}}_{REG_1} &= \overline{y} + b_1\left(\overline{X}_1 - \overline{x}_1\right) = 176.7039 \\
\hat{\overline{Y}}_{REG_2} &= \overline{y} + b_2\left(\overline{X}_2 - \overline{x}_2\right) = 175.1694 \\
\hat{\overline{Y}}_{REG_3} &= \overline{y} + b_3\left(\overline{X}_3 - \overline{x}_3\right) = 171.2094
\end{aligned} \tag{5.16}$$

Regresyon tahmin edicileri yansız olup bu yansız tahmin edicilerin varyansları ise Denklem (5.17)'de verilmiştir.

$$V(\hat{\bar{Y}}_{REG_1}) \cong \nu S_y^{\ 2}(1-\rho_{yx_1}^{\ \ 2}) = 23.5511$$
$$V(\hat{\bar{Y}}_{REG_2}) \cong \nu S_y^{\ 2}(1-\rho_{yx_2}^{\ \ 2}) = 825.2094 \qquad (5.17)$$
$$V(\hat{\bar{Y}}_{REG_3}) \cong \nu S_y^{\ 2}(1-\rho_{yx_3}^{\ \ 2}) = 826.1839$$

$(1-\rho_{yx}^{\ 2})<1$ koşulu 3 tahmin edici için de sağlandığından klasik doğrusal regresyon tahmin edicileri $\bar{y}$ tahmin edicisinden daha etkindir. Regresyon tahmin edicilerinin varyansları klasik oran tahmin edicilerinin HKO'ları ile karşılaştırıldığında her değişken için regresyon tahmin edicisinin klasik oran tahmin edicisinden daha etkin olduğu görülmektedir.

5.2.8 $\bar{Y}$ için Kaur Tahminleri

Belirlenen (X_1,X_2,X_3) yardımcı değişkenleri için 3 farklı Kaur regresyon tahmini yapılmıştır. Kaur regresyon tahminlerinin klasik doğrusal regresyon tahminlerinden daha etkin olması için $\kappa = \dfrac{1+\nu\rho_{yx}^{\ 2}C_y^{\ 2}}{1+\nu C_y^{\ 2}}$ değerleri $\kappa_1 = 0.9996$, $\kappa_2 = 0.9833$, $\kappa_3 = 0.9833$ şeklinde hesaplanmıştır. Elde edilen tahminler Denklem (5.18)'da verilmiştir.

$$\hat{\bar{Y}}_{KAUR_1} = \kappa_1\bar{y} + b\left(\bar{X}_1 - \bar{x}_1\right) = 176.5727$$
$$\hat{\bar{Y}}_{KAUR_2} = \kappa_2\bar{y} + b\left(\bar{X}_2 - \bar{x}_2\right) = 175.0381 \qquad (5.18)$$
$$\hat{\bar{Y}}_{KAUR_3} = \kappa_3\bar{y} + b\left(\bar{X}_3 - \bar{x}_3\right) = 171.0782$$

Tahmin edicilerin yanı ve hata kareler ortalaması ise sırasıyla Denklem (5.19) ve (5.20)'de verilmiştir.

$$\text{Yan}(\hat{\bar{Y}}_{KAUR_1}) = \bar{Y}(\kappa_1 - 1) = -0.0638$$

$$\text{Yan}(\hat{\bar{Y}}_{KAUR_2}) = \bar{Y}(\kappa_2 - 1) = -2.9902 \qquad (5.19)$$

$$\text{Yan}(\hat{\bar{Y}}_{KAUR_3}) = \bar{Y}(\kappa_3 - 1) = -2.9902$$

$$HKO(\hat{\bar{Y}}_{KAUR_1}) = \nu\left[\kappa_1^2 \bar{Y}^2 C_y^2 + \beta^2 \bar{X}_1^2 C_{x_1}^2 - 2\kappa_1 \beta \rho_{yx_1} \bar{X}_1 \bar{Y} C_{x_1} C_y\right] = 23.5344$$

$$HKO(\hat{\bar{Y}}_{KAUR_2}) = \nu\left[\kappa_2^2 \bar{Y}^2 C_y^2 + \beta^2 \bar{X}_2^2 C_{x_2}^2 - 2\kappa_2 \beta \rho_{yx_2} \bar{X}_2 \bar{Y} C_{x_2} C_y\right] = 824.0528 \qquad (5.20)$$

$$HKO(\hat{\bar{Y}}_{KAUR_3}) = \nu\left[\kappa_3^2 \bar{Y}^2 C_y^2 + \beta^2 \bar{X}_3^2 C_{x_3}^2 - 2\kappa_3 \beta \rho_{yx_3} \bar{X}_3 \bar{Y} C_{x_3} C_y\right] = 825.0259$$

(X_1,X_2,X_3)'ün Kaur regresyon tahmin edicileri seçilen κ değerleri için klasik doğrusal regresyon tahmin edicilerinden daha etkin olmakla beraber klasik doğrusal regresyon tahmin edicilerinden farklı olarak yanlıdırlar.

5.2.9 $\bar{Y}$ için GREG Tahminleri

Belirlenen (X_1,X_2,X_3) yardımcı değişkenleri için tek değişkenli klasik doğrusal regresyon tahminleri dışında ayrıca 2 değişkenli 3 GREG tahmini ve 3 değişkenli 1 GREG tahmini yapılmıştır. Elde edilen tahminler Denklem (5.21)'de verilmiştir.

X=(X_1,X_2) iken $\hat{\bar{Y}}_{GREG_{12}} = \bar{y} + \left(\bar{X} - \bar{x}\right)' b_{12} = 176.8864$

X=(X_1,X_3) iken $\hat{\bar{Y}}_{GREG_{13}} = \bar{y} + \left(\bar{X} - \bar{x}\right)' b_{13} = 176.1679$

X=(X_2,X_3) iken $\hat{\bar{Y}}_{GREG_{23}} = \bar{y} + \left(\bar{X} - \bar{x}\right)' b_{23} = 168.7208$ (5.21)

X=(X_1,X_2,X_3) iken $\hat{\bar{Y}}_{GREG_{123}} = \bar{y} + \left(\bar{X} - \bar{x}\right)' b_{123} = 177.2587$

GREG tahmin edicileri yansız olup bu yansız tahmin edicilerin varyansları ise Denklem (5.22)'de verilmiştir.

$$V(\hat{\hat{Y}}_{GREG_{12}}) = \nu S_y^2(1-R_{12}^2) = 21.3117$$

$$V(\hat{\hat{Y}}_{GREG_{13}}) = \nu S_y^2(1-R_{13}^2) = 21.6396$$

$$V(\hat{\hat{Y}}_{GREG_{23}}) = \nu S_y^2(1-R_{23}^2) = 712.4904 \qquad (5.22)$$

$$V(\hat{\hat{Y}}_{GREG_{123}}) = \nu S_y^2(1-R_{123}^2) = 19.4176$$

GREG tahmin edicilerinin varyans sonuçlarına bakıldığında değişken sayısı arttıkça etkinliğin arttığı görülmektedir. X_1 ve X_2'nin yardımcı değişken olduğu $\hat{\hat{Y}}_{GREG_{12}}$ tahmin edicisi, sadece X_1'in yardımcı değişken olduğu $\hat{\hat{Y}}_{REG_1}$ tahmin edicisinden, $\hat{\hat{Y}}_{GREG_{123}}$ tahmin edicisi ise her ikisinden de daha etkindir. $\hat{\hat{Y}}_{GREG_{23}}$ tahmin edicisi ile $\hat{\hat{Y}}_{REG_1}$ tahmin edicileri, yüksek ilişkili bir yardımcı değişkenin düşük ilişkili iki yardımcı değişkenden daha etkin tahminler verebileceğini göstermektedir.

Tablo 5.8 Basit Rasgele Örneklemede $\bar{Y}$ Tahmin Edicilerinin Etkinlik Sıralaması

$\bar{Y}$	178,9818		
	TAHMİN	**YAN**	**HKO**
$\hat{\bar{Y}}_{GREG_{123}}$	177,2587	-	19,4176
$\hat{\bar{Y}}_{GREG_{12}}$	176,8864	-	21,3117
$\hat{\bar{Y}}_{GREG_{13}}$	176,1679	-	21,6396
$\hat{\bar{Y}}_{KAUR_{1}}$	176,5727	-0,0638	23,5344
$\hat{\bar{Y}}_{REG_{1}}$	176,7039	-	23,5511
$\hat{\bar{Y}}_{ORAN_{1}}$	176,6560	-0,5777	36,1373
$\hat{\bar{Y}}_{BEALE_{1}}$	177,2114	-0,5795	36,3649
$\hat{\bar{Y}}_{TIN_{1}}$	177,2261	-0,5795	36,3709
$\hat{\bar{Y}}_{SINTAIL_{1}}$	176,7032	-0,6034	37,4535
$\hat{\bar{Y}}_{GREG_{23}}$	168,7208	-	712,4904
$\hat{\bar{Y}}_{KAUR_{2}}$	175,0381	-2,9902	824,0528
$\hat{\bar{Y}}_{KAUR_{3}}$	171,0782	-2,9902	825,0259
$\hat{\bar{Y}}_{REG_{2}}$	175,1694	-	825,2094
$\hat{\bar{Y}}_{REG_{3}}$	171,2094	-	826,1839
$\hat{\bar{Y}}_{ORAN_{3}}$	169,6846	-0,3607	871,7477
$\hat{\bar{Y}}_{SINTAIL_{3}}$	169,7549	-0,3614	872,4120
$\hat{\bar{Y}}_{BEALE_{3}}$	170,0255	-0,3614	875,2545
$\hat{\bar{Y}}_{TIN_{3}}$	170,0265	-0,3614	875,2646
$\hat{\bar{Y}} = \bar{y}$	184,6025	-	1092,3040
$\hat{\bar{Y}}_{SINTAIL_{2}}$	125,6402	5,1775	1370,3190
$\hat{\bar{Y}}_{TIN_{2}}$	114,7361	7,2115	1552,2600
$\hat{\bar{Y}}_{BEALE_{2}}$	115,0454	7,2309	1560,6400
$\hat{\bar{Y}}_{ORAN_{2}}$	119,7739	7,5281	1691,5640

Tablo 5.8'de basit rasgele örneklemede $\bar{Y}$ tahminlerinin hata kareler ortalamaları küçükten büyüğe sıralanmıştır. Bu sıralama aynı zamanda tahmin edicilerin $\bar{Y}$ tahminlerinin etkinlik sıralamasını da verir. Bu sıralamaya göre $X=(X_1,X_2,X_3)$ yardımcı değişkenlerinin hepsini tahmin aşamasında kullanan GREG tahmin edicisi en etkin tahmin edicidir. Tablo 5.8'deki tahminler N=19923 yığın büyüklüğünden n=2500 örnek hacmi ile basit rasgele olarak seçtiğimiz tek bir örnekten gelen tahminleri vermektedir. Tablo 5.8'de yan ve hata kareler ortalamaları hesaplanırken ise seçtiğimiz örnekten bağımsız olarak; basit rasgele örneklemede tahmin ediciler için belirlenmiş, yığın bilgilerini kullanan Bölüm 2'de verilen eşitlikler kullanılmaktadır. Basit rasgele örnekleme dışında kalan örnekleme yöntemlerinde tahminlerin yan ve hata kareler ortalaması için basit eşitlikler elde etmek zordur. Bu durumda tahmin edicilerin yan ve hata kareler ortalamasını Monte Carlo simülasyon yöntemi ile elde edebiliriz.

5.2.10 Basit Rasgele Örneklemede $\bar{Y}$ Tahmin Edicilerinin Yan ve HKO'larının Monte Carlo Simülasyon Yöntemiyle Elde Edilmesi

Monte Carlo simülasyonu sıklıkla bir tahmin edicinin örnekleme dağılımını tanımlamak zor olduğunda kullanılır. Dağılımı elde etmek için verilen tasarım altında olası bütün örnekler düşünülmelidir. Her bir örnek s için s'nin seçilme olasılığı p(s) ve tahmin edici $\hat{\bar{Y}}$'nin değeri bilinmelidir. $\hat{\bar{Y}}$'nin beklenen değerinin, yanının ve varyansının tam olarak hesaplanması bu durumda olası olur. Ancak olası tüm örneklerin sayısı çok olduğundan bu genellikle imkansızdır. İşte bu yüzden araştırmada kullanılan tahmin edicilerin istatistiksel özellikleri üstüne çalışılırken sıklıkla simülasyon yöntemi kullanılır (Särndal, Swensson, Wretman, 1992:277).

Monte Carlo simülasyonunun prensiplerine dayanarak basit rasgele örneklemedeki tahmin ediciler için şu aşamalar uygulanmıştır:

- Sonlu yığın ve basit rasgele örnekleme tasarımı sabitlenmiştir.
- Verilen tasarım ve yığın için K=5000 örnek seçilmiştir. Seçilen bir örnek diğeri seçilmeden yerine konulur; böylece örneğin seçildiği yığının hep aynı olması sağlanmıştır.
- Seçilen her örnek için basit rasgele örneklemedeki tahmin ediciler hesaplanmıştır. Eğer K yeterince büyükse K tahminlerinin dağılımı (ampirik örnekleme dağılımı), kolaylıkla elde edemediğimiz doğru örnekleme dağılımına çok yaklaşır.
- $\hat{\hat{Y}}_k$ k'ncı örnek için elde ettiğimiz tahmin olmak üzere tahmin ediciler için tahmin, yan ve hata kareler ortalaması sırasıyla Denklem (5.23), (5.24), (5.25)'deki gibi hesaplanmaktadır.

$$\bar{\hat{\hat{Y}}} = \frac{1}{5000}\sum_{k=1}^{5000}\hat{\hat{Y}}_k \tag{5.23}$$

Tahmin edicinin beklenen değeri $E(\hat{\hat{Y}})$'nin tahminidir.

$$Yan(\hat{\hat{Y}}) = E(\hat{\hat{Y}}) - \bar{Y} = \bar{\hat{\hat{Y}}} - \bar{Y} \tag{5.24}$$

Tahmin edici $\hat{\hat{Y}}$'nın yanının tahminidir.

$$S^2_{\hat{\hat{Y}}} = \frac{1}{4999}\sum_{k=1}^{5000}(\hat{\hat{Y}}_k - \bar{\hat{\hat{Y}}})^2 \tag{5.25}$$

Tahmin edicinin varyansı $V(\hat{\hat{Y}})$'nin tahminidir.

Hata kareler ortalamaları ise $HKO_{\hat{\bar{Y}}} = S^2_{\hat{\bar{Y}}} + \left[Yan(\hat{\bar{Y}}) \right]^2$ eşitliği ile hesaplanmıştır. Tablo 5.9'da basit rasgele örneklemede N=19923 büyüklüğündeki yığınımız için n=2500 ve K=5000 olmak üzere yapılan Monte Carlo simülasyon çalışmasının sonuçları verilmiştir.

Tablo 5.9'daki simülasyon sonuçları incelendiğinde Tablo 5.8'deki sonuçlarla benzerlik göstermektedir. K simülasyon sayısı yeterince büyük alındığında bu iki tablonun sonuçlarının yaklaşık olarak aynı olması beklenir. Yapılan simülasyon çalışması sonucunda da regresyon tahmin edicilerinin oran tahmin edicilerinden daha etkin sonuçlar verdiği görülmektedir. Diğer yandan $\hat{\bar{Y}}_{REG_1}$ ve $\hat{\bar{Y}}_{KAUR_1}$ tahmin edicilerinin GREG tahmin edicilerinden daha iyi sonuç vermesinin nedeni; bu tahmin edicilerin sadece Y ile yüksek korelasyona sahip X_1 değişkenini kullanmasıdır. X_1, X_2, X_3 değişkenlerini kullanan GREG tahminleri ise X_2, X_3 değişkenlerinin Y ile ilişkisiz olduğu örneklerde $\hat{\bar{Y}}_k$ tahminini kötü yaparak sonucu olumsuz etkilemektedir. Diğer yandan K yeterince büyük olduğunda bu olumsuz etki ortadan kalkmaktadır.

Sonuç olarak Monte Carlo simülasyon yöntemi kullanılarak tahmin edicilerin etkinlikleri birbirleri ile kıyaslanabilir. Korelasyon, varyans ve kovaryans büyüklüklerine bağlı olarak bu kıyaslamanın doğruluk derecesini arttırmak için K yeterince büyük alınmalıdır. Basit rasgele örneklemede ilgili tahmin ediciler için verilen eşitlikler kullanılarak hesaplanan HKO ile Monte Carlo simülasyonu sonucunda hesaplanan HKO Tablo 5.10'da karşılaştırılmıştır.

Tablo 5.9 Basit Rasgele Örneklemede $\overline{Y}$ Tahmin Edicilerinin Monte Carlo Simülasyon Sonuçları

$\overline{Y}$	178,9818		
	E(TAHMİN)	**YAN**	**HKO**
$\hat{\overline{Y}}_{REG_1}$	177,6234	-1,3584	24,0569
$\hat{\overline{Y}}_{KAUR_1}$	177,4961	-1,4857	24,3611
$\hat{\overline{Y}}_{GREG_{12}}$	177,7474	-1,2344	25,1643
$\hat{\overline{Y}}_{GREG_{13}}$	177,6532	-1,3286	25,1823
$\hat{\overline{Y}}_{GREG_{123}}$	177,7256	-1,2562	26,2020
$\hat{\overline{Y}}_{BEALE_1}$	179,0048	0,0230	30,7346
$\hat{\overline{Y}}_{TIN_1}$	179,0197	0,0379	30,7406
$\hat{\overline{Y}}_{ORAN_1}$	178,4437	-0,5381	30,8312
$\hat{\overline{Y}}_{SINTAIL_1}$	178,4212	-0,5606	31,8502
$\hat{\overline{Y}}_{GREG_{23}}$	171,3508	-7,6310	435,9458
$\hat{\overline{Y}}_{REG_2}$	169,3835	-9,5983	637,0622
$\hat{\overline{Y}}_{KAUR_2}$	169,2561	-9,7257	638,6726
$\hat{\overline{Y}}_{ORAN_3}$	178,8345	-0,1473	872,5360
$\hat{\overline{Y}}_{SINTAIL_3}$	178,8335	-0,1483	873,2706
$\hat{\overline{Y}}_{BEALE_3}$	179,1939	0,2121	876,0691
$\hat{\overline{Y}}_{TIN_3}$	179,1949	0,2131	876,0796
$\hat{\overline{Y}}_{KAUR_3}$	179,4967	0,5149	930,8654
$\hat{\overline{Y}}_{REG_3}$	179,6241	0,6423	932,2165
$\hat{\overline{Y}} = \overline{y}$	179,1441	0,1623	1115,1200
$\hat{\overline{Y}}_{SINTAIL_2}$	184,1037	5,1219	1663,9580
$\hat{\overline{Y}}_{TIN_2}$	178,9005	-0,0813	2078,6230
$\hat{\overline{Y}}_{BEALE_2}$	179,3828	0,4010	2089,9990
$\hat{\overline{Y}}_{ORAN_2}$	186,7556	7,7738	2325,5890

Tablo 5.10 Monte Carlo HKO'larının Doğru HKO'larla Karşılaştırılması

	HKO	HKO_{MC}
$\hat{\hat{Y}}_{GREG_{123}}$	19,4176	26,2020
$\hat{\hat{Y}}_{GREG_{12}}$	21,3117	25,1643
$\hat{\hat{Y}}_{GREG_{13}}$	21,6396	25,1823
$\hat{\hat{Y}}_{KAUR_{1}}$	23,5344	24,3611
$\hat{\hat{Y}}_{REG_{1}}$	23,5511	24,0569
$\hat{\hat{Y}}_{ORAN_{1}}$	36,1373	30,8312
$\hat{\hat{Y}}_{BEALE_{1}}$	36,3649	30,7346
$\hat{\hat{Y}}_{TIN_{1}}$	36,3709	30,7406
$\hat{\hat{Y}}_{SINTAIL_{1}}$	37,4535	31,8502
$\hat{\hat{Y}}_{GREG_{23}}$	712,4904	435,9458
$\hat{\hat{Y}}_{KAUR_{2}}$	824,0528	638,6726
$\hat{\hat{Y}}_{KAUR_{3}}$	825,0259	930,8654
$\hat{\hat{Y}}_{REG_{2}}$	825,2094	637,0622
$\hat{\hat{Y}}_{REG_{3}}$	826,1839	932,2165
$\hat{\hat{Y}}_{ORAN_{3}}$	871,7477	872,5360
$\hat{\hat{Y}}_{SINTAIL_{3}}$	872,4120	873,2706
$\hat{\hat{Y}}_{BEALE_{3}}$	875,2545	876,0691
$\hat{\hat{Y}}_{TIN_{3}}$	875,2646	876,0796
$\hat{\hat{Y}} = \overline{y}$	1092,3040	1115,1200
$\hat{\hat{Y}}_{SINTAIL_{2}}$	1370,3190	1663,9580
$\hat{\hat{Y}}_{TIN_{2}}$	1552,2600	2078,6230
$\hat{\hat{Y}}_{BEALE_{2}}$	1560,6400	2089,9990
$\hat{\hat{Y}}_{ORAN_{2}}$	1691,5640	2325,5890

5.3 TABAKALI RASGELE ÖRNEKLEME YÖNTEMİNDE $\bar{Y}$ TAHMİNİ

Bu bölümdeki uygulamada 2006 yılında İmalat Sanayi'nde 20+ çalışanı olan girişimlerin ortalama cirosunu ($\bar{Y}$) tahmin ederken örneklerimiz tabakalı rasgele örnekleme yöntemiyle seçilmiştir. İBBS2 bölge düzeyindeki 26 bölge tabaka olarak alınmıştır. Tabakalı örnekleme için n=2500 örnek Neyman Dağıtımı ile tabakalara dağıtılmıştır.

$$n_h = n\frac{N_h S_{yh}}{\sum_{h=1}^{26} N_h S_{yh}} \qquad h= 1, 2, ..., 25, 26 \qquad (5.26)$$

Yığındaki tabaka büyüklükleri ve örnekteki tabakalara düşen örnek hacimleri Tablo 5.11'de verilmiştir.

Tablo 5.11 Tabaka Büyüklükleri ve Örnek Hacimleri

i	N_i	n_i	i	N_i	n_i
1	452	31	**14**	13	2
2	1147	81	**15**	70	2
3	295	9	**16**	424	23
4	233	11	**17**	1116	756
5	1757	205	**18**	503	20
6	708	32	**19**	126	2
7	85	3	**20**	671	92
8	305	18	**21**	116	3
9	29	2	**22**	333	8
10	365	20	**23**	24	2
11	264	29	**24**	210	16
12	8758	951	**25**	22	2
13	1738	133	**26**	159	47

Basit rasgele örnekleme uygulamasında X_1, X_2, X_3 yardımcı değişkenlerinden X_2 ve X_3 'ün tek başlarına X_1 kadar iyi birer yardımcı değişken olmadıkları görülmüştür. Bu nedenle; gereksiz hesaplamalardan kaçınmak adına tabakalı rasgele örneklemede $\overline{Y}$'nin tahmini aşamasında X_2, X_3 tek başlarına kullanılmayacaktır.

5.3.1 $\overline{Y}$ için Klasik Ortalama Tahmini

Tabakalı rasgele örneklemede $\overline{Y}$ için klasik ortalama tahmin edicisi yansız olup Denklem (5.27)'deki gibi hesaplanır.

$$\hat{\overline{Y}}_{TABAKA} = \sum_{h=1}^{26} W_h \overline{y}_h = 169.8388 \tag{5.27}$$

$\hat{\overline{Y}}_{TABAKA}$ yansız tahmin edicisinin varyansı ise (5.28)'deki gibi olmaktadır.

$$V(\hat{\overline{Y}}_{TABAKA}) = \frac{1}{N^2} \sum_{h=1}^{26} N_h^2 V(\overline{y}_h) = 404,466 \tag{5.28}$$

5.3.2 $\overline{Y}$ için Bileşik Oran Tahmini

X_1 yardımcı değişkeni kullanılarak yapılan bileşik oran tahmini, bu tahmin edicinin yanı ve hata kareler ortalaması sırasıyla Denklem (5.29), (5.30) ve (5.31)'de hesaplanmıştır.

$$\hat{\overline{Y}}_{ORANBİL} = \left(\frac{\hat{\overline{Y}}_{TABAKA}}{\hat{\overline{X}}_{1_{TABAKA}}} \right) \overline{X}_1 = 175.2318 \tag{5.29}$$

$$Yan(\hat{\bar{Y}}_{ORANBİL}) \cong \frac{1}{\bar{X}_1}\sum_{h=1}^{26} W_h^2 v_h (R_1 S_{x_1h}^2 - S_{x_1yh}) = 1.9249 \tag{5.30}$$

$$HKO(\hat{\bar{Y}}_{ORANBİL}) \cong \sum_{h=1}^{26} W_h^2 v_h (S_{yh}^2 + R_1^2 S_{x_1h}^2 - 2R_1 S_{x_1yh}) = 22.3951 \tag{5.31}$$

Bileşik oran tahmin edicisi yanlı tahmin edici olmakla beraber klasik ortalama tahmin edicisinden daha düşük varyanslı bir tahmin elde etmiştir.

5.3.3 $\bar{Y}$ için Ayrı Oran Tahmini

X_1 yardımcı değişkeni kullanılarak yapılan ayrı oran tahmini, bu tahmin edicinin yanı ve hata kareler ortalaması sırasıyla Denklem (5.32), (5.33) ve (5.34)'de hesaplanmıştır.

$$\hat{\bar{Y}}_{ORANAYRI} = \sum_{h=1}^{26} W_h \frac{\bar{y}_h}{\bar{x}_{1_h}} \bar{X}_{1_h} = 173.7815 \tag{5.32}$$

$$\text{Yan}(\hat{\bar{Y}}_{ORANAYRI}) \cong \sum_{h=1}^{26} W_h \bar{Y}_h v_h (C_{x_1h}^2 - C_{x_1yh}) = -1.1888 \tag{5.33}$$

$$HKO(\hat{\bar{Y}}_{ORANAYRI}) \cong \sum_{h=1}^{26} W_h^2 v_h \bar{Y}_h^{\,2} (C_{yh}^2 + C_{x_1h}^2 - 2C_{x_1yh}) = 20.9717 \tag{5.34}$$

Ayrı oran tahmin edicisi bileşik oran tahmin edicisinden hem daha az yanlı hem de daha düşük varyanslı bir tahmin elde etmektedir.

5.3.4 $\bar{Y}$ için Sisodia-Dwivedi Oran Tahmini

X_1 yardımcı değişkeninin değişim katsayısı bilgisini kullanan Sisodia-Dwivedi oran tahmini, bu tahmin edicinin yanı ve hata kareler ortalaması sırasıyla Denklem (5.35), (5.36) ve (5.37)'de hesaplanmıştır.

$$\hat{\bar{Y}}_{SDTAB} = \hat{\bar{Y}}_{TABAKA} \frac{\sum_{h=1}^{26} W_h (\bar{X}_{1_h} + C_{x_1 h})}{\sum_{h=1}^{26} W_h (\bar{x}_{1_h} + C_{x_1 h})} = 175.0511 \quad (5.35)$$

$$Yan(\hat{\bar{Y}}_{SDTAB}) \cong \frac{1}{X_{1_{SD}}} \sum_{h=1}^{26} W_h^2 \nu_h (R_{1_{SD}} S_{x_1 h}^2 - S_{x_1 yh}) = -0.1513 \quad (5.36)$$

$$HKO(\hat{\bar{Y}}_{SDTAB}) \cong \sum_{h=1}^{26} W_h^2 \nu_h (S_{yh}^2 + R_{1_{SD}}^2 S_{x_1 h}^2 - 2R_{1_{SD}} S_{x_1 yh}) = 23.8452 \quad (5.37)$$

$M_1 > \dfrac{2L_1}{(R_1 + R_{1_{SD}})}$ koşulu sağlanmadığından Sisodia-Dwivedi oran tahmin edicisi bileşik oran tahmin edicisinden daha iyi sonuç vermemektedir.

5.3.5 $\bar{Y}$ için Singh-Kakran Oran Tahmini

X_1 yardımcı değişkeninin basıklık katsayısı bilgisini kullanan Singh-Kakran oran tahmini, bu tahmin edicinin yanı ve hata kareler ortalaması sırasıyla Denklem (5.38), (5.39) ve (5.40)'da hesaplanmıştır.

$$\hat{\bar{Y}}_{SKTAB} = \hat{\bar{Y}}_{TABAKA} \frac{\sum_{h=1}^{26} W_h (\bar{X}_{1_h} + G_{2h}(x_1))}{\sum_{h=1}^{26} W_h (\bar{x}_{1_h} + G_{2h}(x_1))} = 170.8580 \quad (5.38)$$

$$Yan(\hat{\bar{Y}}_{SKTAB}) \cong \frac{1}{X_{1_{SK}}} \sum_{h=1}^{26} W_h^2 \nu_h (R_{1_{SK}} S_{x_1 h}^2 - S_{x_1 yh}) = -0.3225 \quad (5.39)$$

$$HKO(\hat{\bar{Y}}_{SKTAB}) \cong \sum_{h=1}^{26} W_h^2 \nu_h (S_{yh}^2 + R_{1_{SK}}^2 S_{x_1 h}^2 - 2R_{1_{SK}} S_{x_1 yh}) = 275.9139 \quad (5.40)$$

Basıklık katsayısı kullanılarak elde edilen bu tahmin edici $M_1 > \dfrac{2L_1}{(R_1 + R_{1_{SK}})}$ eşitsizliğini sağlamadığı için bileşik oran tahmin edicisinden daha az etkin olduğu HKO değerlerinden açıkça görülmektedir.

5.3.6 $\overline{Y}$ için Upadhyaya-Singh Oran Tahminleri

X_1 yardımcı değişkeninin basıklık katsayısı ve değişim katsayısı bilgilerini birlikte kullanan Upadhyaya-Singh oran tahminleri, yanları ve hata kareler ortalamaları şu şekildedir:

$$\hat{\overline{Y}}_{US1TAB} = \hat{\overline{Y}}_{TABAKA} \frac{\sum_{h=1}^{26} W_h (\overline{X}_{1_h} G_{2h}(x_1) + C_{x_1h})}{\sum_{h=1}^{26} W_h (\overline{x}_{1_h} G_{2h}(x_1) + C_{x_1h})} = 184.2048 \tag{5.41}$$

$$Yan(\hat{\overline{Y}}_{US1TAB}) \cong \frac{1}{X_{1_{US1}}} \sum_{h=1}^{26} W_h^2 v_h (R_{1_{US1}} S_{x_1h}^2 - S_{x_1yh}) = -0.003 \tag{5.42}$$

$$HKO(\hat{\overline{Y}}_{US1TAB}) \cong \sum_{h=1}^{26} W_h^2 v_h (S_{yh}^2 + R_{1_{US1}}^2 S_{x_1h}^2 - 2R_{1_{US1}} S_{x_1yh}) = 403.3926 \tag{5.43}$$

$$\hat{\overline{Y}}_{US2TAB} = \hat{\overline{Y}}_{TABAKA} \frac{\sum_{h=1}^{26} W_h (\overline{X}_{1_h} C_{x_1h} + G_{2h}(x_1))}{\sum_{h=1}^{26} W_h (\overline{x}_{1_h} C_{x_1h} + G_{2h}(x_1))} = 177.3177 \tag{5.44}$$

$$Yan(\hat{\overline{Y}}_{US2TAB}) \cong \frac{1}{X_{1_{US2}}} \sum_{h=1}^{26} W_h^2 v_h (R_{1_{US2}} S_{x_1h}^2 - S_{x_1yh}) = -0.1815 \tag{5.45}$$

$$HKO(\hat{\overline{Y}}_{US2TAB}) \cong \sum_{h=1}^{26} W_h^2 v_h (S_{yh}^2 + R_{1_{US2}}^2 S_{x_1h}^2 - 2R_{1_{US2}} S_{x_1yh}) = 336.1249 \tag{5.46}$$

$\hat{\bar{Y}}_{US1TAB}$ ve $\hat{\bar{Y}}_{US2TAB}$ tahmin edicilerinin HKO'ları bileşik oran tahmin edicisinin HKO'sundan büyüktür. Dolayısıyla bileşik oran tahmin edicisi Upadhyaya-Singh oran tahmin edicilerinden etkindir. Bu sonuç $M_1 > \dfrac{2L_1}{(R_1 + R_{1_{US1}})}$ ve $M_1 > \dfrac{2L_1}{(R_1 + R_{1_{US2}})}$ koşulları incelenerek de gösterilebilir.

5.3.7 $\bar{Y}$ için Bileşik Regresyon Tahmini

X_1 yardımcı değişkeni kullanılarak elde edilen bileşik regresyon tahmini, bu tahminin yanı ve hata kareler ortalaması sırasıyla Denklem (5.47), (5.48) ve (5.49)'daki gibi hesaplanmıştır.

$$\hat{\bar{Y}}_{REGBİL} = \hat{\bar{Y}}_{TABAKA} + b_c(\bar{X}_1 - \hat{\bar{X}}_{1_{TABAKA}}) = 175.9140 \quad (5.47)$$

$$Yan(\hat{\bar{Y}}_{REGBİL}) \cong -\beta_c \sum_{h=1}^{26} W_h^2 v_h (\alpha_h^* \mu_{12h} - \alpha_h \mu_{03h}) = -0.9247 \quad (5.48)$$

$$HKO(\hat{\bar{Y}}_{REGBİL}) \cong \sum_{h=1}^{26} W_h^2 v_h S_{yh}^2 (1 - \rho_c^2) = 21.6007 \quad (5.49)$$

Bileşik regresyon tahmin edicisinde $\rho_c \neq 0$ koşulu sağlandığından klasik ortalama tahmin edicisinden daha etkindir. HKO değerlerine bakıldığında ayrı oran tahmin edicisi dışındaki tüm tahmin edicilerden de daha etkin olduğu görülmektedir.

5.3.8 $\bar{Y}$ için Ayrı Regresyon Tahmini

X_1 yardımcı değişkeni kullanılarak elde edilen ayrı regresyon tahmini, bu tahmin edicinin yanı ve hata kareler ortalaması sırasıyla Denklem (5.50), (5.51) ve (5.52)'deki gibi hesaplanmıştır.

$$\hat{\bar{Y}}_{REGAYRI} = \sum_{h=1}^{26} W_h \left[\bar{y}_h + b_h (\bar{X}_{1_h} - \bar{x}_{1_h}) \right] = 172.8299 \tag{5.50}$$

$$Yan(\hat{\bar{Y}}_{REGAYRI}) \cong \sum_{h=1}^{26} W_h \beta_h \bar{X}_{1_h} v_h C_{x_1 h} \left(\lambda_{03h} - \frac{\lambda_{12h}}{\rho_{x_1 yh}} \right) = -0.6478 \tag{5.51}$$

$$HKO(\hat{\bar{Y}}_{REGAYRI}) \cong \sum_{h=1}^{H} W_h^2 v_h \left[S_{yh}^2 + \beta_h^2 S_{x_1 h}^2 - 2\beta_h S_{x_1 yh} \right] = 18.8025 \tag{5.52}$$

Ayrı regresyon tahmin edicisi, ayrı oran tahmin edicisinden de daha etkin sonuç vermektedir. Ayrı regresyon tahmin edicisi ve ayrı oran tahmin edicisinin bileşik oran ve bileşik regresyon tahmin edicilerinden daha iyi sonuç vermelerinin nedeni $\bar{Y}$'nin X_1 ile olan ilişkisinin tabaka bazında değerlendirilmesidir. Bu sayede yardımcı değişkenin ilgili değişkene olan katkısı daha detay düzeyde olup tahmin edicinin etkinliği arttırılmaktadır.

Tablo 5.12 Tabakalı Rasgele Örneklemede $\bar{Y}$ Tahmin Edicilerinin Etkinlik Sıralaması

$\bar{Y}$	178,9818		
	TAHMİN	**YAN**	**HKO**
$\hat{\bar{Y}}_{REGAYRI}$	172,8299	-0,6478	18,8025
$\hat{\bar{Y}}_{ORANAYRI}$	173,7815	-1,1888	20,9717
$\hat{\bar{Y}}_{REGBİL}$	175,9140	-0,9247	21,6007
$\hat{\bar{Y}}_{ORANBİL}$	175,2318	1,9249	22,3951
$\hat{\bar{Y}}_{SDTAB}$	175,0511	-0,1512	23,8452
$\hat{\bar{Y}}_{SKTAB}$	170,8580	-0,3225	275,9139
$\hat{\bar{Y}}_{US2TAB}$	177,3177	-0,1815	336,1249
$\hat{\bar{Y}}_{US1TAB}$	184,2048	-0,0030	403,3926
$\hat{\bar{Y}}_{TABAKA}$	169,8388	-	404,4660

Tabakalı rasgele örneklemede kullanılan tahmin edicilerin $\overline{Y}$ tahminleri, bu tahmin edicilerin yanları ve hata kareler ortalamaları Tablo 5.12'de karşılaştırılmıştır.

5.3.9 $\overline{Y}$ için GREG Tahminleri

Seçtiğimiz tabakalı rasgele örneklemede Denklem (4.8) ve (4.9)'u kullanarak GREG tahminlerini elde etmemiz mümkündür. Diğer yandan GREG tahmin edicileri için basit rasgele örnekleme yönteminde olduğu gibi varyansı, R^2'nin basit bir fonksiyonu biçiminde hesaplayamayız. Bu nedenle tabakalı rasgele örneklemede GREG tahmin edicilerinin etkinliklerini diğer tahmin edicilerle kıyaslayabilmek için Monte Carlo simülasyon yöntemi kullanılmıştır.

5.3.10 Tabakalı Rasgele Örneklemede $\overline{Y}$ Tahmin Edicilerinin Yan ve HKO'larının Monte Carlo Simülasyon Yöntemiyle Elde Edilmesi

Basit rasgele örnekleme yönteminde olduğu gibi tabakalı rasgele örnekleme yönteminde de tahmin ediciler için yan ve hata kareler ortalamalarını Monte Carlo simülasyon yöntemiyle elde edebiliriz. GREG tahmin edicilerinin varyanslarını elde edemediğimiz için bu varyansları yaklaşık olarak tahmin etmemizi sağlayacak simülasyon yönteminden faydalanılmıştır. Basit rasgele örnekleme yönteminde uygulanan simülasyon aşamalarının aynısı burada da uygulanmıştır. Tabakalı örnekleme çekiminin basit rasgele örneklemeden daha fazla işlem gerektirmesi nedeniyle simülasyondaki K tekrar sayısı 1000 olarak alınmıştır. Tablo 5.11'de verilen tabaka örnek hacimleri dikkate alınarak K=1000 olmak üzere yapılan örnek çekimlerinin Monte Carlo simülasyon çalışmasının sonuçları Tablo 5.13'te verilmiştir.

Tablo 5.13 Tabakalı Rasgele Örneklemede $\overline{Y}$ Tahmin Edicilerinin Monte Carlo Simülasyon Sonuçları

	E(TAHMİN)	YAN	HKO
$\hat{\bar{Y}}_{ORANAYRI}$	178,8947	-0,0871	18,7776
$\hat{\bar{Y}}_{REGAYRI}$	177,6150	-1,3668	20,9278
$\hat{\bar{Y}}_{ORANBİL}$	179,1001	0,1183	22,1981
$\hat{\bar{Y}}_{REGBİL}$	178,9938	0,0120	22,9167
$\hat{\bar{Y}}_{GREG_{13}}$	178,5136	-0,4682	23,8660
$\hat{\bar{Y}}_{SDTAB}$	179,0536	0,0718	23,9723
$\hat{\bar{Y}}_{GREG_{123}}$	178,8905	-0,0913	25,0395
$\hat{\bar{Y}}_{GREG_{12}}$	179,3143	0,3325	25,1056
$\hat{\bar{Y}}_{US1TAB}$	179,7484	0,7666	95,5000
$\hat{\bar{Y}}_{US2TAB}$	178,9633	-0,0185	106,1493
$\hat{\bar{Y}}_{SKTAB}$	179,1563	0,1745	283,5287
$\hat{\bar{Y}}_{TABAKA}$	179,5466	0,5648	418,6796

Tabakalı rasgele örneklemede Monte Carlo simülasyon sonuçları incelendiğinde GREG tahmin edicilerinin $\hat{\bar{Y}}_{REGAYRI}$ tahmin edicisinden daha az etkin sonuçlar verdiği görülmektedir. Diğer yandan gerçekte $\hat{\bar{Y}}_{ORANAYRI}$ tahmin edicisinden daha etkin olan $\hat{\bar{Y}}_{REGAYRI}$ tahmin edicisi simülasyon sonucunda daha az etkin gözükmektedir. Bu farklı örneklerdeki değişkenlikten kaynaklanıp K=1000 örnek çekim sayısı arttırıldığında gerçek sonuçlara daha fazla yaklaşılacaktır. X_1 değişkeninin Y ile yüksek korelasyonlu olması nedeni ile GREG tahmin edicilerinin Monte Carlo simülasyon sonuçlarının bu veri seti için $\hat{\bar{Y}}_{REGAYRI}$ tahmin edicileri ile karşılaştırılması sağlıklı olmayabilir. Sonuç olarak veri setimizde $\overline{\mathrm{Y}}$ parametresinin GREG tahmin edicilerinin etkinlikleri için net bir ifade kullanamayız. Tabakalı rasgele örnekleme için bulduğumuz en etkin tahmin edici $\hat{\bar{Y}}_{REGAYRI}$ tahmin edicisidir.

SONUÇ VE DEĞERLENDİRME

Örnek araştırmalarında istatistikçilerin en önemli amacı, ilgilenilen değişkene ilişkin parametre tahmininin etkinliğini arttırmaktır. Daha etkin ve daha iyi tahminler elde etmek için yardımcı değişkenlerden faydalanılmaktadır. İlgilenilen değişken Y'nin ortalaması tahmin edilirken, yardımcı değişkene ilişkin bilgiler kullanılarak oran tahmin edicileri ve regresyon tahmin edicileri hesaplanmaktadır. Oran tahmin edicileri ve regresyon tahmin edicilerinin hesaplanması örnekleme yöntemlerine göre farklılıklar gösterir. Çalışmada basit rasgele örnekleme ve tabakalı rasgele örnekleme yöntemlerinde kullanılan bazı oran ve regresyon tahmin edicileri incelenmiştir.

Basit rasgele örnekleme yönteminde incelenen oran tahmin edicilerinin tahmin değerlerinin, yanlarının ve hata kareler ortalamalarının hesaplanabilmesi için, yardımcı değişkenin yığın ortalaması ve değişim katsayısı bilgisine ihtiyaç duyulmaktadır. Oran tahmin edicileri yanlı tahminler üretmekle beraber Tablo 6.1'deki koşulların sağlandığı durumlarda klasik oran tahmin edicisi, ortalama tahmin edicisinden; diğer oran tahmin edicileri de klasik oran tahmin edicisinden daha etkin ve iyi sonuçlar verir.

Tablo 6.1 Basit Rasgele Örneklemede Oran Tahmin Edicileri Etkinlik Koşulları

Karşılaştırma	Koşul
$HKO(\hat{\bar{Y}}_{ORAN}) < Var(\bar{y})$	$\frac{C_x}{2C_y} < \rho_{yx}$
$HKO(\hat{\bar{Y}}_{BEALE}) < HKO(\hat{\bar{Y}}_{ORAN})$	$-1 < \left(\frac{1+vC_{yx}}{1+vC_x^2}\right) < 1$
$HKO(\hat{\bar{Y}}_{TIN}) < HKO(\hat{\bar{Y}}_{ORAN})$	$-1 < [1+v(C_{yx} - C_x^2)] < 1$
$HKO(\hat{\bar{Y}}_{SINTAIL}) < HKO(\hat{\bar{Y}}_{ORAN})$	$\rho_{yx} < \frac{C_x}{2C_y}(1+\omega)$

Basit rasgele örnekleme yönteminde klasik doğrusal regresyon tahmin edicisi $\hat{\bar{Y}}_{REG}$'e ait tahmin değerinin elde edilebilmesi için örneğe çıkan birimlerde yardımcı değişken değerlerinin bilinmesi gereklidir. $\hat{\bar{Y}}_{REG}$, korelasyon katsayısı $\rho_{yx} \neq 0$ olduğu sürece ortalama tahmin edicisinden daha etkindir. $\hat{\bar{Y}}_{REG}$, $\beta \neq \bar{Y} / \bar{X}$ olduğu sürece de klasik oran tahmin edicisi $\hat{\bar{Y}}_{ORAN}$'dan daha etkindir. Kaur regresyon tahmin edicisi ise $2\rho_{yx}^{2} - 1 < \kappa < 1$ koşulunun sağlandığı durumlarda $\hat{\bar{Y}}_{REG}$ tahmin edicisinden biraz daha etkindir. $\hat{\bar{Y}}_{KAUR}$ az miktardaki etkinlik kazancını sağlarken yanlı tahmin üretmektedir. Regresyon tahmin edicileri genel olarak oran tahmin edicilerinden daha etkin sonuçlar vermektedir. Zaten klasik oran tahmin edicisi de kesme noktasının (0,0) olduğu bir tip klasik doğrusal regresyon tahmin edicisidir. $\hat{\bar{Y}}_{ORAN}$ tahmin edicisi çok küçük örnek hacimlerinde $\hat{\bar{Y}}_{REG}$ tahmin edicisinden daha etkin sonuç verebilir.

Tabakalı rasgele örnekleme yönteminde; ilgilenilen değişken ile yardımcı değişken her tabakada ayrı ayrı ilişkilendirilerek tahmin yapılıyorsa "ayrı tahmin", ilişkilendirme genel düzeyde yapılıyorsa "bileşik tahmin" elde edilir. Bileşik oran tahmin edicisi, tabakalı rasgele örnekleme için klasik ortalama tahmin edicisi $\hat{\bar{Y}}_{TABAKA}$'yı yardımcı değişken bilgisi $\hat{\bar{X}}_{TABAKA}$ ile oranlar. Ayrı oran tahmin edicisi ise her tabakadaki ortalamaları ayrı ayrı oranlar $\frac{\bar{y}_h}{\bar{x}_h}$. Kadılar ve Çıngı (2003) basit rasgele örnekleme için önerilmiş Sisodia-Dwivedi, Singh-Kakran ve Upadhyaya-Singh tahmin edicilerini tabakalı rasgele örnekleme için genelleştirmişlerdir. Bu tahmin edicilerin bileşik oran tahmin edicisinden daha etkin sonuç verebilmeleri için gerekli koşullar Tablo 6.2'de verilmiştir.

Tablo 6.2 Tabakalı Rasgele Örneklemede Oran Tahmin Edicileri Etkinlik Koşulları

Karşılaştırma	Koşul
$HKO(\hat{\hat{Y}}_{SDTAB}) < HKO(\hat{\hat{Y}}_{ORANBİL})$	$(R-R_{SD})(R+R_{SD})>0 \rightarrow M > \frac{2L}{(R+R_{SD})}$ $(R-R_{SD})(R+R_{SD})<0 \rightarrow M < \frac{2L}{(R+R_{SD})}$
$HKO(\hat{\hat{Y}}_{SKTAB}) < HKO(\hat{\hat{Y}}_{ORANBİL})$	$(R-R_{SK})(R+R_{SK})>0 \rightarrow M > \frac{2L}{(R+R_{SK})}$ $(R-R_{SK})(R+R_{SK})<0 \rightarrow M < \frac{2L}{(R+R_{SK})}$
$HKO(\hat{\hat{Y}}_{US1TAB}) < HKO(\hat{\hat{Y}}_{ORANBİL})$	$(R-R_{US1})(R+R_{US1})>0 \rightarrow M > \frac{2L}{(R+R_{US1})}$ $(R-R_{US1})(R+R_{US1})<0 \rightarrow M < \frac{2L}{(R+R_{US1})}$
$HKO(\hat{\hat{Y}}_{US2TAB}) < HKO(\hat{\hat{Y}}_{ORANBİL})$	$(R-R_{US2})(R+R_{US2})>0 \rightarrow M > \frac{2L}{(R+R_{US2})}$ $(R-R_{US2})(R+R_{US2})<0 \rightarrow M < \frac{2L}{(R+R_{US2})}$

Tabakalı rasgele örneklemede regresyon tahmin edicileri için β parametresi her tabakada ayrı ayrı tahmin edilerek tahmin edici hesaplanırsa ayrı regresyon tahmini, veri seti için genel bir β parametresi tahmin edilerek tahmin hesaplanırsa bileşik regresyon tahmini elde edilir. Regresyon tahmin edicileri ve oran tahmin edicileri arasındaki etkinliği karşılaştıracak genel bir koşul elde edilememiştir.

Genelleştirilmiş regresyon (GREG) tahmin edicisi; birden fazla yardımcı değişkenin kullanılabilir olduğu durumlarda, çoklu doğrusal regresyon modelini kullanarak ilgilenilen değişkeni tahmin eder. Basit rasgele örneklemede $R^2 > \rho_{yx}^{\;2}$ koşulunun sağlandığı durumda $\hat{\hat{Y}}_{GREG}$ tahmin edicisi $\hat{\hat{Y}}_{REG}$ tahmin edicisinden daha etkindir. GREG tahmin edicisi tasarım ağırlıkları kullanılarak da elde edilebilir. Bu nedenle tasarım ağırlıklarının bilindiği tüm örnekleme yöntemlerinde kullanılabilir. GREG tahmin edicisinin diğer bir avantajı tasarım ağırlıkları ve yardımcı değişkenler kullanılarak GREG ağırlıklarının elde edilebilmesidir. GREG ağırlıkları kullanılarak Y değişkeni dışında örnek araştırmasından elde edilen diğer değişkenlere ilişkin tahminlerin etkinlikleri de arttırılabilir. Böylece; bir araştırmada farklı değişkenlere ilişkin tahminlerin etkinliklerini arttırmak için, birden fazla oran veya regresyon tahmin edicisi kullanmak yerine tek bir GREG tahmin edicisi kullanılabilir.

Uygulamada 2005 ve 2006 yılına ait Yıllık Sanayi ve Hizmet İstatistikleri Araştırmasındaki İmalat Sanayi'nde 20+ çalışanı olan 19923 girişime ait veriler kullanılmıştır. Yıllık Sanayi ve Hizmet İstatistikleri Araştırmasında bu girişimlere ilişkin bilgiler tamsayım yöntemiyle toplandığından 2006 yılında yaptıkları ortalama ciro ($\overline{Y}$) parametresi bilinmektedir. Yaptığımız uygulamada 2006 ciro bilgisinin N=19923 girişimden tamsayım yöntemi ile alınmak yerine, n=2500 girişimden örnekleme yöntemi ile alınması durumunda, ortalama cironun 2005 yılındaki uygun değişkenleri yardımcı değişken olarak kullandığı farklı tahmin ediciler incelenmiştir.

Örneklerin basit rasgele örnekleme yöntemiyle seçilmesi durumunda 2005 yılına ait ciro, stok değişimi ve çalışan sayısı değişkenlerini yardımcı değişken olarak kullanan $\hat{\overline{Y}}_{GREG_{123}}$ tahmin edicisi en etkin tahmini vermektedir. Tablo 5.8 incelendiğinde GREG tahmin edicilerinin, regresyon tahmin edicilerinden, regresyon tahmin edicilerinin de oran tahmin edicilerinden daha etkin sonuçlar verdiği görülmüştür.

Tabakalı rasgele örnekleme yönteminde GREG tahmin edicilerinin varyansları kesin olarak hesaplanamadığı için Monte Carlo simülasyon yöntemiyle HKO'lar tahmin edilmiştir. Simülasyon sonuçlarında GREG tahmin edicilerinin HKO'su yaklaşık olarak elde edilmiştir; ancak bu sonuçların diğer tahmin edicilerle karşılaştırılması 2005 ciro yardımcı değişkeninin Y ile yüksek korelasyonlu olmasından dolayı sağlıklı görülmemiştir. Diğer yandan GREG tahmin edicileri dışında kalan tahmin edicilerden en etkin olanı $\hat{\bar{Y}}_{REGAYRI}$ ayrı regresyon tahmin edicisidir. Tablo 5.12 incelendiğinde tabakalı rasgele örnekleme yönteminde ayrı tahmin edicilerin bileşik tahmin edicilerden daha etkin sonuçlar verdiği görülmektedir. Bu sonuçlar uygulama yaptığımız veri seti için geçerli olup farklı verilerde farklı sonuçlar çıkabilir.

Tahmin aşamasında uygun yardımcı değişken kullanımı tahminin etkinliğini arttırmaktadır. Yardımcı değişken veya değişkenler kullanarak tahmin üreten oran tahmin edicileri ve regresyon tahmin edicileri arasından tercih yapılırken hata kareler ortalamaları karşılaştırılarak en etkin olan tahmin belirlenebilir. Tahmin edicilerin hata kareler ortalaması örnekleme tasarımının karmaşıklığı vb. gibi nedenlerle elde edilemeyebilir. Bu durumda, tahmin edicilerin etkinliğini karşılaştırmak için Monte Carlo simülasyon yönteminin kullanılması önerilir. Monte Carlo simülasyon yöntemi, hata kareler ortalamalarını yaklaşık olarak tahmin ederek tahmin edicilerin etkinlikleri hakkında fikir vermektedir.

KAYNAKÇA

ALKAYA, A., ESİN A.; Ayarlama Tahmin Edicisi, G.Ü. Fen Bilimleri Dergisi 18(4), 2005, s. 591-601.

BEALE, E.M.L.; "Some Use of Computers in Operational Research", **Industrielle Organisation**, 31, 1962, s.27-28.

BETHLEHEM, J.,COBBEN, F., SCHOUTEN B.; **Nonresponse in Household Surveys**, Statistics Netherlands, Version 2, 2007.

CASSEL, C.M., SÄRNDAL, C.E., WRETMAN J.H.; "Some results on generalized difference estimation and generalized regression estimation for finite populations", **Biometrika**, 63, 1976, s. 615-620.

CHAKRABARTY, R.P.; "Some Ratio-type Estimators",**Journal of the Indian Society of Agricultural Statistics**, XXXI, 1, 1979, s.49-62.

COCHRAN, W.G.;" Sampling theory when the sampling units are of unequal sizes", **Journal of the American Statistical Association**, 37, 1942, s. 199-212.

COCHRAN, W.G.; **Sampling Techniques**, New York, John Wiley&Sons, 2th ed., 1964.

COCHRAN, W.G.; **Sampling Techniques**, New York, John Wiley&Sons, 3th ed., 1977.

ÇINGI, H.; **Örnekleme Kuramı**, Beytepe, H.Ü. Fen Fakültesi Basımevi, 1994.

DEMING, W. E.; "On Errors in Surveys", **American Sociological Review**, 9, 1944, s. 359-369.

DEMING, W. E.; **Some Theory of Sampling**, New York, Wiley, 1950.

DEMING, W. E.; **Some Elementary Theory for Design**, New York, John Wiley, 1960.

DEVILLE, J., SÄRNDAL C.E.; "Calibration estimators in survey sampling", **Journal of the American Statistical Association**, 87, 1992, s. 376-382.

ESİN, A.; "Örnekleme Metotları ve Bir Uygulama", **Ankara İktisadi ve Ticari İlimler Akademisi**; yayın no: 97, 1975.

ESTEVAO, V.M., SÄRNDAL, C. E.; “A Functional Form Approach to Calibration”, **Journal of Official Statistics**, Vol.16, No. 4, 2000, s. 379-399.

FULLER, W.A.; **Regression for sample surveys**, Austria, Paper presented at meeting of International Statistical Institute, Vienna, 1973.

FULLER, W.A.; “Regression analysis for sample survey”, **Sankhyā,** Series C, 37, 1975, s. 117-132.

FULLER, W.; “Regression Estimation for Survey Samples, Survey Methodology”, **Statistics Canada**, Volume 28, No 1, 2002, s. 5-23.

HANSEN, M.H., HURWITZ, W.N, GURNEY, M.; “Problems and methods of the sample survey of business”, **Journal of American Statistical Association**, 41, 1946, s.173- 189.

HANSEN, M. H., HURWITZ, W.N., MADOW, W.G; **Sample Survey Methods and Theory**, New York, John Wiley and Sons, 1964.

HEDEYAT A.S., SINHA, B.K.; **Design and Inference in Finite Population Sampling**, John Wiley&Sons, New York, 1991.

HORVITZ, D.G., THOMPSON, D.J.; “A Generalization of Sampling Without Replacement from a Finite Universe**.” Journal of the American Statistical Association 47**, 1952, s. 663-685.

JESSEN, R.J.; **Statistical investigation of a sample survey for obtaining farm facts**, Iowa Agriculture Experiment Station Research Bulletin, 304, 1942.

KADILAR, C., ÇINGI, H.; “Ratio Estimators in Stratified Random Sampling”, **Biometrical Journal,** 45, 2, 2003, s. 218–225.

KADILAR, C., ÇINGI, H.; “An Improvement in Estimating the Population Mean by Using the Correlation Coefficient”, **Hacettepe Journal of Mathematics and Statistics**, Vol. 35 (1), 2006, s. 103 -109.

KALTON, G., FLORES-CERVANTES I.; “Weighting Methods”, **Journal of Official Statistics**, Vol.19, No. 2, 2003, s. 81-97.

KARAKÜLAH, Ü.H.; Basit rasgele örnekleme yönteminde oransal tahmin ediciler, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 2006.

KISH, L.; **Survey Sampling**, New York, John Wiley&Sons, 1965

KOYUNCU, N.; Tabakalı Rasgele Örneklemede Yardımcı Değişken Bilgisi Kullanılarak Kitle Ortalaması ve Varyansının Tahmin Edilmesi, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 2007.

MENENDEZ, E., REYES, A.; "On an Efficient Regression Type Estimator", **Biometrical Journal**, 40, 1, 1998, s. 79-84.

MUSSA, S. Abeidi; "A New Bias-reducing Modification of the Finite Population Ratio Estimator and a Comparison Among Proposed Alternatives", **Journal of Official Statistics**, Vol.15, 1999, s. 25-38.

NEYMAN, J.; "On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection" **Journal of the Royal Statistical Society**, 97, 1934,s. 558–625.

QUENOUILLE, M.H.; "Notes on Bias Estimation", **Biometrika**, 43, 1956, s. 353-360.

RAY, S.K., R.K.SINGH; "Difference cum product type estimators", **Journal of Indian Statistics Association**, 19, 1981, s. 147-151.

ROYALL, R.M., CUMBERLAND W.G.; "The finite population linear regression estimator and estimators of its variance, an empirical study" **Journal of the American Statistical Association**, 76, 1981, s. 924-930.

SÄRNDAL, C.E, SWENSSON, B., WRETMAN,J.; **Model Assisted Survey Sampling**, New York, Springer-Verlag, 1992.

SCHEAFFER, R.L, MENDENHALL, W., OTT, L.; **Elementary Survey Sampling**, Belmont-California, PWS-KENT Publishing, 1990.

SERPER, Ö., AYTAÇ, M.; **Örnekleme**, Bursa, Ezgi Kitabevi, 2. Baskı, 2000.

SINGH, H. P., KAKRAN, M.S.; A Modified Ratio Estimator Using Known Coefficient of Kurtosis of an Auxiliary Character, 1993, (yayımlanmamış)

SINGH, H.P., TAILOR, R.; "Use of Known Correlation Coefficient in Estimating the Finite Population Mean", **Statistics in Transition**, 6, 4, 2003, s. 555-560.

SISODIA, B.V.S., DWIVEDI, V.K.; "A modified ratio estimator using coefficient of variation of auxiliary variable", **Journal of Indian Society Agricultural Statistics**, 33, 1981, s. 13-18.

SRIVASTAVA, S.K.; "An Estimate of The Mean of a Finite Population Using Several Auxiliary Variables", **Journal of Indian Statistical Association**, 3, 1965, s. 189-194.

STEPHAN, F. F.; “History of the uses of modern sampling procedures”, **Journal of the American Statistical Association**, 43, 1948, s. 12–39.

SUKHATME, P.V., SUKHATME, B.V.; **Sampling Theory of Survey with Application**, Iowa State University Pres, 1984.

SUKHATME, P., SUKHATME, B., SUKHATME, S., ASOK, C.; **Sampling Theory of Surveys with Applications,** Iowa State Uni., New Delphi, Ames**,** Iowa and Indian Society of Agricultural Statistics, 1970.

TIN, M.; “Comparison of Some Ratio Estimators”, **Journal of American Statistical Assosiation**, 60, 1965, s. 294-307.

UPADHYAYA, L.N., SINGH, H.P.; “Use of Transformed Auxiliary Variable in Estimating the Finite Population Mean”, **Biometrical Journal**, 41, 5, 1999, s. 627–636.

UPADHYAYA, L.N., SINGH, H.P.; “A Note on The Estimation of Mean Using Auxiliary İnformation”, **Statistics in Transition**, 6, 4, 2003, s. 571-575.

VERMA, V.; **Sampling Methods**, Manual for Statistical Trainers Number 2 Statistical Institute for Asia and The Pacific Tokyo, 1998.

YAMANE, T.; **Elementary Sampling Theory,** Engleweed Cliffs, N.J. Prentice-Hall Inc., 1967.

YATES, F.; “The Design of Rotation Experiments, Commonwealth Bureau of Soil Science”, **Technical Communications,** No 46, 1949, s. 142-155.

EK

MONTE CARLO SİMÜLASYONU İÇİN YAZILAN PROGRAMLAR

```
### KİTLENİN PROGRAMA OKUTULMASI ###

N<-19923
tab<-read.table("tab.txt")
names(tab)=c("y","x1","x2","x3","x1y","x2y","x3y","TABAKA")
n<-2500
d<-N/n
f<-n/N
v<-(1-f)/n
w1<-0.5
w2<-0.5
y<-tab$y
x1<-tab$x1
x2<-tab$x2
x3<-tab$x3
x1y<-tab$x1y
x2y<-tab$x2y
x3y<-tab$x3y
kitle<-c(sum(x1),sum(x2),sum(x3))

### DEĞİŞİM KATSAYILARININ HESAPLANMASI ###

CVy<-sqrt(var(y))/mean(y)
CVx1<-sqrt(var(x1))/mean(x1)
CVx2<-sqrt(var(x2))/mean(x2)
CVx3<-sqrt(var(x3))/mean(x3)
CVyx1<-(var(x1,y))/(mean(x1)*mean(y))
CVyx2<-(var(x2,y))/(mean(x2)*mean(y))
CVyx3<-(var(x3,y))/(mean(x3)*mean(y))

### KORELASYONLARIN HESAPLANMASI ###

CORx1y<-var(x1,y)/(sqrt(var(x1))*sqrt(var(y)))
CORx2y<-var(x2,y)/(sqrt(var(x2))*sqrt(var(y)))
CORx3y<-var(x3,y)/(sqrt(var(x3))*sqrt(var(y)))
CORx1x2<-var(x1,x2)/(sqrt(var(x1))*sqrt(var(x2)))
CORx1x3<-var(x1,x3)/(sqrt(var(x1))*sqrt(var(x3)))
CORx2x3<-var(x2,x3)/(sqrt(var(x2))*sqrt(var(x3)))


### BASİT RASGELE ÖRNEKLMEDE MONTE CARLO ###

r<-5000
```

```
### ÖRNEK ÇEKİMİ ###

ornekler<- matrix(0,n,r)
yardim1<-matrix(0,n,r)
yardim2<-matrix(0,n,r)
yardim3<-matrix(0,n,r)
s<-matrix(0,N,r)
for(i in 1:r){
	s[,i]<-srswor(n,N)
	ornekler[,i]<-as.vector(y[s[,i]==1])}
for(i in 1:r){
	yardim1[,i]<-as.vector(x1[s[,i]==1])
	yardim2[,i]<-as.vector(x2[s[,i]==1])
	yardim3[,i]<-as.vector(x3[s[,i]==1])}

### ORTALAMA TAHMİN EDİCİSİ VE VARYANSI ###

HT<-apply(ornekler,2,mean)
YHT<-mean(HT)
BIASYHT<-YHT-mean(y)
VARYHT<-sum((HT-YHT)^2)/(r-1)
MSEYHT<-VARYHT+BIASYHT^2

### KLASİK ORAN TAHMİN, YAN, VARYANS ve HKO ###

RATIOson1<-matrix(0,1,r)
RATIOson2<-matrix(0,1,r)
RATIOson3<-matrix(0,1,r)
yort<-HT

for(m in 1:r){
	RATIOson1[,m]<-(yort[m]/mean(yardim1[,m]))*mean(x1)}
YRATIOx1<-mean(RATIOson1)
BIASYRATIOx1<-YRATIOx1-mean(y)
VARYRATIOx1<-sum((RATIOson1-YRATIOx1)^2)/(r-1)
MSEYRATIOx1<-VARYRATIOx1+BIASYRATIOx1^2

for(o in 1:r){
	RATIOson2[,o]<-(yort[o]/mean(yardim2[,o]))*mean(x2)}
YRATIOx2<-mean(RATIOson2)
BIASYRATIOx2<-YRATIOx2-mean(y)
VARYRATIOx2<-sum((RATIOson2-YRATIOx2)^2)/(r-1)
MSEYRATIOx2<-VARYRATIOx2+BIASYRATIOx2^2

for(p in 1:r){
	RATIOson3[,p]<-(yort[p]/mean(yardim3[,p]))*mean(x3)}
YRATIOx3<-mean(RATIOson3)
BIASYRATIOx3<-YRATIOx3-mean(y)
VARYRATIOx3<-sum((RATIOson3-YRATIOx3)^2)/(r-1)
MSEYRATIOx3<-VARYRATIOx3+BIASYRATIOx3^2
```

```
### BEALE ORAN TAHMİN, YAN, VARYANS ve HKO ###

YBEALEson1<-matrix(0,1,r)
YBEALEson2<-matrix(0,1,r)
YBEALEson3<-matrix(0,1,r)

for(m in 1:r){
	YBEALEson1[,m]<-
(yort[m]/mean(yardim1[,m]))*((1+v*CVyx1)/(1+v*CVx1^2))*mean(x1)}
YBEALEx1<-mean(YBEALEson1)
BIASYBEALEx1<-YBEALEx1-mean(y)
VARYBEALEx1<-sum((YBEALEson1-YBEALEx1)^2)/(r-1)
MSEYBEALEx1<-VARYBEALEx1+BIASYBEALEx1^2

for(o in 1:r){
	YBEALEson2[,o]<-
(yort[o]/mean(yardim2[,o]))*((1+v*CVyx2)/(1+v*CVx2^2))*mean(x2)}
YBEALEx2<-mean(YBEALEson2)
BIASYBEALEx2<-YBEALEx2-mean(y)
VARYBEALEx2<-sum((YBEALEson2-YBEALEx2)^2)/(r-1)
MSEYBEALEx2<-VARYBEALEx2+BIASYBEALEx2^2

for(p in 1:r){
	YBEALEson3[,p]<-
(yort[p]/mean(yardim3[,p]))*((1+v*CVyx3)/(1+v*CVx3^2))*mean(x3)}
YBEALEx3<-mean(YBEALEson3)
BIASYBEALEx3<-YBEALEx3-mean(y)
VARYBEALEx3<-sum((YBEALEson3-YBEALEx3)^2)/(r-1)
MSEYBEALEx3<-VARYBEALEx3+BIASYBEALEx3^2

### TIN ORAN TAHMİN,YAN,VARYANS ve HKO ###

YTINson1<-matrix(0,1,r)
YTINson2<-matrix(0,1,r)
YTINson3<-matrix(0,1,r)

for(m in 1:r){
	YTINson1[,m]<-(yort[m]/mean(yardim1[,m]))*(1+v*(CVyx1-CVx1^2))*mean(x1)}
YTINx1<-mean(YTINson1)
BIASYTINx1<-YTINx1-mean(y)
VARYTINx1<-sum((YTINson1-YTINx1)^2)/(r-1)
MSEYTINx1<-VARYTINx1+BIASYTINx1^2

for(o in 1:r){
	YTINson2[,o]<-(yort[o]/mean(yardim2[,o]))*(1+v*(CVyx2-CVx2^2))*mean(x2)}
YTINx2<-mean(YTINson2)
BIASYTINx2<-YTINx2-mean(y)
VARYTINx2<-sum((YTINson2-YTINx2)^2)/(r-1)
MSEYTINx2<-VARYTINx2+BIASYTINx2^2
```

```
for(p in 1:r){
        YTINson3[,p]<-(yort[p]/mean(yardim3[,p]))*(1+v*(CVyx3-CVx3^2))*mean(x3)}
YTINx3<-mean(YTINson3)
BIASYTINx3<-YTINx3-mean(y)
VARYTINx3<-sum((YTINson3-YTINx3)^2)/(r-1)
MSEYTINx3<-VARYTINx3+BIASYTINx3^2

### SINGH TAILOR ORAN TAHMİN, YAN, VARYANS ve HKO ###

YSINTAson1<-matrix(0,1,r)
YSINTAson2<-matrix(0,1,r)
YSINTAson3<-matrix(0,1,r)

for(m in 1:r){
        YSINTAson1[,m]<-yort[m]*(mean(x1)+CORx1y)/(mean(yardim1[,m])+CORx1y)}
YSINTAx1<-mean(YSINTAson1)
BIASYSINTAx1<-YSINTAx1-mean(y)
VARYSINTAx1<-sum((YSINTAson1-YSINTAx1)^2)/(r-1)
MSEYSINTAx1<-VARYSINTAx1+BIASYSINTAx1^2

for(o in 1:r){
        YSINTAson2[,o]<-yort[o]*(mean(x2)+CORx2y)/(mean(yardim2[,o])+CORx2y)}
YSINTAx2<-mean(YSINTAson2)
BIASYSINTAx2<-YSINTAx2-mean(y)
VARYSINTAx2<-sum((YSINTAson2-YSINTAx2)^2)/(r-1)
MSEYSINTAx2<-VARYSINTAx2+BIASYSINTAx2^2

for(p in 1:r){
        YSINTAson3[,p]<-yort[p]*(mean(x3)+CORx3y)/(mean(yardim3[,p])+CORx3y)}
YSINTAx3<-mean(YSINTAson3)
BIASYSINTAx3<-YSINTAx3-mean(y)
VARYSINTAx3<-sum((YSINTAson3-YSINTAx3)^2)/(r-1)
MSEYSINTAx3<-VARYSINTAx3+BIASYSINTAx3^2

### TEK DEĞİŞKENLİ REGRESYON TAHMİN, YAN, VARYANS ve HKO  ###

YREG1son<-matrix(0,1,r)
YREG2son<-matrix(0,1,r)
YREG3son<-matrix(0,1,r)

for(a in 1:r){
B1<-sum((yardim1[,a]-mean(yardim1[,a]))*(ornekler[,a]-
mean(ornekler[,a])))/sum((yardim1[,a]-mean(yardim1[,a]))^2)
YREG1son[a]<-mean(ornekler[,a])+B1*(mean(x1)-mean(yardim1[,a]))
B2<-sum((yardim2[,a]-mean(yardim2[,a]))*(ornekler[,a]-
mean(ornekler[,a])))/sum((yardim2[,a]-mean(yardim2[,a]))^2)
YREG2son[a]<-mean(ornekler[,a])+B2*(mean(x2)-mean(yardim2[,a]))
B3<-sum((yardim3[,a]-mean(yardim3[,a]))*(ornekler[,a]-
mean(ornekler[,a])))/sum((yardim3[,a]-mean(yardim3[,a]))^2)
YREG3son[a]<-mean(ornekler[,a])+B3*(mean(x3)-mean(yardim3[,a]))
}
```

```
YREG1<-mean(YREG1son)
YREG2<-mean(YREG2son)
YREG3<-mean(YREG3son)

BIASYREG1<-YREG1-mean(y)
VARYREG1<-sum((YREG1son-YREG1)^2)/(r-1)
MSEYREG1<-VARYREG1+BIASYREG1^2

BIASYREG2<-YREG2-mean(y)
VARYREG2<-sum((YREG2son-YREG2)^2)/(r-1)
MSEYREG2<-VARYREG2+BIASYREG2^2

BIASYREG3<-YREG3-mean(y)
VARYREG3<-sum((YREG3son-YREG3)^2)/(r-1)
MSEYREG3<-VARYREG3+BIASYREG3^2

### KAUR REGRESYON TAHMİN,YAN,VARYANS ve HKO  ###

alfa<-(1+v*CORx1y^2*CVy^2)/(1+v*CVy^2)
YKAUR1son<-matrix(0,1,r)
YKAUR2son<-matrix(0,1,r)
YKAUR3son<-matrix(0,1,r)

for(a in 1:r){
B1<-sum((yardim1[,a]-mean(yardim1[,a]))*(ornekler[,a]-
mean(ornekler[,a])))/sum((yardim1[,a]-mean(yardim1[,a]))^2)
YKAUR1son[a]<-alfa*mean(ornekler[,a])+B1*(mean(x1)-mean(yardim1[,a]))
B2<-sum((yardim2[,a]-mean(yardim2[,a]))*(ornekler[,a]-
mean(ornekler[,a])))/sum((yardim2[,a]-mean(yardim2[,a]))^2)
YKAUR2son[a]<-alfa*mean(ornekler[,a])+B2*(mean(x2)-mean(yardim2[,a]))
B3<-sum((yardim3[,a]-mean(yardim3[,a]))*(ornekler[,a]-
mean(ornekler[,a])))/sum((yardim3[,a]-mean(yardim3[,a]))^2)
YKAUR3son[a]<-alfa*mean(ornekler[,a])+B3*(mean(x3)-mean(yardim3[,a]))
}

YKAUR1<-mean(YKAUR1son)
YKAUR2<-mean(YKAUR2son)
YKAUR3<-mean(YKAUR3son)

BIASYKAUR1<-YKAUR1-mean(y)
VARYKAUR1<-sum((YKAUR1son-YKAUR1)^2)/(r-1)
MSEYKAUR1<-VARYKAUR1+BIASYKAUR1^2

BIASYKAUR2<-YKAUR2-mean(y)
VARYKAUR2<-sum((YKAUR2son-YKAUR2)^2)/(r-1)
MSEYKAUR2<-VARYKAUR2+BIASYKAUR2^2

BIASYKAUR3<-YKAUR3-mean(y)
VARYKAUR3<-sum((YKAUR3son-YKAUR3)^2)/(r-1)
MSEYKAUR3<-VARYKAUR3+BIASYKAUR3^2
```

```
### GENELLEŞTİRİLMİŞ REGRESYON TAHMİN, YAN, VARYANS ve HKO  ###

YTOP<-apply(ornekler,2,sum)*d
GREGsonx1x2x3<-matrix(0,1,r)
for(i in 1:r){
	GREGset<-cbind(ornekler[,i],yardim1[,i],yardim2[,i],yardim3[,i])
	xx<-GREGset[,2:4]*d
	orn<-apply(xx,2,sum)
	qq<-kitle-orn
	xxx<-t(GREGset[,2:4])
	zzz<-t(GREGset[,1])
	yy<-matrix(0,3,3)
	zz<-matrix(0,3,1)
	for(j in 1:n){
		yy<-yy+d*xxx[,j]%*%t(xxx[,j])}
		yyy<-solve(yy)
		for(k in 1:n){
			zz<-zz+d*xxx[,k]*zzz[,k]}
		Beta<-yyy%*%zz
GREGsonx1x2x3[i]<-YTOP[i]+t(qq)%*%Beta}
GREGsonx1x2x3<-GREGsonx1x2x3/N
YGREGx1x2x3<-mean(GREGsonx1x2x3)
BIASYGREGx1x2x3<-YGREGx1x2x3-mean(y)
VARYGREGx1x2x3<-sum((GREGsonx1x2x3-YGREGx1x2x3)^2)/(r-1)
MSEYGREGx1x2x3<-VARYGREGx1x2x3+BIASYGREGx1x2x3^2

### GREG TAHMİN, YAN, VARYANS ve HKO  (X2X3) ###

YTOP<-apply(ornekler,2,sum)*d
GREGsonx2x3<-matrix(0,1,r)
for(i in 1:r){
	GREGset<-cbind(ornekler[,i],yardim1[,i],yardim2[,i],yardim3[,i])
	xx<-GREGset[,3:4]*d
	orn<-apply(xx,2,sum)
	qq<-kitle[2:3]-orn
	xxx<-t(GREGset[,3:4])
	zzz<-t(GREGset[,1])
	yy<-matrix(0,2,2)
	zz<-matrix(0,2,1)
	for(j in 1:n){
		yy<-yy+d*xxx[,j]%*%t(xxx[,j])}
		yyy<-solve(yy)
		for(k in 1:n){
			zz<-zz+d*xxx[,k]*zzz[,k]}
		Beta<-yyy%*%zz
GREGsonx2x3[i]<-YTOP[i]+t(qq)%*%Beta}
GREGsonx2x3<-GREGsonx2x3/N
YGREG<-mean(GREGson)
BIASYGREG<-YGREG-mean(y)
VARYGREG<-sum((GREGson-YGREG)^2)/(r-1)
MSEYGREG<-VARYGREG+BIASYGREG^2
```

```
### GREG TAHMİN, YAN, VARYANS ve HKO  (X1X2) ###

YTOP<-apply(ornekler,2,sum)*d
GREGsonx1x2<-matrix(0,1,r)
for(i in 1:r){
        GREGset<-cbind(ornekler[,i],yardim1[,i],yardim2[,i],yardim3[,i])
        xx<-GREGset[,2:3]*d
        orn<-apply(xx,2,sum)
        qq<-kitle[1:2]-orn
        xxx<-t(GREGset[,2:3])
        zzz<-t(GREGset[,1])
        yy<-matrix(0,2,2)
        zz<-matrix(0,2,1)
        for(j in 1:n){
                yy<-yy+d*xxx[,j]%*%t(xxx[,j])}
                yyy<-solve(yy)
                for(k in 1:n){
                        zz<-zz+d*xxx[,k]*zzz[,k]}
                Beta<-yyy%*%zz
GREGsonx1x2[i]<-YTOP[i]+t(qq)%*%Beta}
GREGsonx1x2<-GREGsonx1x2/N
YGREGx1x2<-mean(GREGsonx1x2)
BIASYGREGx1x2<-YGREGx1x2-mean(y)
VARYGREGx1x2<-sum((GREGsonx1x2-YGREGx1x2)^2)/(r-1)
MSEYGREGx1x2<-VARYGREGx1x2+BIASYGREGx1x2^2

### GREG TAHMİN, YAN, VARYANS ve HKO  (X1X3) ###

YTOP<-apply(ornekler,2,sum)*d
GREGsonx1x3<-matrix(0,1,r)
for(i in 1:r){
        GREGset<-cbind(ornekler[,i],yardim1[,i],yardim2[,i],yardim3[,i])
        xx<-cbind(GREGset[,2],GREGset[,4])*d
        orn<-apply(xx,2,sum)
        qq<-cbind(kitle[1],kitle[3])-orn
        xxx<-t(cbind(GREGset[,2],GREGset[,4]))
        zzz<-t(GREGset[,1])
        yy<-matrix(0,2,2)
        zz<-matrix(0,2,1)
        for(j in 1:n){
                yy<-yy+d*xxx[,j]%*%t(xxx[,j])}
                yyy<-solve(yy)
                for(k in 1:n){
                        zz<-zz+d*xxx[,k]*zzz[,k]}
                Beta<-yyy%*%zz
GREGsonx1x3[i]<-YTOP[i]+qq%*%Beta}
GREGsonx1x3<-GREGsonx1x3/N
YGREGx1x3<-mean(GREGsonx1x3)
BIASYGREGx1x3<-YGREGx1x3-mean(y)
VARYGREGx1x3<-sum((GREGsonx1x3-YGREGx1x3)^2)/(r-1)
MSEYGREGx1x3<-VARYGREGx1x3+BIASYGREGx1x3^2
```

```
### TABAKALI RASGELE ÖRNEKLMEDE MONTE CARLO ###

tab<-read.table("tabst.txt")
names(tab)=c("y","x1","x2","x3","x1y","x2y","x3y","TABAKA")
r<-1000
kitle<-c(sum(tab$x1),sum(tab$x2),sum(tab$x3))

### KİTLE BİLGİLERİ ###

AA<-tab[tab$TABAKA=="1",]
BB<-tab[tab$TABAKA=="6",]
CC<-tab[tab$TABAKA=="7",]
DD<-tab[tab$TABAKA=="10",]
EE<-tab[tab$TABAKA=="16",]
FF<-tab[tab$TABAKA=="20",]
GG<-tab[tab$TABAKA=="21",]
HH<-tab[tab$TABAKA=="22",]
JJ<-tab[tab$TABAKA=="25",]
KK<-tab[tab$TABAKA=="27",]
LL<-tab[tab$TABAKA=="31",]
MM<-tab[tab$TABAKA=="34",]
NN<-tab[tab$TABAKA=="35",]
OO<-tab[tab$TABAKA=="36",]
PP<-tab[tab$TABAKA=="37",]
RR<-tab[tab$TABAKA=="38",]
SS<-tab[tab$TABAKA=="41",]
TT<-tab[tab$TABAKA=="42",]
UU<-tab[tab$TABAKA=="44",]
VV<-tab[tab$TABAKA=="45",]
WW<-tab[tab$TABAKA=="50",]
XX<-tab[tab$TABAKA=="55",]
YY<-tab[tab$TABAKA=="56",]
ZZ<-tab[tab$TABAKA=="61",]
QQ<-tab[tab$TABAKA=="65",]
II<-tab[tab$TABAKA=="67",]


### ORTALAMALAR ###
Yi<
rbind(mean(AA$y),mean(BB$y),mean(CC$y),mean(DD$y),mean(EE$y),mean(FF$y),mean(
GG$y),mean(HH$y),mean(JJ$y),mean(KK$y),mean(LL$y),mean(MM$y),mean(NN$y),mea
n(OO$y),mean(PP$y),mean(RR$y),mean(SS$y),mean(TT$y),mean(UU$y),mean(VV$y),me
an(WW$y),mean(XX$y),mean(YY$y),mean(ZZ$y),mean(QQ$y),mean(II$y))
X1i<-
cbind(mean(AA$x1),mean(BB$x1),mean(CC$x1),mean(DD$x1),mean(EE$x1),mean(FF$x1
),mean(GG$x1),mean(HH$x1),mean(JJ$x1),mean(KK$x1),mean(LL$x1),mean(MM$x1),m
ean(NN$x1),mean(OO$x1),mean(PP$x1),mean(RR$x1),mean(SS$x1),mean(TT$x1),mean(
UU$x1),mean(VV$x1),mean(WW$x1),mean(XX$x1),mean(YY$x1),mean(ZZ$x1),mean(Q
Q$x1),mean(II$x1))
X2i<-
cbind(mean(AA$x2),mean(BB$x2),mean(CC$x2),mean(DD$x2),mean(EE$x2),mean(FF$x2
```

```
),mean(GG$x2),mean(HH$x2),mean(JJ$x2),mean(KK$x2),mean(LL$x2),mean(MM$x2),m
ean(NN$x2),mean(OO$x2),mean(PP$x2),mean(RR$x2),mean(SS$x2),mean(TT$x2),mean(
UU$x2),mean(VV$x2),mean(WW$x2),mean(XX$x2),mean(YY$x2),mean(ZZ$x2),mean(Q
Q$x2),mean(II$x2))
X3i<-
cbind(mean(AA$x3),mean(BB$x3),mean(CC$x3),mean(DD$x3),mean(EE$x3),mean(FF$x3
),mean(GG$x3),mean(HH$x3),mean(JJ$x3),mean(KK$x3),mean(LL$x3),mean(MM$x3),m
ean(NN$x3),mean(OO$x3),mean(PP$x3),mean(RR$x3),mean(SS$x3),mean(TT$x3),mean(
UU$x3),mean(VV$x3),mean(WW$x3),mean(XX$x3),mean(YY$x3),mean(ZZ$x3),mean(Q
Q$x3),mean(II$x3))
X1Yi<-
cbind(mean(AA$x1y),mean(BB$x1y),mean(CC$x1y),mean(DD$x1y),mean(EE$x1y),mean(
FF$x1y),mean(GG$x1y),mean(HH$x1y),mean(JJ$x1y),mean(KK$x1y),mean(LL$x1y),mea
n(MM$x1y),mean(NN$x1y),mean(OO$x1y),mean(PP$x1y),mean(RR$x1y),mean(SS$x1y),
mean(TT$x1y),mean(UU$x1y),mean(VV$x1y),mean(WW$x1y),mean(XX$x1y),mean(YY$
x1y),mean(ZZ$x1y),mean(QQ$x1y),mean(II$x1y))
X2Yi<-
cbind(mean(AA$x2y),mean(BB$x2y),mean(CC$x2y),mean(DD$x2y),mean(EE$x2y),mean(
FF$x2y),mean(GG$x2y),mean(HH$x2y),mean(JJ$x2y),mean(KK$x2y),mean(LL$x2y),mea
n(MM$x2y),mean(NN$x2y),mean(OO$x2y),mean(PP$x2y),mean(RR$x2y),mean(SS$x2y),
mean(TT$x2y),mean(UU$x2y),mean(VV$x2y),mean(WW$x2y),mean(XX$x2y),mean(YY$
x2y),mean(ZZ$x2y),mean(QQ$x2y),mean(II$x2y))
X3Yi<-
cbind(mean(AA$x3y),mean(BB$x3y),mean(CC$x3y),mean(DD$x3y),mean(EE$x3y),mean(
FF$x3y),mean(GG$x3y),mean(HH$x3y),mean(JJ$x3y),mean(KK$x3y),mean(LL$x3y),mea
n(MM$x3y),mean(NN$x3y),mean(OO$x3y),mean(PP$x3y),mean(RR$x3y),mean(SS$x3y),
mean(TT$x3y),mean(UU$x3y),mean(VV$x3y),mean(WW$x3y),mean(XX$x3y),mean(YY$
x3y),mean(ZZ$x3y),mean(QQ$x3y),mean(II$x3y))

### VARYANSLAR ###

varYi<-
cbind(var(AA$y),var(BB$y),var(CC$y),var(DD$y),var(EE$y),var(FF$y),var(GG$y),var(HH
$y),var(JJ$y),var(KK$y),var(LL$y),var(MM$y),var(NN$y),var(OO$y),var(PP$y),var(RR$y
),var(SS$y),var(TT$y),var(UU$y),var(VV$y),var(WW$y),var(XX$y),var(YY$y),var(ZZ$y),
var(QQ$y),var(II$y))
varX1i<-
cbind(var(AA$x1),var(BB$x1),var(CC$x1),var(DD$x1),var(EE$x1),var(FF$x1),var(GG$x1
),var(HH$x1),var(JJ$x1),var(KK$x1),var(LL$x1),var(MM$x1),var(NN$x1),var(OO$x1),var
(PP$x1),var(RR$x1),var(SS$x1),var(TT$x1),var(UU$x1),var(VV$x1),var(WW$x1),var(XX
$x1),var(YY$x1),var(ZZ$x1),var(QQ$x1),var(II$x1))
varX2i<-
cbind(var(AA$x2),var(BB$x2),var(CC$x2),var(DD$x2),var(EE$x2),var(FF$x2),var(GG$x2
),var(HH$x2),var(JJ$x2),var(KK$x2),var(LL$x2),var(MM$x2),var(NN$x2),var(OO$x2),var
(PP$x2),var(RR$x2),var(SS$x2),var(TT$x2),var(UU$x2),var(VV$x2),var(WW$x2),var(XX
$x2),var(YY$x2),var(ZZ$x2),var(QQ$x2),var(II$x2))
varX3i<-
cbind(var(AA$x3),var(BB$x3),var(CC$x3),var(DD$x3),var(EE$x3),var(FF$x3),var(GG$x3
),var(HH$x3),var(JJ$x3),var(KK$x3),var(LL$x3),var(MM$x3),var(NN$x3),var(OO$x3),var
(PP$x3),var(RR$x3),var(SS$x3),var(TT$x3),var(UU$x3),var(VV$x3),var(WW$x3),var(XX
$x3),var(YY$x3),var(ZZ$x3),var(QQ$x3),var(II$x3))
```

```
covX1Yi<-
cbind(var(AA$x1,AA$y),var(BB$x1,BB$y),var(CC$x1,CC$y),var(DD$x1,DD$y),var(EE$x
1,EE$y),var(FF$x1,FF$y),var(GG$x1,GG$y),var(HH$x1,HH$y),var(JJ$x1,JJ$y),var(KK$x
1,KK$y),var(LL$x1,LL$y),var(MM$x1,MM$y),var(NN$x1,NN$y),var(OO$x1,OO$y),var(
PP$x1,PP$y),var(RR$x1,RR$y),var(SS$x1,SS$y),var(TT$x1,TT$y),var(UU$x1,UU$y),var(
VV$x1,VV$y),var(WW$x1,WW$y),var(XX$x1,XX$y),var(YY$x1,YY$y),var(ZZ$x1,ZZ$
y),var(QQ$x1,QQ$y),var(II$x1,II$y))
covX2Yi<-
cbind(var(AA$x2,AA$y),var(BB$x2,BB$y),var(CC$x2,CC$y),var(DD$x2,DD$y),var(EE$x
2,EE$y),var(FF$x2,FF$y),var(GG$x2,GG$y),var(HH$x2,HH$y),var(JJ$x2,JJ$y),var(KK$x
2,KK$y),var(LL$x2,LL$y),var(MM$x2,MM$y),var(NN$x2,NN$y),var(OO$x2,OO$y),var(
PP$x2,PP$y),var(RR$x2,RR$y),var(SS$x2,SS$y),var(TT$x2,TT$y),var(UU$x2,UU$y),var(
VV$x2,VV$y),var(WW$x2,WW$y),var(XX$x2,XX$y),var(YY$x2,YY$y),var(ZZ$x2,ZZ$
y),var(QQ$x2,QQ$y),var(II$x2,II$y))
covX3Yi<-
cbind(var(AA$x3,AA$y),var(BB$x3,BB$y),var(CC$x3,CC$y),var(DD$x3,DD$y),var(EE$x
3,EE$y),var(FF$x3,FF$y),var(GG$x3,GG$y),var(HH$x3,HH$y),var(JJ$x3,JJ$y),var(KK$x
3,KK$y),var(LL$x3,LL$y),var(MM$x3,MM$y),var(NN$x3,NN$y),var(OO$x3,OO$y),var(
PP$x3,PP$y),var(RR$x3,RR$y),var(SS$x3,SS$y),var(TT$x3,TT$y),var(UU$x3,UU$y),var(
VV$x3,VV$y),var(WW$x3,WW$y),var(XX$x3,XX$y),var(YY$x3,YY$y),var(ZZ$x3,ZZ$
y),var(QQ$x3,QQ$y),var(II$x3,II$y))

### BASIKLIKLAR ###

kurX1i<-
cbind(kurtosis(AA$x1),kurtosis(BB$x1),kurtosis(CC$x1),kurtosis(DD$x1),kurtosis(EE$x1),
kurtosis(FF$x1),kurtosis(GG$x1),kurtosis(HH$x1),kurtosis(JJ$x1),kurtosis(KK$x1),kurtosis
(LL$x1),kurtosis(MM$x1),kurtosis(NN$x1),kurtosis(OO$x1),kurtosis(PP$x1),kurtosis(RR$
x1),kurtosis(SS$x1),kurtosis(TT$x1),kurtosis(UU$x1),kurtosis(VV$x1),kurtosis(WW$x1),k
urtosis(XX$x1),kurtosis(YY$x1),kurtosis(ZZ$x1),kurtosis(QQ$x1),kurtosis(II$x1))
kurX2i<-
cbind(kurtosis(AA$x2),kurtosis(BB$x2),kurtosis(CC$x2),kurtosis(DD$x2),kurtosis(EE$x2),
kurtosis(FF$x2),kurtosis(GG$x2),kurtosis(HH$x2),kurtosis(JJ$x2),kurtosis(KK$x2),kurtosis
(LL$x2),kurtosis(MM$x2),kurtosis(NN$x2),kurtosis(OO$x2),kurtosis(PP$x2),kurtosis(RR$
x2),kurtosis(SS$x2),kurtosis(TT$x2),kurtosis(UU$x2),kurtosis(VV$x2),kurtosis(WW$x2),k
urtosis(XX$x2),kurtosis(YY$x2),kurtosis(ZZ$x2),kurtosis(QQ$x2),kurtosis(II$x2))
kurX3i<-
cbind(kurtosis(AA$x3),kurtosis(BB$x3),kurtosis(CC$x3),kurtosis(DD$x3),kurtosis(EE$x3),
kurtosis(FF$x3),kurtosis(GG$x3),kurtosis(HH$x3),kurtosis(JJ$x3),kurtosis(KK$x3),kurtosis
(LL$x3),kurtosis(MM$x3),kurtosis(NN$x3),kurtosis(OO$x3),kurtosis(PP$x3),kurtosis(RR$
x3),kurtosis(SS$x3),kurtosis(TT$x3),kurtosis(UU$x3),kurtosis(VV$x3),kurtosis(WW$x3),k
urtosis(XX$x3),kurtosis(YY$x3),kurtosis(ZZ$x3),kurtosis(QQ$x3),kurtosis(II$x3))

### TABAKALI ÖRNEK ÇEKİMİ ####

N<-19923
n<-2500
Ni<-
cbind(452,1147,295,233,1757,708,85,305,29,365,264,8758,1738,13,70,424,1116,503,126,67
1,116,333,24,210,22,159)
ni<-cbind(31,81,9,11,205,32,3,18,2,20,29,951,133,2,2,23,756,20,2,92,3,8,2,16,2,47)
```

```
wi<-Ni/ni
di<-
c(rep(wi[1],ni[1]),rep(wi[2],ni[2]),rep(wi[3],ni[3]),rep(wi[4],ni[4]),rep(wi[5],ni[5]),rep(wi[6]
,ni[6]),rep(wi[7],ni[7]),rep(wi[8],ni[8]),rep(wi[9],ni[9]),rep(wi[10],ni[10]),rep(wi[11],ni[11])
,rep(wi[12],ni[12]),rep(wi[13],ni[13]),rep(wi[14],ni[14]),rep(wi[15],ni[15]),rep(wi[16],ni[16
]),rep(wi[17],ni[17]),rep(wi[18],ni[18]),rep(wi[19],ni[19]),rep(wi[20],ni[20]),rep(wi[21],ni[2
1]),rep(wi[22],ni[22]),rep(wi[23],ni[23]),rep(wi[24],ni[24]),rep(wi[25],ni[25]),rep(wi[26],ni[
26]))
fi<-ni/Ni
lamdai<-(1-fi)/ni

YTOP<-matrix(0,r,1)
YSTRATA<-matrix(0,r,1)
x1STRATA<-matrix(0,r,1)
YCOMSTRA<-matrix(0,r,1)
YSEPSTRA<-matrix(0,r,1)
YSDSTRA<-matrix(0,r,1)
YSKSTRA<-matrix(0,r,1)
YUS1STRA<-matrix(0,r,1)
YUS2STRA<-matrix(0,r,1)
YREGCOMSTRA<-matrix(0,r,1)
YREGSEPSTRA<-matrix(0,r,1)
GREGson<-matrix(0,r,1)
GREGsonx1x2<-matrix(0,r,1)
GREGsonx1x3<-matrix(0,r,1)

for(i in 1:r){

s<-
strata(tab,"TABAKA",size=c(31,81,9,11,205,32,3,18,2,20,29,951,133,2,2,23,756,20,2,92,3,8
,2,16,2,47),method="srswor")
veri<-getdata(tab,s)
names(veri)=c("ornekler","yardim1","yardim2","yardim3","x1y","x2y","x3y","TABAKA","
ID_unit","Prob","Stratum")

A<-veri[veri$TABAKA=="1",]
B<-veri[veri$TABAKA=="6",]
C<-veri[veri$TABAKA=="7",]
D<-veri[veri$TABAKA=="10",]
E<-veri[veri$TABAKA=="16",]
F<-veri[veri$TABAKA=="20",]
G<-veri[veri$TABAKA=="21",]
H<-veri[veri$TABAKA=="22",]
J<-veri[veri$TABAKA=="25",]
K<-veri[veri$TABAKA=="27",]
L<-veri[veri$TABAKA=="31",]
M<-veri[veri$TABAKA=="34",]
NZ<-veri[veri$TABAKA=="35",]
O<-veri[veri$TABAKA=="36",]
P<-veri[veri$TABAKA=="37",]
R<-veri[veri$TABAKA=="38",]
```

```
S<-veri[veri$TABAKA=="41",]
T<-veri[veri$TABAKA=="42",]
U<-veri[veri$TABAKA=="44",]
V<-veri[veri$TABAKA=="45",]
W<-veri[veri$TABAKA=="50",]
X<-veri[veri$TABAKA=="55",]
Y<-veri[veri$TABAKA=="56",]
Z<-veri[veri$TABAKA=="61",]
Q<-veri[veri$TABAKA=="65",]
I<-veri[veri$TABAKA=="67",]

### TABAKALAR İÇİN İSTATİSTİK HESAPLAMALARI ###

yi<-
rbind(mean(A$ornekler),mean(B$ornekler),mean(C$ornekler),mean(D$ornekler),mean(E$or
nekler),mean(F$ornekler),mean(G$ornekler),mean(H$ornekler),mean(J$ornekler),mean(K$o
rnekler),mean(L$ornekler),mean(M$ornekler),mean(NZ$ornekler),mean(O$ornekler),mean(
P$ornekler),mean(R$ornekler),mean(S$ornekler),mean(T$ornekler),mean(U$ornekler),mean
(V$ornekler),mean(W$ornekler),mean(X$ornekler),mean(Y$ornekler),mean(Z$ornekler),me
an(Q$ornekler),mean(I$ornekler))
x1i<-
rbind(mean(A$yardim1),mean(B$yardim1),mean(C$yardim1),mean(D$yardim1),mean(E$y
ardim1),mean(F$yardim1),mean(G$yardim1),mean(H$yardim1),mean(J$yardim1),mean(K$
yardim1),mean(L$yardim1),mean(M$yardim1),mean(NZ$yardim1),mean(O$yardim1),mean
(P$yardim1),mean(R$yardim1),mean(S$yardim1),mean(T$yardim1),mean(U$yardim1),mea
n(V$yardim1),mean(W$yardim1),mean(X$yardim1),mean(Y$yardim1),mean(Z$yardim1),
mean(Q$yardim1),mean(I$yardim1))
x2i<-
rbind(mean(A$yardim2),mean(B$yardim2),mean(C$yardim2),mean(D$yardim2),mean(E$y
ardim2),mean(F$yardim2),mean(G$yardim2),mean(H$yardim2),mean(J$yardim2),mean(K$
yardim2),mean(L$yardim2),mean(M$yardim2),mean(NZ$yardim2),mean(O$yardim2),mean
(P$yardim2),mean(R$yardim2),mean(S$yardim2),mean(T$yardim2),mean(U$yardim2),mea
n(V$yardim2),mean(W$yardim2),mean(X$yardim2),mean(Y$yardim2),mean(Z$yardim2),
mean(Q$yardim2),mean(I$yardim2))
x3i<-
rbind(mean(A$yardim3),mean(B$yardim3),mean(C$yardim3),mean(D$yardim3),mean(E$y
ardim3),mean(F$yardim3),mean(G$yardim3),mean(H$yardim3),mean(J$yardim3),mean(K$
yardim3),mean(L$yardim3),mean(M$yardim3),mean(NZ$yardim3),mean(O$yardim3),mean
(P$yardim3),mean(R$yardim3),mean(S$yardim3),mean(T$yardim3),mean(U$yardim3),mea
n(V$yardim3),mean(W$yardim3),mean(X$yardim3),mean(Y$yardim3),mean(Z$yardim3),
mean(Q$yardim3),mean(I$yardim3))

tabsay<-length(ni)
varyi<-
cbind(var(A$ornekler),var(B$ornekler),var(C$ornekler),var(D$ornekler),var(E$ornekler),var
(F$ornekler),var(G$ornekler),var(H$ornekler),var(J$ornekler),var(K$ornekler),var(L$ornekl
er),var(M$ornekler),var(NZ$ornekler),var(O$ornekler),var(P$ornekler),var(R$ornekler),var(
S$ornekler),var(T$ornekler),var(U$ornekler),var(V$ornekler),var(W$ornekler),var(X$ornek
ler),var(Y$ornekler),var(Z$ornekler),var(Q$ornekler),var(I$ornekler))
varx1i<-
cbind(var(A$yardim1),var(B$yardim1),var(C$yardim1),var(D$yardim1),var(E$yardim1),var
```

```
(F$yardim1),var(G$yardim1),var(H$yardim1),var(J$yardim1),var(K$yardim1),var(L$yardi
m1),var(M$yardim1),var(NZ$yardim1),var(O$yardim1),var(P$yardim1),var(R$yardim1),va
r(S$yardim1),var(T$yardim1),var(U$yardim1),var(V$yardim1),var(W$yardim1),var(X$yard
im1),var(Y$yardim1),var(Z$yardim1),var(Q$yardim1),var(I$yardim1))
varx2i<-
cbind(var(A$yardim2),var(B$yardim2),var(C$yardim2),var(D$yardim2),var(E$yardim2),var
(F$yardim2),var(G$yardim2),var(H$yardim2),var(J$yardim2),var(K$yardim2),var(L$yardi
m2),var(M$yardim2),var(NZ$yardim2),var(O$yardim2),var(P$yardim2),var(R$yardim2),va
r(S$yardim2),var(T$yardim2),var(U$yardim2),var(V$yardim2),var(W$yardim2),var(X$yard
im2),var(Y$yardim2),var(Z$yardim2),var(Q$yardim2),var(I$yardim2))
varx3i<-
cbind(var(A$yardim3),var(B$yardim3),var(C$yardim3),var(D$yardim3),var(E$yardim3),var
(F$yardim3),var(G$yardim3),var(H$yardim3),var(J$yardim3),var(K$yardim3),var(L$yardi
m3),var(M$yardim3),var(NZ$yardim3),var(O$yardim3),var(P$yardim3),var(R$yardim3),va
r(S$yardim3),var(T$yardim3),var(U$yardim3),var(V$yardim3),var(W$yardim3),var(X$yard
im3),var(Y$yardim3),var(Z$yardim3),var(Q$yardim3),var(I$yardim3))
covx1yi<-
cbind(var(A$yardim1,A$ornekler),var(B$yardim1,B$ornekler),var(C$yardim1,C$ornekler),
var(D$yardim1,D$ornekler),var(E$yardim1,E$ornekler),var(F$yardim1,F$ornekler),var(G$
yardim1,G$ornekler),var(H$yardim1,H$ornekler),var(J$yardim1,J$ornekler),var(K$yardim1
,K$ornekler),var(L$yardim1,L$ornekler),var(M$yardim1,M$ornekler),var(NZ$yardim1,NZ
$ornekler),var(O$yardim1,O$ornekler),var(P$yardim1,P$ornekler),var(R$yardim1,R$ornekl
er),var(S$yardim1,S$ornekler),var(T$yardim1,T$ornekler),var(U$yardim1,U$ornekler),var(
V$yardim1,V$ornekler),var(W$yardim1,W$ornekler),var(X$yardim1,X$ornekler),var(Y$ya
rdim1,Y$ornekler),var(Z$yardim1,Z$ornekler),var(Q$yardim1,Q$ornekler),var(I$yardim1,I
$ornekler))
covx2yi<-
cbind(var(A$yardim2,A$ornekler),var(B$yardim2,B$ornekler),var(C$yardim2,C$ornekler),
var(D$yardim2,D$ornekler),var(E$yardim2,E$ornekler),var(F$yardim2,F$ornekler),var(G$
yardim2,G$ornekler),var(H$yardim2,H$ornekler),var(J$yardim2,J$ornekler),var(K$yardim2
,K$ornekler),var(L$yardim2,L$ornekler),var(M$yardim2,M$ornekler),var(NZ$yardim2,NZ
$ornekler),var(O$yardim2,O$ornekler),var(P$yardim2,P$ornekler),var(R$yardim2,R$ornekl
er),var(S$yardim2,S$ornekler),var(T$yardim2,T$ornekler),var(U$yardim2,U$ornekler),var(
V$yardim2,V$ornekler),var(W$yardim2,W$ornekler),var(X$yardim2,X$ornekler),var(Y$ya
rdim2,Y$ornekler),var(Z$yardim2,Z$ornekler),var(Q$yardim2,Q$ornekler),var(I$yardim2,I
$ornekler))
covx3yi<-
cbind(var(A$yardim3,A$ornekler),var(B$yardim3,B$ornekler),var(C$yardim3,C$ornekler),
var(D$yardim3,D$ornekler),var(E$yardim3,E$ornekler),var(F$yardim3,F$ornekler),var(G$
yardim3,G$ornekler),var(H$yardim3,H$ornekler),var(J$yardim3,J$ornekler),var(K$yardim3
,K$ornekler),var(L$yardim3,L$ornekler),var(M$yardim3,M$ornekler),var(NZ$yardim3,NZ
$ornekler),var(O$yardim3,O$ornekler),var(P$yardim3,P$ornekler),var(R$yardim3,R$ornekl
er),var(S$yardim3,S$ornekler),var(T$yardim3,T$ornekler),var(U$yardim3,U$ornekler),var(
V$yardim3,V$ornekler),var(W$yardim3,W$ornekler),var(X$yardim3,X$ornekler),var(Y$ya
rdim3,Y$ornekler),var(Z$yardim3,Z$ornekler),var(Q$yardim3,Q$ornekler),var(I$yardim3,I
$ornekler))

x1STRATA[i]<-(Ni%*%x1i)/N
```

```
### TABAKALI ÖRNEKLEMEDE KLASİK TAHMİN ###

YSTRATA[i]<-(Ni%*%yi)/N

### TABAKALIDA BİLEŞİK ORAN TAHMİN ###

YCOMSTRA[i]<-(YSTRATA[i]/x1STRATA[i])*mean(tab$x1)

### TABAKALI ÖRNEKLEMEDE AYRI ORAN TAHMİN ###

ARAHESAP7<-matrix(0,tabsay,1)
for(j in 1:tabsay){
ARAHESAP7[j,]<-(Ni[j]/N)*yi[j]/x1i[j]*X1i[j]}
TOP7<-sum(ARAHESAP7)
YSEPSTRA[i]<-TOP7

### TABAKALI ÖRNEKLEMEDE SISODIA-DWIVEDI ORAN TAHMİN ###

ARAHESAP10<-matrix(0,tabsay,1)
for(j in 1:tabsay){
ARAHESAP10[j,]<-(Ni[j]/N)*(X1i[j]+(sqrt(varX1i[j])/X1i[j]))}
TOP10<-sum(ARAHESAP10)
ARAHESAP11<-matrix(0,tabsay,1)
for(j in 1:tabsay){
ARAHESAP11[j,]<-(Ni[j]/N)*(x1i[j]+(sqrt(varX1i[j])/X1i[j]))}
TOP11<-sum(ARAHESAP11)
YSDSTRA[i]<-YSTRATA[i]*TOP10/TOP11

### TABAKALI ÖRNEKLEMEDE SINGH-KAKRAN ORAN TAHMİN ###

ARAHESAP14<-matrix(0,tabsay,1)
for(j in 1:tabsay){
ARAHESAP14[j,]<-(Ni[j]/N)*(X1i[j]+kurX1i[j])}
TOP14<-sum(ARAHESAP14)
ARAHESAP15<-matrix(0,tabsay,1)
for(j in 1:tabsay){
ARAHESAP15[j,]<-(Ni[j]/N)*(x1i[j]+kurX1i[j])}
TOP15<-sum(ARAHESAP15)
YSKSTRA[i]<-YSTRATA[i]*TOP14/TOP15

### TABAKALI ÖRNEKLEMEDE UPADHYAYA-SINGH 1 ORAN TAHMİN ###

ARAHESAP18<-matrix(0,tabsay,1)
for(j in 1:tabsay){
ARAHESAP18[j,]<-(Ni[j]/N)*(X1i[j]*kurX1i[j]+(sqrt(varX1i[j])/X1i[j]))}
TOP18<-sum(ARAHESAP18)
ARAHESAP19<-matrix(0,tabsay,1)
for(j in 1:tabsay){
ARAHESAP19[j,]<-(Ni[j]/N)*(x1i[j]*kurX1i[j]+(sqrt(varX1i[j])/X1i[j]))}
TOP19<-sum(ARAHESAP19)
YUS1STRA[i]<-YSTRATA[i]*TOP18/TOP19
```

```
### TABAKALI ÖRNEKLEMEDE UPADHYAYA-SINGH 2 ORAN TAHMİN ###

ARAHESAP22<-matrix(0,tabsay,1)
for(j in 1:tabsay){
ARAHESAP22[j,]<-(Ni[j]/N)*(X1i[j]*(sqrt(varX1i[j])/X1i[j])+kurX1i[j])}
TOP22<-sum(ARAHESAP22)
ARAHESAP23<-matrix(0,tabsay,1)
for(j in 1:tabsay){
ARAHESAP23[j,]<-(Ni[j]/N)*(x1i[j]*(sqrt(varX1i[j])/X1i[j])+kurX1i[j])}
TOP23<-sum(ARAHESAP23)
YUS2STRA[i]<-YSTRATA[i]*TOP22/TOP23

### TABAKALI ÖRNEKLEMEDE BİLEŞİK REGRESYON TAHMİN ###

ARAHESAP24<-matrix(0,tabsay,1)
for(j in 1:tabsay){
ARAHESAP24[j,]<-(Ni[j]/N)^2*lamdai[j]*covx1yi[j]}
TOP24<-sum(ARAHESAP24)
ARAHESAP25<-matrix(0,tabsay,1)
for(j in 1:tabsay){
ARAHESAP25[j,]<-(Ni[j]/N)^2*lamdai[j]*varx1i[j]}
TOP25<-sum(ARAHESAP25)
bc<-TOP24/TOP25
YREGCOMSTRA[i]<-YSTRATA[i]+bc*(mean(tab$x1)-x1STRATA[i])

### TABAKALI ÖRNEKLEMEDE AYRI REGRESYON TAHMİN ###

ARAHESAP26<-matrix(0,tabsay,1)
for(j in 1:tabsay){
ARAHESAP26[j,]<-(Ni[j]/N)*(yi[j]+covx1yi[j]/varx1i[j]*(X1i[j]-x1i[j]))}
TOP26<-sum(ARAHESAP26)
YREGSEPSTRA[i]<-TOP26

### TABAKALI ÖRNEKLEMEDE GENEL REGRESYON TAHMİN ###

YTOP[i]<-sum(di*veri$ornekler)
GREGset<-cbind(veri$ornekler,veri$yardim1,veri$yardim2,veri$yardim3)
xx<-GREGset[,2:4]*di
orn<-apply(xx,2,sum)
qq<-kitle-orn
xxx<-t(GREGset[,2:4])
zzz<-t(GREGset[,1])
yy<-matrix(0,3,3)
zz<-matrix(0,3,1)
        for(j in 1:n){
                yy<-yy+di[j]*xxx[,j]%*%t(xxx[,j])}
                yyy<-solve(yy)
                for(k in 1:n){
                        zz<-zz+di[k]*xxx[,k]*zzz[,k]}
                Beta<-yyy%*%zz
```

```
GREG<-YTOP[i]+t(qq)%*%Beta
GREGson[i]<-GREG/N

### TABAKALI ÖRNEKLEMEDE GENEL REGRESYON TAHMİN (X1X2) ###

YTOP[i]<-sum(di*veri$ornekler)
GREGset<-cbind(veri$ornekler,veri$yardim1,veri$yardim2,veri$yardim3)
xx<-GREGset[,2:3]*di
orn<-apply(xx,2,sum)
qq<-kitle[1:2]-orn
xxx<-t(GREGset[,2:3])
zzz<-t(GREGset[,1])
yy<-matrix(0,2,2)
zz<-matrix(0,2,1)
        for(j in 1:n){
                yy<-yy+di[j]*xxx[,j]%*%t(xxx[,j])}
                yyy<-solve(yy)
                for(k in 1:n){
                        zz<-zz+di[k]*xxx[,k]*zzz[,k]}
                Beta<-yyy%*%zz
GREGx1x2<-YTOP[i]+t(qq)%*%Beta
GREGsonx1x2[i]<-GREGx1x2/N

### TABAKALI ÖRNEKLEMEDE GENEL REGRESYON TAHMİN (X1X3) ###

YTOP[i]<-sum(di*veri$ornekler)
GREGset<-cbind(veri$ornekler,veri$yardim1,veri$yardim2,veri$yardim3)
xx<-cbind(GREGset[,2],GREGset[,4])*di
orn<-apply(xx,2,sum)
qq<-cbind(kitle[1],kitle[3])-orn
xxx<-t(cbind(GREGset[,2],GREGset[,4]))
zzz<-t(GREGset[,1])
yy<-matrix(0,2,2)
zz<-matrix(0,2,1)
        for(j in 1:n){
                yy<-yy+di[j]*xxx[,j]%*%t(xxx[,j])}
                yyy<-solve(yy)
                for(k in 1:n){
                        zz<-zz+di[k]*xxx[,k]*zzz[,k]}
                Beta<-yyy%*%zz
GREGx1x3<-YTOP[i]+qq%*%Beta
GREGsonx1x3[i]<-GREGx1x3/N

}
```

```
### TAB. ÖRNEKLEMEDE TAHMİNLER, YANLAR, VARYANSLAR ve HKO  ###

YST<-mean(YSTRATA)
BIASYST<-YST-mean(tab$y)
VARYST<-sum((YSTRATA-YST)^2)/(r-1)
MSEYST<-VARYST+BIASYST^2

YCOMST<-mean(YCOMSTRA)
BIASYCOMST<-YCOMST-mean(tab$y)
VARYCOMST<-sum((YCOMSTRA-YCOMST)^2)/(r-1)
MSEYCOMST<-VARYCOMST+BIASYCOMST^2

YSEPST<-mean(YSEPSTRA)
BIASYSEPST<-YSEPST-mean(tab$y)
VARYSEPST<-sum((YSEPSTRA-YSEPST)^2)/(r-1)
MSEYSEPST<-VARYSEPST+BIASYSEPST^2

YSDST<-mean(YSDSTRA)
BIASYSDST<-YSDST-mean(tab$y)
VARYSDST<-sum((YSDSTRA-YSDST)^2)/(r-1)
MSEYSDST<-VARYSDST+BIASYSDST^2

YSKST<-mean(YSKSTRA)
BIASYSKST<-YSKST-mean(tab$y)
VARYSKST<-sum((YSKSTRA-YSKST)^2)/(r-1)
MSEYSKST<-VARYSKST+BIASYSKST^2

YUS1ST<-mean(YUS1STRA)
BIASYUS1ST<-YUS1ST-mean(tab$y)
VARYUS1ST<-sum((YUS1STRA-YUS1ST)^2)/(r-1)
MSEYUS1ST<-VARYUS1ST+BIASYUS1ST^2

YUS2ST<-mean(YUS2STRA)
BIASYUS2ST<-YUS2ST-mean(tab$y)
VARYUS2ST<-sum((YUS2STRA-YUS2ST)^2)/(r-1)
MSEYUS2ST<-VARYUS2ST+BIASYUS2ST^2

YREGCOMST<-mean(YREGCOMSTRA)
BIASYREGCOMST<-YREGCOMST-mean(tab$y)
VARYREGCOMST<-sum((YREGCOMSTRA-YREGCOMST)^2)/(r-1)
MSEYREGCOMST<-VARYREGCOMST+BIASYREGCOMST^2

YREGSEPST<-mean(YREGSEPSTRA)
BIASYREGSEPST<-YREGSEPST-mean(tab$y)
VARYREGSEPST<-sum((YREGSEPSTRA-YREGSEPST)^2)/(r-1)
MSEYREGSEPST<-VARYREGSEPST+BIASYREGSEPST^2

YGREG<-mean(GREGson)
BIASYGREG<-YGREG-mean(tab$y)
VARYGREG<-sum((GREGson-YGREG)^2)/(r-1)
MSEYGREG<-VARYGREG+BIASYGREG^2
```

```
YGREGx1x2<-mean(GREGsonx1x2)
BIASYGREG x1x2<-YGREG x1x2-mean(tab$y)
VARYGREG x1x2<-sum((GREGson x1x2-YGREG x1x2)^2)/(r-1)
MSEYGREG x1x2<-VARYGREG x1x2+BIASYGREG x1x2^2

YGREGx1x3<-mean(GREGsonx1x3)
BIASYGREG x1x3<-YGREG x1x3-mean(tab$y)
VARYGREG x1x3<-sum((GREGson x1x3-YGREG x1x3)^2)/(r-1)
MSEYGREG x1x3<-VARYGREG x1x3+BIASYGREG x1x3^2
```

ÖZET

METİN, Cenker Burak. Örnekleme Yöntemlerinde Yardımcı Değişkenlerin Kullanımı ve Regresyon Tahmin Edicisi ile Bir Uygulama, Yüksek Lisans Tezi, Ankara, 2010.

Bu tez çalışmasının amacı; örnek araştırması sonrasındaki tahmin aşamalarında, daha iyi ve daha etkin sonuçlara ulaşmak için yardımcı değişken kullanımının nasıl bir etki yaptığını incelemektir. Bu amaç doğrultusunda, basit rasgele örnekleme ve tabakalı rasgele örnekleme yöntemlerinde kitle ortalamasının tahmini için yardımcı değişkenlerden faydalanan, çeşitli oran ve regresyon tahmin edicileri ile bu tahmin edicilerin yan ve hata kareler ortalamaları incelenmiştir. Ayrıca farklı örneklem tasarımları için de kullanılabilen Genelleştirilmiş Regresyon (GREG) Tahmin Edicisi de ele alınmıştır. Karşılaştırılabilir olan tahmin edicilerin hangi koşullarda birbirlerine tercih edilmesi gerektiği irdelenmiştir.

Uygulamada, 2006 yılında İmalat Sanayi'nde 20+ çalışanı olan girişimlerin ortalama cirosu; 2005 Yıllık Sanayi ve Hizmet İstatistiklerinden alınan yardımcı değişkenler kullanılarak bazı oran tahmin edicileri, regresyon tahmin edicileri ve GREG Tahmin Edicisi ile tahmin edilmiştir. Basit rasgele örneklemede en iyi sonucu GREG Tahmin Edicisi vermiştir. Tabakalı rasgele örneklemede hata kareler ortalaması hesaplanabilen tahmin ediciler arasından en iyi sonucu Ayrı Regresyon Tahmin Edicisi vermiştir. Hata kareler ortalamasının kesin olarak hesaplanamadığı durumlar için Monte Carlo simülasyon sonuçlarından faydalanabileceği gösterilmiştir.

Anahtar Sözcükler:

1. Yardımcı Değişken
2. Oran Tahmin Edicisi
3. Regresyon Tahmin Edicisi
4. Genelleştirilmiş Regresyon (GREG) Tahmin Edicisi
5. Monte Carlo Simülasyonu

ABSTRACT

METİN, Cenker Burak. Using Auxiliary Variables in Sampling Methods and An Application with Regression Estimator, Master Thesis, Ankara, 2010.

The aim of this thesis is to study the effect of using auxiliary variable to obtain more accurate and more efficient results in estimation stages of survey sampling. In this sense; various ratio and regression estimators, which use auxiliary variables, their biases and mean square errors are investigated for the estimation of population mean in simple random sampling and stratified random sampling. Moreover, Generalized Regression (GREG) Estimator, which can be used in different sampling design, is considered. For comparable estimators, in which conditions which of them should be chosen is analysed.

In application, using auxiliary variables taken from 2005 Annual Industry and Service Statistics average turnover of enterprises which work in Manifacturing Industry and have 20+ employees at 2006 are estimated by various ratio, regression estimators and GREG Estimator. In simple random sampling GREG Estimator gives the best result. In stratified random sampling Seperate Regression Estimator gives the best result through the estimators that mean square errors can be evaluated. It is shown that Monte Carlo simulation results may be taken when the mean square errors can not be evaluated exactly.

Key Words:

1. Auxiliary Variable
2. Ratio Estimator
3. Regression Estimator
4. Generalized Regression (GREG) Estimator
5. Monte Carlo Simulation