

**BAĞ BİLGİSİ OLDUĞUNDA SIRALI KÜME
ÖRNEKLEMESİNDE YENİ TAHMİN EDİCİLER**

**NEW ESTIMATORS IN RANKED SET SAMPLING IN
THE PRESENCE OF TIE INFORMATION**

EDA GİZEM KOÇYİĞİT

PROF. DR. CEM KADILAR

Tez Danışmanı

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

İstatistik Anabilim Dalı için Öngördüğü

DOKTORA TEZİ olarak hazırlanmıştır.

ÖZET

BAĞ BİLGİSİ OLDUĞUNDA SIRALI KÜME ÖRNEKLEMESİNDE YENİ TAHMİN EDİCİLER

EDA GİZEM KOÇYİĞİT

Doktora, İstatistik Bölümü

Tez Danışmanı: Prof. Dr. Cem KADILAR

Şubat 2021, 67 sayfa

Sıralı küme örnekleme, basit rastgele örnekleme alternatif olarak geliştirilmiş sıklıkla kullanılan bir örnekleme yöntemidir. Örnekleme yönteminde sıralayıcılardan düşük bir güvenle bile küme içindeki birimleri doğru bir şekilde sıralaması beklenir. Ayrıca kümeler içinde iki veya daha fazla sayıda birbirine çok benzer ya da aynı birim var ise bu durum sıralamayı daha da zorlaştırır ve kümedeki birimlerin yanlış sıralanmasına neden olur. Tez çalışmasında, sıralı küme örneklemesinde yardımcı değişken yardımıyla sıralama yapılırken oluşan sıralama hatasının bağ bilgisi kullanılarak azaltıldığı bir yöntem (SKÖ-b) altında, ortalama tahmin edicileri incelenmektedir. Çalışma, sıralamada bağ olması durumunda kullanılan yöntemi incelemeyi ve bağları sıralama hatasını en aza indirerek daha güvenilir sonuçlar elde etmek üzere ortalama tahminlerinde kullanabilmeyi amaçlamaktadır. Yöntem ve literatürdeki tahmin ediciler incelenmiş, yöntem için oransal tahmin edicilerin kullanılmadığı görülmüştür. Dolayısıyla tez çalışmasında modifiye edilmiş yeni oransal

tahmin ediciler önerilmiştir. Tahmin edicilerin etkinlikleri önce simülasyonla türetilen büyük ve küçük kitleler üzerinde hesaplanmış daha sonra yapılan simülasyon çalışmaları önerilen oransal tahmin edicilerin literatürde yer alan diğer tahmin edicilerden daha etkin olduğu gösterilmiştir. Ayrıca son zamanlarda ortaya çıkan COVID-19 salgınına ait gerçek veri setlerinde teşhis edilmiş hasta sayısı ve ölen hasta sayısı değişkenleri incelendiğinde, verilerin bağ bilgisi yapısına uygun olduğu görülmüştür. Simülasyon sonuçlarına benzer şekilde gerçek veriler üzerinde de önerilen oransal tahmin ediciler daha iyi sonuçlar vermiştir.

Anahtar Kelimeler: Sıralı küme örnekleme, bağ bilgisi, oransal tahmin ediciler, ortalama tahmini, izotonik tahmin edici.

ABSTRACT

NEW ESTIMATORS IN RANKED SET SAMPLING IN THE PRESENCE OF TIE INFORMATION

EDA GİZEM KOÇYİĞİT

Doctor of Philosophy, Department of Statistics

Supervisor: Prof. Dr. Cem KADILAR

February 2021, 67 pages

Ranked set sampling is a frequently used sampling method developed as an alternative to the simple random sampling. In this sampling method, rankers are expected to rank the units within the set correctly, even with low confidence. Also, if there are two or more very similar or identical units in a set, this makes ranking more difficult and it causes the units in the set to be ranked incorrectly. In this thesis, the mean estimators are examined under a method (RSS-t) in which the ranking error, occurred while ranking with the aid of the auxiliary variable, is reduced by using the tie information under the Ranked Set Sampling. The study aims to examine the ties in the ranking and to use these ties in the population mean estimators for more reliable estimates by minimizing the ranking error. After examining the method and the estimators in the literature, it is seen that the ratio estimators have not been examined under this method and therefore new modified ratio estimators are proposed in this thesis study. The effectiveness of the estimators is first calculated for the samples drawn from large and

small populations derived from simulation. Simulation studies have shown that the proposed estimators are more effective than other estimators in the literature. In addition, when the variables of the number of diagnosed patients and the number of patients who died are examined in a real data set of the recently emerging COVID-19 epidemic, it is seen that the data set is suitable for the tie information structure. Similar to the simulation results, we can also see from the real data set that the proposed estimators give better results.

Keywords: Ranked set sampling, tie information, ratio estimators, mean estimation, isotonic estimators.

TEŐEKKÜR

Doktora alıőmam sırasında ihtiyacım olan her anda bana yol gsteren, yeni fikirlere teővik eden ve gven veren deęerli danıőmanım Sayın Prof. Dr. Cem KADILAR hocama, yksek lisans danıőmanım ve rnekleme yntemlerini ğreten ok kıymetli hocam Sayın Prof. Dr. Hlyla INGI' ya, lisans dzeyinden itibaren bana yardımcı olan ve doktora srecimde de benden desteęini ve fikirlerini esirgemeyen Prof. Dr. Sevil BACANLI hocama, tez izleme srecimde bilgi ve deęerlendirmeleri ile nemli katkıda bulunan jri yem Prof. Dr. Yaprak Arzu ZDEMİR' e, istatistięe dair her Őeyi ğreten, gcn ve desteęini her zaman arkamda hissettięim, Hacettepe niversitesi İstatistik Blm' ne ve oradaki kıymetli hocalarıma, her trl yardım ve desteęinden dolayı Prof. Dr. Aylin ALIN' a ve inanlarını benden eksik etmeyen ve zerimde en byk pay sahibi olan anne ve babama teőekkr bor bilirim.

Eda Gizem KOYIęIT

Őubat 2021, İzmir.

İÇİNDEKİLER

ÖZET	iii
ABSTRACT	iii
TEŞEKKÜR	v
İÇİNDEKİLER	vi
ŞEKİLLER DİZİNİ	viii
ÇİZELGELER DİZİNİ	ix
SİMGELER VE KISALTMALAR	x
1. GİRİŞ	1
2. GENEL BİLGİLER	4
2.1. Sıralı İstatistikler	4
2.2. SKÖ Yöntemi ve Amacı	6
2.3. Sıralı Küme Örneklemesinde Bağ Bilgisi Olması Durumu	8
3. TAHMİN EDİCİLER	11
3.1. Tahmin Edicilerin Özellikleri	11
3.2. Sıralı Küme Örneklemesinde Oransal Tahmin Edicisi	15
3.3. SKÖ-b İçin Literatürde Yer Alan Tahmin Ediciler	19
3.3.1. Bağ Matrisi İçeren Basit Tahmin Edici	19
3.3.2. Basit İzotonik Tahmin Edicileri	20
3.3.3. Rao- Blackwell Tahmin Edicileri	21
3.4. Önerilen Tahmin Ediciler	22
3.4.1. İzotonik Basit Oransal Tahmin Edici	22
3.4.2. İzotonik Üstel Oransal Tahmin Edici	23
3.4.3. Genelleştirilmiş İzotonik Oransal Tahmin Edici	24
3.5. Sayısal Örnek	24

4. SİMÜLASYON	29
4.1. Simülasyonla Veri Setinde Bağ Üretme Yöntemleri.....	29
4.2. Simülasyon Çalışması.....	31
5. GERÇEK VERİ UYGULAMASI.....	41
6. SONUÇ VE TARTIŞMA	46
7. KAYNAKLAR	48
EKLER	53
EK 1 – Çin'e Ait COVID-19 Veri Seti	53
EK 2 – Dünya Geneli COVID-19 Veri Seti.....	54
EK 3 – Tezden Üretilen Yayın.....	58
EK 4 – Simülasyonda Kullanılan R Program Kodları.....	59
EK 5 – Tez Orijinallik Raporu	66
ÖZGEÇMİŞ	67

ŞEKİLLER DİZİNİ

Şekil 4.1. Büyük bir kitlede korelasyon katsayısının e 'ye göre değişimleri (N=1000). 31	31
Şekil 4.2. Küçük bir kitlede korelasyon katsayısının e 'ye göre değişimleri (N=50) 31	31
Şekil 4.3. N(5,1) ve $\rho=0.75$ iken farklı e değerleri için X histogramları. 32	32
Şekil 4.4.N(5,1) ve $\rho=0.75$ iken farklı e değerleri için X-Y saçılım grafikleri.....32	32
Şekil 4.5. N(5,1) ve $\rho=0.9$ iken farklı e değerleri için X-Y saçılım grafikleri..... 32	32
Şekil 4.6. Göreli etkinliğin korelasyon katsayısına, küme büyüklüğüne ve e değerine göre değişimleri. 34	34
Şekil 5.1. COVID-19 verisindeki yardımcı değişken histogram grafiği. 41	41
Şekil 5.2. Dünya geneli COVID-19 verisindeki yardımcı değişken histogram grafiği...43	43

ÇİZELGELER DİZİNİ

Çizelge 2.1. Sıralı küme örnekleme seçimi süreci	7
Çizelge 3.1. COVID-19 verileri için kitle parametreleri	25
Çizelge 3.2. Örnekleme birimlerinin sıralanışı	25
Çizelge 3.3. Örnekleme birimlerinin sıralanışı çizelgesindeki bağların kırılabilmesi için tüm farklı durumlar	27
Çizelge 4.1. Göreceli etkinlik bilgileri	34
Çizelge 4.2. $\mu=5$, $\sigma^2=1$ parametrelili ve $N=1000$ olan Normal dağılımlı kitle için göreceli etkinlik değerleri	35
Çizelge 4.3. $\mu=5$, $\sigma^2=1$ parametrelili ve $N=50$ olan Normal dağılımlı kitle için göreceli etkinlik değerleri	36
Çizelge 4.4. $\mu=5$, $\sigma^2=1$ parametrelili ve $N=50$ olan Normal dağılımlı kitle için yan değerleri	37
Çizelge 4.5. $\mu=5$, $\sigma^2=1$ parametrelili ve $N=1000$ olan Normal dağılımlı kitle için yan değerleri	38
Çizelge 5.1. Çin'e ait COVID-19 verileri için göreceli etkinlik değerleri ($N=34$, $\rho=0.84$)	42
Çizelge 5.2. Çin'e ait COVID-19 verileri için yan değerleri ($N=34$, $\rho=0.84$)	42
Çizelge 5.3. Dünya genelinde COVID-19 verileri için kitle parametreleri	43
Çizelge 5.4. Dünya genelinde COVID-19 veri seti için göreceli etkinlik değerleri ($N=212$, $\rho=0.92$)	44
Çizelge 5.5. Dünya genelinde COVID-19 veri seti için yan değerleri ($N=212$, $\rho=0.92$)	44

SİMGELER VE KISALTMALAR

Simgeler

X_T	Bağ bilgisi olan yardımcı değişken X
ρ	Kitle ilişki Katsayısı
e	Kesikli Seçilen Boyut (DPS) yöntem parametresi
σ^2	Kitle varyansı
μ	Kitle ortalaması
$\hat{\mu}$	Kitle ortalaması için tahmin edici

Kısaltmalar

BRÖ	Basit Rastgele Örnekleme
SKÖ	Sıralı Küme Örnekleme
SKÖ-b	Sıralı Küme Örneklemesinde Bağ Bilgisi
DPS	Kesikli Seçilen Boyut Yöntemi
TIC	Yakınsa Bağlı Yöntemi
GE	Görel Etkinlik
B	Yan

1. GİRİŞ

Bilimsel arařtırmalarda, temel istatistiksel ilkeler hayati rol oynar. İlkelerden en önemlisi arařtırılan konu hakkında geçerli yargılarda bulunabilmek için öncelikle doğru verileri elde edebilmektir. İnsanların mantık çerçevesinde karar verebilmek için ve veri elde edebilmek için kullandıkları yöntemlerden biri örnekleme yöntemidir. Günümüzde fen, mühendislik, sađlık bilimleri, sosyal bilimler gibi pek çok alanda yapılan arařtırmalarda, kamuoyu ve pazara arařtırmalarında hatta günlük yařantıda bile örnekleme yöntemlerinden yararlanılır (Çingı, 2009).

İstatistiklerde kullanılacak verilerin elde edilmesi genellikle maliyet ve zaman gibi kısıtlı imkânlarla bađlıdır. Tam sayım arařtırmaları ise pahalı ve uzun süren çalışmalar olmakla birlikte çođu zaman ilgilenilen kitlenin tamamına ulaşmak çok zor hatta imkânsız olabilir. Kitlenin tamamına ulaşmak yerine; daha az insan gücü, daha düşük maliyet ve kısa bir zaman içerisinde güvenilir verilerin toplanabilmesi nedeniyle örnekleme tercih edilir (Semiz, 2007).

Kitlenin özelliklerini temsil edebilecek nitelikteki daha az birimin oluşturduđu alt gruba örneklem, kitleden örneklem seçme işine ise örnekleme adı verilir. Örneklemde elde edilen bilgiler ışığında kitle hakkında tahminler yapılır. Örneklem seçmek için kullanılan yöntemlere ise örnekleme yöntemleri denir (Çingı, 2009).

Örnekleme kuramında amaç kitlenin yapısına en uygun yöntemi kullanarak örnekleme seçmek ve kullanılan yöntem sürecine göre tahminler yapmaktır. Örnekleme çalışmalarında ana problem ortalama, varyans ve oran gibi parametreleri tahmin etmektir (Yamane, 2001). Kitlelerin çok farklı yapılara sahip olması nedeniyle her farklı kitle yapısı ve tahmin edilecek konu için farklı örnekleme yöntemleri geliştirilmiştir. Kullanılabilecek en basit yöntem basit rastgele örnekleme (BRÖ) yöntemidir.

BRÖ yönteminde sonlu büyüklükteki bir kitleden seçilebilecek tüm örneklemelere eşit seçilme şansı verilerek ve seçilen birim yerine konulmaksızın ya da koyularak n büyüklüğünde örneklem seçilir (Karakulah, 2006). Sıralı Küme Örneklemesi (SKÖ) ise BRÖ' ye alternatif olarak geliştirilen örnekleme yöntemidir.

SKÖ yöntemi ilk kez 1952 yılında McIntyre tarafından Avustralya'daki meraların ortalama ürün miktarını tahmin etmek için kullanılmıştır ve SKÖ yönteminin kitle ortalamasının tahmininde BRÖ' ye göre daha duyarlı bir yöntem olduğu görülmüştür. McIntyre'nin çalışması daha sonra tarım (Coby ve diğ., 1985; Kvam, 2003, Halls ve Dell, 1966), biyometri (Samawi ve diğ., 2001, Mahdizadeh ve Zamanzade 2019) ve tıp (Chen ve diğerleri 2005) gibi pek çok alanda uygulamaları motive etmiştir. Patil (1995), çevresel izlemede çeşitli SKÖ uygulamaları önermiştir. Daha sonra birçok çalışma bu alana odaklanmıştır (Nussbaum ve Sinha, 1997; Kvam, 2003; Öztürk ve ark., 2005) ve SKÖ' nün BRÖ' ye göre avantajları olduğu gösterilmiştir. SKÖ 'de literatür taraması üzerine yapılmış en önemli ve ayrıntılı çalışma Patil ve diğerleri tarafından yapılmıştır. SKÖ parametrik ve parametrik olmayan pek çok istatistiksel yöntem için duyarlı parametre tahminlerinin yapılması amacıyla kullanılmaya devam etmiştir (Zamanzade, 2018; Özdemir, 2005; Çıngı, 2009; Frey,2013).

Yöntemin teorik incelemelerinde verilerdeki sıralama hatası göz ardı edildiğinde, SKÖ'de kitle ortalaması tahmin edicisinin yansız olduğu (Takahaski ve Wahimoto, 1968), sıralama hatası da göz önüne alarak SKÖ'de kitle ortalaması tahmin edicisinin yansız olduğu ve varyansının, rastgele örneklemenin varyansından, daha küçük olduğu gösterilmiştir (Dell ve Clutter, 1972). SKÖ' deki kitle varyansı için tahmin ediciler önerilmiş ve tahmin ediciler sıralama hatası olması durumunda da incelemiştir (Stokes, 1980).

2012 yılında Frey tarafından ortalama tahmini için, 2018 yılında Zamanzade ve Wang tarafından oran tahmini için bağ bilgisi olması durumunda çeşitli tahmin ediciler önerilmiştir. Bağ bilgisi olması durumunda parametrik olmayan dağılım fonksiyonu kestiricileri çalışılmış (Zamanzade, 2020a), biyoistatistik alanında ROC eğrisinin altında kalan alan incelenmiştir (Zamanzade, 2020b).

Tez çalışmasında amaç, SKÖ' de bağ bilgisi olması durumunu incelemek ve bu bağları daha güvenilir sonuçlar elde etmek üzere ortalama tahminlerinde kullanabilmektir. Bölümlerde sırasıyla ihtiyaç duyulabilecek genel bilgiler verildikten sonra yöntem tanıtılmış, literatürde yer alan ortalama tahmin edicileri verilip önerilen tahmin ediciler sunulduktan sonra simülasyon ve gerçek veriler üzerinden hesaplamalar yapılarak önerilen tahmin edicilerin performansı incelenmiştir.

Genel Bilgiler bölümünde, sıralı küme örnekleme yöntemi ve yönteme ilişkin detaylı bilgiler verilmiştir. Daha sonra SKÖ' de bağ durumunun nasıl oluştuğu ve bağların nasıl kaydedildiği gösterilmiştir.

Tahmin ediciler bölümünde tahmin edicilerin özellikleri, sağlaması beklenen özellikler açıklandıktan ve simülasyon çalışması sırasında kullanılan tanımlamalar yapıldıktan sonra literatürde yer alan SKÖ' de bağ bilgisi olması durumu için tahmin ediciler verilmiş ve önerilen tahmin ediciler tanıtılmıştır.

Dördüncü bölümde simülasyon çalışmasıyla ilgili bilgiler verilip, literatürde yer alan ve önerilen tahmin ediciler etkinlikleri bakımından karşılaştırılmış, Beşinci Bölümde güncel veri setleri üzerinde uygulama yapılmıştır. Son bölümde ise Dördüncü ve Beşinci Bölümlerde elde edilen sonuçların genel yorumu yapılmıştır.

2. GENEL BİLGİLER

Bölümde, SKÖ yöntemine ilişkin gerekli bilgiler ve bağ durumundaki SKÖ ile ilgili temel bilgiler verilecektir.

2.1. Sıralı İstatistikler

Sıralı istatistikler istatistik teorisinde önemli konularından biridir ve SKÖ yöntemi için de büyük önem taşımaktadır. Son zamanlarda çıkarsamalı istatistik kuramında da sıralı istatistiklerin önemli olduğu görülmüştür. Çünkü sıralı istatistiklerin dağılımı örneklemelerin çekildiği dağılımlardan bağımsızdır (İnal ve Günay, 2009).

$a < x < b$ olmak üzere $f(x)$ olasılık yoğunluk fonksiyonuna sahip kitleden seçilmiş n büyüklüğündeki rastgele örneklem X_1, X_2, \dots, X_n ya da başka bir gösterimle $X_i, i = 1, 2, \dots, n$ olsun. Rastgele örneklemin küçükten büyüğe sıralanmış biçimi $Y_1 < Y_2 < \dots < Y_n$ 'dir. Bu durumda Y_i için, rastgele örneklemin i -inci sıralı istatistiği denir ve

$$Y_1 = \min\{X_1, X_2, \dots, X_n\}, Y_2 = \min\{\{X_1, X_2, \dots, X_n\} - Y_1\}, \dots, Y_n = \max\{X_1, X_2, \dots, X_n\} \quad (1)$$

olmuş olur (Haki, 2012).

Y_1, \dots, Y_n sıralı istatistiklerinin bileşik olasılık yoğunluk fonksiyonu

$$f(y_1, \dots, y_n) = n! f(y_1) \dots f(y_n), \quad a < y_1 < \dots < y_n < b \quad (2)$$

$= 0$, diğer y_1, \dots, y_n değerleri için

olarak yazılır. $F(x)$ dağılım fonksiyonunu göstermek üzere, en küçük sıralı istatistiğe ait olasılık yoğunluk fonksiyonu,

$$\begin{aligned} f_{Y_1}(y_1) &= n[1-F(y_1)]^{n-1} f(y_1), \quad a < y_1 < b \\ &= 0 \quad , \text{ diğ}er \ y_1 \ \text{değ}erleri için \end{aligned} \quad (3)$$

en büyük sıralı istatistiğe ait olasılık yoğunluk fonksiyonu,

$$\begin{aligned} f_{Y_n}(y_n) &= n[F(y_n)]^{n-1} f(y_n), \quad a < y_n < b \\ &= 0 \quad , \text{ diğ}er \ y_n \ \text{değ}erleri için \end{aligned} \quad (4)$$

ve genel olarak k -ıncı sıralı istatistiğın olasılık yoğunluk fonksiyonu,

$$\begin{aligned} f_{Y_k}(y_k) &= \frac{n!}{(k-1)!(n-k)!} [F(y_k)]^{k-1} [1-F(y_k)]^{n-k} f(y_k), \quad a < y_k < b \\ &= 0 \quad , \text{ diğ}er \ y_k \ \text{değ}erleri için \end{aligned} \quad (5)$$

biçiminde tanımlanır. Herhangi iki Y_i ve Y_j ($1 < i < j < n$) sıralı istatistiklerinin bileşik olasılık yoğunluk fonksiyonu ise,

$$\begin{aligned} f_{Y_i, Y_j}(y_i, y_j) &= \frac{n}{(i-1)!(i-j-1)!(n-j)!} [F(y_i)]^{i-1} [1-F(y_j)]^{n-j} \\ &\quad [F(y_j) - F(y_i)]^{j-i-1} f(y_i) f(y_j) \quad , a < y_i < y_j < b \\ &= 0 \quad , \text{ diğ}er \ y_i, y_j \ \text{değ}erleri için \end{aligned} \quad (6)$$

olarak verilir (İnal ve Günay, 2009).

2.2. SKÖ Yöntemi ve Amacı

İstatistiksel arařtırmalarda en önemli amaç kitlenin bilinmeyen parametrelerini tahmin edebilmektir. Kullanılacak yöntem ne olursa olsun toplanan veri kitleyi temsil edemiyorsa elde edilecek sonuçlar gerçeęi yansıtmayacaktır. Bu nedenle seçilecek olan az sayıdaki birimin, örneklemin, bütünü yani kitleyi en iyi temsil eden alt grup olması çok önemlidir. Örneklem genişlięi arttıkça, örneklemin kitleyi temsil edebilme kabiliyetinin arttıęı da bilinmektedir. Ancak arařtırmalarda örneklem büyüklüęü bir kısıttır ve her zaman büyük örneklem genişlikleriyle çalışmak mümkün deęildir. Çaęımızda özellikle tarım, saęlık, ekoloji gibi pek çok çalışma alanında arařtırılan deęişkenin ölçümü maliyetli, zahmetli, zaman alıcı hatta imkansız olabilir. Bu durum kitleyi en az örneklem büyüklüęü ile temsil edebilen bir örnekleme yöntemi ihtiyacını doğurmuştur.

SKÖ yoluyla yüksek düzeyli gözlem sayısı azaltılabildięinden, maliyet ve zaman gibi koşullar da optimize edilmiş olur. Ayrıca SKÖ kitleyi, mükemmel sıralama koşulu altında, yargısal katmanlar içinde homojen gruplara böler (Patil, Sinha, Taillie, 1994). Örneęin, bir üniversitedeki öğrencilerin boy uzunluęu ortalamasını üç öğrenciden oluşan bir örneklem kullanarak tahmin edilmek istensin. BRÖ yöntemine göre tüm üniversite öğrencilerinin arasından üç tanesini rastgele seçip boylarını ölçmek yeterli olacaktır. Elde edilen üç ölçümün ortalaması, kitle ortalamasının yansız bir nokta tahmini olmasına rağmen, örnekleme giren üç öğrenci tamamen kontrol dışında rastgele olarak örnekleme alındıęı için, örneęin iki çok kısa ve bir tane çok uzun öğrenci seçilmiş olabilir ve elde edilen güven aralıęı çok geniş olabilir. Bu istenilen bir durum deęildir ve belirsizlięin göstergesidir. BRÖ kullanarak bu sorunu gidermenin tek yolu örneklem genişlięi n 'yi artırmaktır. Öte yandan üç öğrenci SKÖ yöntemiyle seçilmek istenir ise, ilk önce üç öğrenci davet edip görsel olarak en kısa olduęunu düşündüğümüz öğrencinin gerçek boyu ölçülür ve daha sonra bu işlemi iki kez daha tekrarlayıp ikincisinde orta boylu ve üçüncüsünde en uzun boylu öğrenciyi seçip boy ölçüsü alınır ve sıralı küme örneęini elde edilmiş olur. Aynı şekilde SKÖ ile elde edilen üç ölçümün ortalaması da kitle ortalaması için yansız bir tahmin olacaktır fakat SKÖ ile elde edilen örneklem BRÖ ile elde edilen örnekleme göre daha düzenli olduęundan, SKÖ ile elde edilen tahminin güven aralıęının BRÖ ile elde edilenden daha dar olması beklenir. SKÖ yöntemi aslında, kitleyi örneklem düzeyinde kısa, orta ve uzun boylu şekilde

tabakalamış ve her gruptan gözlem almıştır. Bu nedenle, SKÖ'nün kitleyi temsil kabiliyeti BRÖ'ye göre daha yüksektir.

SKÖ tasarımı dengeli ve dengesiz olmak üzere iki farklı şekilde oluşturulabilir. Bütün kümelerin içinde eşit sayıda birim olması durumu dengeli, farklı sayılarda birim olması durumu dengesiz tasarımı ifade eder. Dengeli tasarımlarda örneklem çekim işlemi iki aşamada gerçekleştirilir. İlk aşamada N büyüklüğündeki kitleden m^2 genişliğinde rastgele örneklem seçilir ve birimler m tane kümeye rastgele olarak dağıtılır. Daha sonra, her bir küme içerisindeki gözlemler kullanılarak ilgilenilen Y değişkeni,, ilgilenilen değişkenle ilişkili olduğu düşünülen ve daha az hassaslıktaki X değişkeni dikkate alınarak ölçülür ve küçükten büyüğe sıralanır. Bu X değişkeni üzerinden yapılan ölçüm, düşük maliyetli ya da daha düşük düzeyli bir ölçüm olabileceği gibi, daha önceki deneyimlerden elde edilen bilgiler ya da görsel bir ölçüm de olabilir. Bir sonraki aşamada i -inci kümenin i -inci elemanı, yani ilk kümenin en küçük değeri, ikinci kümenin ikinci en küçük değeri ve bu şekilde devam edilerek en son kümenin en büyük değeri alınarak ilgilenilen değişkenin ölçümleri yapılır. SKÖ'de örneklem seçim işlemi Çizelge (2.1)'de gösterilmiştir.

Çizelge 2.1. SKÖ' de örneklem seçim süreci

Küme	Kitleden Seçilip Kümelere Rastgele Dağıtılan Birimler	Küme İçinde Sıralanan ve Örneklem Alınan Birimler	Gerçek Ölçümü Yapılan Birimler
1	X_{11} X_{12} ... X_{1m}	$X_{[1:m]}$ $X_{[2:m]}$... $X_{[m:m]}$	$Y_{[1:m]}$ * ... *
2	X_{21} X_{22} ... X_{2m}	$X_{[1:m]}$ $X_{[2:m]}$... $X_{[m:m]}$	* $Y_{[2:m]}$... *
.	.	.	.
.	.	.	.
.	.	.	.
m	X_{m1} X_{m2} ... X_{mm}	$X_{[1:m]}$ $X_{[2:m]}$... $X_{[m:m]}$	* * ... $Y_{[m:m]}$

Sıralamanın kolay yapılabilmesi için, küme büyüklüğü genellikle 3, 4 veya 5 olarak alınır ve 5'ten büyük olması tercih edilmez. Ancak istenilen örneklem büyüklüğü bu

küme büyüklükleri ile sağlanamazsa örneklem seçim işlemi r kez tekrar edilebilir. Böylece r tekrarın sonunda, m^2r genişliğindeki rastgele örneklemden $n=mr$ tanesi üzerinde hassas ölçüm yapıp sıralı küme örnekleme elde edilmiş olur. SKÖ ile edilen bu örneklem, aynı büyüklükteki BRÖ örnekleminin alternatifi olur (Özdemir, 2005).

2.3. Sıralı Küme Örneklemesinde Bağ Bilgisi Olma Durumu

SKÖ'de örneklem seçim işlemi sırasında ilgilenilen Y değişkenine göre tüm birimlerin hassas ölçümlerinin yapıp sıralanması yerine, daha düşük düzeyli bir ölçüm kullanılması sıralama hatasına neden olur. Sıralama hatasına neden olabilecek iki durum vardır. Bunlardan birincisi görsel yolla sıralama diğeri ise yardımcı değişken bilgisi kullanarak sıralamadır (Özdemir, 2005). Tez çalışmasında ilgilenilen Y değişkeninden daha düşük düzeyli olan değişken X olarak tanımlanmıştır. Bu durumda X değişkeni sıralama aşamasında kullanılan yardımcı bilgi değişkenidir ve X görsel yolla ya da düşük düzeyli bir ölçümle elde edilmiş olabilir. Bağ bilgisi durumu X değişkeni üzerinde olduğu için X_T olarak gösterilir.

SKÖ kullanılan bir uygulamada, sıralayıcıdan m büyüklükteki bir kümenin içindeki tüm birimlerin ölçümü olmaksızın, görsel yolla sıralama gibi, küçükten büyüğe sıralaması istenildiğinde veya kesikli nicel verilerle çalışıldığında, değerler doğal olarak birbirine eşit ya da çok benzer olabilir. Sıralayıcının tüm gözlemleri tamamen emin olarak sıralaması mümkün ya da güvenilir olmayabilir. Araştırmacının kişisel yargılarının ön planda olduğu bir sıralama yönteminin kullanılması ise sıralama hatasının ciddi biçimde artmasına neden olur. Bununla birlikte, kümedeki birimler için kısmi bir sıralama sağlanabilir. Örneğin, 4 genişliğindeki bir kümede, sıralayıcı kesin şekilde en büyük ve en küçük gözlemi ayırt edebilirken, ikinci ve üçüncü gözlemi seçerken kararsız kalabilir. Burada ikinci ve üçüncü gözlemler bağlı (tied) olarak adlandırılır. SKÖ'nün kullanılabilmesi için, sıralayıcıdan emin olmasa bile kümedeki her bir birim için bir sıralama yapması beklenildiği için veri içindeki bağ yapısı rastgele kırılır. Önceden belirtildiği gibi, araştırmacının, kişisel yargılarının ön planda olduğu bir sıralama yöntemini kullanması, sıralama hatasının ciddi biçimde artmasına neden olur. Bu hatayı azaltabilmek için bağların rastgele kırılmasından kaynaklanan güçlükler açıklanmaya çalışılmış ve sıralayıcıların bağ bilgisini kullanabileceği bir takım yenilikler önerilmiştir

(Öztürk, 2011). İlk ve en genel öneride, tüm bağlara izin verilecek şekilde, sıralama süreci her setteki birimleri S_1, \dots, S_k ($k \leq m$) şeklinde alt kümelere bölünmüştür. Kümelerin içerisinde yer alan her birim bağlıdır ve kümeler $S_i < S_j$, $i < j$ koşulunu sağlar. Yargısal alt kümelerin oluşturulmasından sonra her kümeden bir birim rastgele seçilerek kesin ölçümü yapılır. Kısmi sıralanmış küme (PROS) adı verilen bu yöntem SKÖ sürecinde sıralayıcılara bazı esneklikler sağlar. SKÖ'de, küme sayısının artması genellikle sıralama hatasına neden olur, ancak bu yöntem, her kümedeki birimlerin kısmen sıralanması nedeniyle küme boyutunun artırılmasına izin verir. Bu, her kümede tüm birimlerin tam sıralaması gerekmediği anlamına gelir. Ancak yapılan bu çalışma, alt küme gruplaması yapma aşamasında çok az ayrıntı sunmaktadır ve alt grup seçim işlemleri de yargısal olarak yapıldığı için bu durum da yine sıralama hatalarına neden olmaktadır. Ayrıca küme sayısını artırmak da her koşulda mümkün olmayabilir. Başka bir problem ise teorik sonuçların, bağlı birimlere ait alt kümeden seçilen birimin o kümeye ait ortalama değeri temsil edemeyeceği iddiasına dayanmasıdır. Bahsedilen aksaklıkların giderilmesi adına tahmin edicilerin içerisinde bağ bilgisinin tamamen kullanılabilmesi başka bir yöntem önerilmiştir (Frey, 2012).

SKÖ yönteminde bağ bilgisi olmadığı varsayılarak her birime kesin bir sıra değeri (rank) verilir. Bağ bilgisi olduğunda SKÖ yönteminin (SKÖ-b) örnekleme şeması SKÖ yöntemine benzerdir ama farklı olarak bağlılık bilgisini kaydeder ve bağ yapısını bozmadan gerçek ölçümünü alacağı birimi bağlı olan birimler arasından rastgele seçer. Eğer, j -inci tekrarın i -nci kümesinde, sırasına karar verilemeyen gözlemler, yani bağ varsa, sıralayıcı bu gözlemlerden birini gerçek değerini ölçümünü almak üzere seçer (seçilen birim Y_{ij} olsun). Sonra, $m \times n$ boyutlu,

$$T^j = \begin{bmatrix} I_{[1:1]}^j & \dots & I_{[1:m]}^j \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ I_{[m:1]}^j & \dots & I_{[m:m]}^j \end{bmatrix} \quad (7)$$

matrisi içine bağa sahip olan bu gözlemler kaydedilir. Burada, $I_{[i:k]}^j$ j -inci tekrarda i -inci ve k -inci sıra değeri almış bağa sahip birimler için bir, diğer durumlar da sıfır değerini alan gösterge değişkenidir ($i = 1, \dots, m; k = 1, \dots, m$). $\sum_{k=1}^m I_{[i:k]}^j \geq 1$ ve eğer sıra değeri i sadece bir birime atanmışsa eşitsizliğin “=” olduğuna dikkat edilmelidir.

SKÖ-b için tüm veriler sadece, bağ bilgisi rastgele çözüldükten sonra j -inci tekrarda i -inci yargısal sıra ile ölçülen birimin kesin ölçü değerleri kümesini $\{Y_{ij}, i = 1, \dots, m, j = 1, \dots, m\}$ değil, aynı zamanda bağ bilgisini içeren T^1, \dots, T^n matrislerini de kapsar.

SKÖ'de Y ve X değişkeni bakımından sıralama yapıldığında formül kullanımları aynıdır fakat gösterim olarak, değişkenler arası korelasyon 1 ise ya da kesin bir sıralama varsa formüller () indisi ile ifade edilirken; ilişkili olan başka değişkene göre sıralama yapıldıysa ve sıralama hatası söz konusuysa formüller [] indisi ile gösterilir. Bağ bilgisi olması durumunda mükemmel sıralama söz konusu olmadığı için bütün formüller [] indisi ile gösterilmiştir.

3. TAHMİN EDİCİLER

Kitleden uygun örneklem yöntemi kullanılarak seçilen örneklem ile, kitle özelliklerini (parametreleri) tahmin etmek amacıyla tanımlanan matematiksel eşitliklere tahmin edici adı verilir. Kullanılacak tahmin edicilerin tutarlılık, yansızlık ve etkinlik özelliklerini sağlamaları beklenir (İnal ve Günay, 2009). Bu bölümde tahmin edicilerin istenilen özellikleri ele alındıktan sonra bu bölümde, BRÖ ve SKÖ'de basit ve oransal tahmin ediciler, literatürde bağ durumunda kullanılan SKÖ-b tahmin edicilerine ve önerilen tahmin edicilere yer verilecektir.

3.1. Tahmin Edicilerin Özellikleri

Örneklem büyüklüğü n , yeteri kadar büyük bir değer alındığında tahmin (\hat{Y}) ile parametre (Y) arasındaki farkın, ε gibi çok küçük pozitif bir sayıdan daha küçük kalma olasılığı 1 ise, o tahmine tutarlı tahmin denir. Matematiksel olarak \hat{Y} tahmini için tutarlılık,

$$\lim_{n \rightarrow \infty} P(|Y - \hat{Y}| < \varepsilon) = 1 \quad (8)$$

olarak gösterilir. Sonlu kitle birimi üzerinden yapılacak tahminlerde, örneklem büyüklüğü kitle büyüklüğüne eşit olduğunda tahmin edicinin değeri parametre değerine eşitse, o tahmin tutarlıdır denir. Yani örneklem büyüklüğü arttıkça tahmin edicinin değeri parametreye yaklaşmalıdır.

Bir tahmin edicinin beklenen değeri kitle parametresine eşit ise o tahmin edicinin yansız olduğu söylenir. Tahmin edicinin beklenen değeri ile parametre arasındaki farka ise yan adı verilir ve

$$B(\hat{Y}) = E(\hat{Y}) - Y \quad (9)$$

olarak gösterilir. Yanın küçük olması istenir. Tahmin edici ile parametre arasındaki farkın karesinin beklenen değerine ise hata kareler ortalaması (HKO) denir ve

$$HKO(\hat{Y}) = E(\hat{Y} - Y)^2 \quad (10)$$

olarak yazılır. Eşitlik (10) içerisine $E(\hat{Y})$ terimi bir kez eklenip çıkartılarak ifade yeniden düzenlenirse HKO,

$$HKO(\hat{Y}) = E\left[(\hat{Y} - E(\hat{Y})) + (E(\hat{Y}) - Y)\right]^2 \quad (11)$$

$$= V(\hat{Y}) + [E(\hat{Y}) - Y]^2 \quad (12)$$

$$= V(\hat{Y}) + B(\hat{Y})^2 \quad (13)$$

tahmin edicinin varyansı ile yanının karesinin toplamı şeklinde elde edilir. Buradan yansız tahmin edicilerde HKO' nun varyansa eşit olacağı görülebilir. İyi bir tahmin edicide aranan özellik HKO' nun küçük olmasıdır. Tahmin edicinin varyansının tersi ise etkinlik olarak adlandırılır. Bir tahmin edici ne kadar küçük varyanslı yani etkin ise o derece duyarlı olduğu söylenir. Tahmin edicilerin birbirlerine karşı etkinlikleri ise göreceli etkinlik (GE) kavramıyla açıklanır. GE değeri iki tahmin edicinin varyanslarının, ya da HKO' larının, birbirine bölünmesi ile aşağıdaki gibi elde edilir.

$$GE = \frac{HKO(\hat{\mu}_i)}{HKO(\hat{\mu}_j)}, \quad i \neq j \quad (14)$$

Eşitlik (14)'e göre, GE değeri 1'den küçükse $\hat{\mu}_i$ tahmin edicisi $\hat{\mu}_j$ tahmin edicisinden daha etkin, değer 1'den büyükse $\hat{\mu}_j$ tahmin edicisi $\hat{\mu}_i$ tahmin edicisinden daha etkindir denir.

BRÖ yöntemi kullanılarak kitle ortalaması tahmin edilmek istendiğinde tahmin edici,

$$\hat{\mu}_0 = \frac{\sum_{i=1}^n y_i}{n} \quad (15)$$

olarak tanımlanır. Burada y_i , i. örneklem biriminin karakteristiğini gösterir. Eşitlik (15)'te verilen yansız tahmine ilişkin varyans ise,

$$V(\hat{\mu}_0) = \frac{\sigma_y^2}{n} \quad (16)$$

biçiminde tanımlanır (Çingı, 2009).

SKÖ' de bağ olması durumunda, gösterge değişkenlerin kullanılmadığı ve bağ bilgisini kullanmayan, SKÖ' nün parametrik olmayan yansız ortalama tahmin edicisi aşağıdaki gibi tanımlanabilir:

$$\hat{\mu}_1 = \frac{1}{m} \sum_{i=1}^m \bar{Y}_{[i]} \quad (17)$$

Burada,

$$\bar{Y}_{[i]} = \frac{1}{n} \sum_{j=1}^n Y_{[i]j} \quad (18)$$

şeklinde tanımlanır ve i -inci sıralı istatistiğin örneklem ortalamasını göstermektedir. Tahmin edicinin varyansı ise,

$$V(\hat{\mu}_1) = \frac{1}{mr} \left[\sigma_y^2 - \frac{1}{m} \sum_{i=1}^m (\bar{Y}_{[i]} - \mu_y)^2 \right] \quad (19)$$

olarak verilir (Özdemir, 2005).

SKÖ kullanılarak ortalamaya ilişkin yapılan bir tahminin BRÖ ile yapılan ortalama tahmininden daha etkin olduğu Eşitlik (19) ve (16) karşılaştırılarak görülebilir. Eşitlik (16)'da $n=mr$ olduğu düşünülürse,

$$V(\hat{\mu}_1) = V(\hat{\mu}_0) - \frac{1}{m^2 r} \sum_{i=1}^m (\bar{Y}_{[i]} - \mu_y)^2 \quad (20)$$

biçiminde yazılabilir ve SKÖ ortalama tahmin edicisinin varyansının BRÖ ortalama tahmin edicisinin varyansından küçük olduğu görülür. SKÖ' de GE değeri kitlenin dağılımından, küme büyüklüğünden ve sıralama hatasından etkilenmektedir. Küme büyüklüğü arttıkça SKÖ'de GE değerleri artmaktadır (Özdemir, 2005).

3.2. Sıralı Küme Örneklemesinde Oransal Tahmin Edicisi

Örneklemeye çalışmalarıda toplam, ortalama ve varyans tahminleri sırasında, tahmin edicilerin duyarlılıklarını ve etkinliklerini artırmak için yardımcı değişken bilgisinin kullanımı çok yaygındır. Yardımcı değişken bilgisi oransal, çarpımsal ve regresyon tahmin edicilerinde kolay uygulanabilirlik ve duyarlılık nedeniyle kullanılmaktadır. Bu tahmin ediciler, yardımcı değişken ile ilgilenilen değişken arasındaki korelasyon yeterli olduğunda, basit tahmin edicilere göre daha küçük varyanslı, daha duyarlı tahminler vermektedir.

Oransal tahmin ediciler, iki değişken arasındaki korelasyon değeri pozitif olduğunda, yardımcı değişken bilgisi yardımıyla kitle toplamı, ortalaması ve varyansının tahmin edilmesinde yaygın olarak kullanılan tahmin edicilerdendir. Değişkenler arasındaki korelasyon negatif yönlü olduğunda ise, çarpımsal tahmin ediciler kullanılmaktadır. Oransal ve çarpımsal tahmin edicilerin, bazı koşullar altında klasik tahmin ediciden daha etkin oldukları bilinmektedir (Çıngı ve Kadılar, 2009).

Yardımcı değişken bilgisini kullanan temel tahmin yöntemlerinden biri olan oransal tahmin ediciler ile ilgili çalışmalar günümüze kadar devam etmiştir. Literatürde, klasik oransal tahmin edicilerin devamında kitlenin ortalaması, toplamı ve varyansı için çeşitli yeni oransal tahmin ediciler geliştirilmiştir.

İlk olarak, 1956 yılında, Quenouille, örneklemin iki rastgele yarısından elde edilen oran tahminlerini kullanarak tahmin edicinin yanlılık miktarını azaltmaya çalışmıştır. Sonrasında Shuckla geliştirilen bu yöntemden yola çıkarak 1971 yılında çarpımsal tahmin ediciyi önermiştir. Yardımcı değişken tarafından sağlanan bilginin tahmin edici içerisinde kullanımı, kapsamlı olarak ilk kez 1977 yılında, Cochran tarafından incelenmiştir. Günümüze kadar BRÖ, sistematik örnekleme, tabakalı örnekleme gibi farklı yöntemler altında çeşitli oransal ve çarpımsal tahmin ediciler geliştirilmiş ve birbirlerine karşı etkinlikleri karşılaştırılmıştır (Prasad, 1989; Beale 1962; Tin, 1965; Ray ve Singh, 1981; Sampath ve Dureirajan, 1988; Upadhyaya ve Singh, 1999; Kadılar ve Çıngı, 2003; Singh ve Tailor, 2003; Shabbir ve Yaab, 2003).

BRÖ için önerilen basit oransal tahmin edici

$$\hat{\mu}_2 = \frac{\bar{y}_{BRÖ}}{\bar{x}_{BRÖ}} \bar{X} \quad (21)$$

biçiminde tanımlanmıştır (Cochran, 1977). Burada, $\bar{x}_{BRÖ}$ ve $\bar{y}_{BRÖ}$ sırasıyla yardımcı değişkene ve ilgilenilen değişkene ait örneklem ortalamalarını, \bar{X} ise yardımcı değişkene ait kitle ortalamasını göstermektedir. Bu tahmin yanlıdır ve yan değeri,

$$B(\mu_2) \cong \frac{\left(\frac{1-f}{n}\right)}{\bar{X}} (RS_x^2 - \rho_{xy} S_y S_x) \quad (22)$$

hata kareler ortalaması ise,

$$HKO(\mu_2) \cong \frac{1}{n} (S_y^2 - 2RS_{yx} + R^2 S_x^2) \quad (23)$$

olarak tanımlanır. Burada, R kitle oranını, S_y^2 ve S_x^2 sırasıyla ilgilenilen değişken ve yardımcı değişken için kitle varyansını, S_{yx} ise ilgilenilen değişken ile yardımcı değişken arasındaki kitle kovaryansını ve ρ_{xy} kitle ilişki katsayısını göstermektedir.

Büyüklüğü N olan kitleden SKÖ yöntemi ile n büyüklüğünde bir örneklem seçilmiş olsun. Örneklem birimlerine ilişkin x_i ve y_i ölçümlerinin elde edilebildiği, (x_i / y_i) oran değişiminin küçük olduğu ve değişkenler arasında pozitif yüksek ilişki olduğu varsayalım. Bu durumda ilgilenilen değişken Y ile yardımcı değişken X 'in SKÖ'de örneklem ortalaması sırasıyla,

$$\bar{Y}_{[n]} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n Y_{[n]ij} \quad (24)$$

ve

$$\bar{X}_{[n]} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n X_{[n]ij} \quad (25)$$

olarak gösterilsin. X ile Y deęişkenleri arasındaki doğrusal baęımtı,

$$\bar{Y}_{[n]} = b \bar{X}_{[n]} \quad , \quad \bar{Y} = b \bar{X} \quad (26)$$

şeklinde yazılabilir. Örneklem ortalaması ve kitle ortalaması birbirine oranlanarak,

$$\frac{\bar{Y}}{\bar{Y}_{[n]}} = \frac{\bar{X}}{\bar{X}_{[n]}} \quad (27)$$

elde edilmiş olur. Bu eşitlikten yola çıkarak, ilgilenilen deęişken ile yardımcı deęişken arasındaki ilişki pozitif olduęu zaman, ilgilenilen deęişkene ilişkin kitle ortalamasının SKÖ için oransal tahmin edicisi aşığıdaki gibi tanımlanmıştır:

$$\hat{\mu}_3 = \frac{\bar{Y}_{[n]}}{\bar{X}_{[n]}} \bar{X} \quad (28)$$

Burada, $\bar{Y}_{[n]}$ ve $\bar{X}_{[n]}$ Eşitlik (24) ve (25)' te tanımlandığı üzere SKÖ ile elde edilen örneklem ortalamaları olarak tanımlanmıştır (Samawi ve Mutlak, 1996).

SKÖ için önerilen basit oransal tahmin edici için sırasıyla HKO ve yan,

$$HKO(\hat{\mu}_3) \cong \frac{I}{mr} (S_y^2 - 2RS_{yx} + R^2S_x^2) - \frac{I}{m^2r} \left[\sum_{i=1}^m (\bar{Y}_{[n]} - \bar{Y})^2 - 2R \sum_{i=1}^m (\bar{Y}_{[n]} - \bar{Y})(\bar{X}_{[n]} - \bar{X}) + R^2 \sum_{i=1}^m (\bar{X}_{[n]} - \bar{X})^2 \right] \quad (29)$$

ve

$$B(\hat{\mu}_3) \cong \bar{Y} \left[\frac{I}{mr} (C_x^2 - \rho_{xy} C_x C_y) - (W_{x(i)}^2 - W_{yx(i)}) \right] \quad (30)$$

olarak elde edilmiştir (Samawi ve Muttlak, 1996). Eşitliklerde C_y ve C_x sırasıyla ilgilenilen değişkenin ve yardımcı değişkenin kitle değişim katsayılarını,

$$W_{yx(i)} = \frac{I}{m^2 r \bar{X} \bar{Y}} \sum_{i=1}^m (\bar{Y}_{[i]} - \bar{Y})(\bar{X}_{[i]} - \bar{X}) \quad (31)$$

ve

$$W_{x(i)} = \frac{I}{m^2 r \bar{X}^2} \sum_{i=1}^m (\bar{X}_{[i]} - \bar{X})^2 \quad (32)$$

göstermektedir.

3.3. SKÖ-b için Literatürde Yer Alan Tahmin Ediciler

Kesim 3.2' de tanımlanan tahmin edicilerin varyansları ve yanları teorik olarak formüllerle tanımlanabiliyorken bu kesimde verilen tahmin edicilerin kapalı varyans formülleri yoktur. Bunun yerine HKO ve yan sayısal olarak hesaplanabilmektedir. Bunun nedeni, tahmin edicilerin içinde matrislerin ve algoritmaların kullanılması nedeniyle kompleks olmaları ve fark yöntemi ya da Taylor serisi gibi yöntemlerle varyans/ yan formüllerinin bulunamamasıdır.

3.3.1. Bağ Matrisi İçeren Basit Tahmin Edici

Bağ bilgisini tahmin edicilere dâhil etmenin bir yolu, bağlı olan her birimin, ait olduğu sıralara karşılık gelen tabakalar arasında bölmektir. Eğer ölçülen bir birim r farklı sıraya bağlıysa, o zaman her r tabakada 1 yerine $1 / r$ ağırlıkla atanmalıdır. Bunu bağları kaydedilen T matrisi üzerinden düşünecek olursak, matrisin her bir satırı bir kümeyi gösterirken her bir sütunu da sıralı istatistikleri (tabakaları) göstermektedir. Sütun toplamı alınıp, sütun üzerindeki her birim bu toplama bölünerek tabakada $1 / r$ ağırlıkları kolaylıkla hesaplanmış olur. Bu strateji MacEachern ve diğerlerinin 2004 yılında önerdiği daha genel bir yöntemin özel halidir ve tahmin edici aşağıdaki gibi yazılabilir:

$$\hat{\mu}_4 = \frac{1}{m} \sum_{i=1}^m \bar{Y}'_{[i]} \quad (33)$$

Burada,

$$\bar{Y}'_{[i]} = \frac{\sum_{\ell=1}^m \sum_{j=1}^{n_\ell} Y_{[\ell]j} \frac{I_{[\ell]ji}}{\sum_{k=1}^m I_{[\ell]jk}}}{\sum_{\ell=1}^m \sum_{j=1}^{n_\ell} \frac{I_{[\ell]ji}}{\sum_{k=1}^m I_{[\ell]jk}}} \quad (34)$$

şeklinde tanımlanır ve tabakalar için ağırlıklı örneklem ortalamasıdır. $I_{[i:k]}^j$, Eşitlik (7)' de verilen T matrisi içerisinde yer alan gösterge değişkendir (MacEachern, 2004).

3.3.2. Basit İzotonik Tahmin Ediciler

Yardımcı değişken ile ölçülen değişken arasındaki korelasyon 1'e eşit olmadığı için ve yardımcı değişkenler yargısal olarak sıralandığı için oluşacak olan sıralama hatalarından dolayı, sıralı istatistiklere ait örneklem ortalamaları $\bar{Y}_{[1]}, \dots, \bar{Y}_{[m]}$ ya da $\bar{Y}'_{[1]}, \dots, \bar{Y}'_{[m]}$ olması gerektiği gibi artan sırada olmayabilir. Bu durumda $\bar{Y}_{[1]}, \dots, \bar{Y}_{[m]}$ ya da $\bar{Y}'_{[1]}, \dots, \bar{Y}'_{[m]}$ tahminleri, örneklem büyüklükleri n_1, \dots, n_m ya da ağırlıklı örneklem büyüklükleri

$$n'_i \equiv \sum_{\ell=1}^m \sum_{j=1}^{n_\ell} \frac{I_{[\ell]ji}}{\sum_{k=1}^m I_{[\ell]jk}}, \quad i = 1, 2, \dots, m \quad (35)$$

kullanılarak izotonize edilebilir. Eşitlik (36) ve (37)' de Wang ve diğerleri tarafından 2008 yılında tanımlanan sırasıyla $\bar{Y}_{[1]}, \dots, \bar{Y}_{[m]}$ ve $\bar{Y}'_{[1]}, \dots, \bar{Y}'_{[m]}$ tahminlerinin izotonik versiyonları

$$\bar{Y}_{[i].iso} = \max_{1 \leq r \leq i} \min_{i \leq s \leq m} \frac{\sum_{\ell=r}^s n_\ell \bar{Y}_{[\ell]}}{\sum_{\ell=r}^s n_\ell}, \quad i = 1, \dots, m \quad (36)$$

$$\bar{Y}'_{[i].iso} = \max_{1 \leq r \leq i} \min_{i \leq s \leq m} \frac{\sum_{\ell=r}^s n'_\ell \bar{Y}'_{[\ell]}}{\sum_{\ell=r}^s n_\ell}, \quad i = 1, \dots, m \quad (37)$$

olarak tanımlanmaktadır.

Eşitlik (36) ve (37)' de izotonize edilmiş tabaka içi ortalamalar, Robertson ve arkadaşlarının 1988'de elde ettiği PAVA algoritması kullanılarak da elde edilebilir. PAVA algoritması, R programında “logcondens” gibi kütüphaneler kullanılarak hesaplanabilmektedir (Mair, Hornik ve Leeuw, 2009). Genel ortalama, tahmini elde edilen izotonik ortalamaların ortalaması olacaktır. Kullanılan izotonizasyon stratejisiyle, Öztürk (2007) tarafından dengesiz SKÖ için ve Wang ve diğ. (2008) tarafından yargısal tabakalama sonrası (judgment post stratified- JPS) parametrik olmayan ortalama tahmin edicileri sırasıyla aşağıdaki gibi önerilmiştir:

$$\hat{\mu}_5 = \frac{1}{m} \sum_{i=1}^m \bar{Y}_{[i],iso} \quad (38)$$

$$\hat{\mu}_6 = \frac{1}{m} \sum_{i=1}^m \bar{Y}'_{[i],iso} \quad (39)$$

Burada $\hat{\mu}_5$, $\hat{\mu}_1$ 'in ve $\hat{\mu}_6$, $\hat{\mu}_4$ ' ün izotonize edilmiş versiyonudur (Zamanzade, 2018).

3.3.3. Rao-Blackwell Tahmin Edicileri

Bağ bilgisini tahmin ediciye dâhil etmenin bir başka yolu da bağların rastgele kırılabilceği tüm farklı yolları dikkate almaktır. Tanımlanan tahmin ediciler, $\hat{\mu}_1$ ve $\hat{\mu}_5$ tahmin edicilerinin Rao-Blackwell versiyonu olacaktır ve HKO' ları değerlendirildiğinde en az $\hat{\mu}_1$ ve $\hat{\mu}_5$ kadar iyi sonuçlar vermesi beklenir (Frey, 2012). Rao-Blackwell yöntemi $\hat{\mu}_4$ ve $\hat{\mu}_6$ için yeni tahmin edicilere yol açmaz, çünkü bu tahminler bağların rastgele nasıl kırıldığına bağlı değildir. $\hat{\mu}_1$ tahmin edicisinin Rao-Blackwell versiyonu $\hat{\mu}_{1,Rao}$, $\hat{\mu}_5$ tahmin edicisinin Rao-Blackwell versiyonu ise $\hat{\mu}_{5,Rao}$ olarak ifade edilmektedir. Literatürde yer alan çalışmalarda Rao-Blackwell tahminleri ortalama tahminlerinde de oran tahminlerinde de kullanılmış fakat GE değeri üzerinde etkisi olmadığı, Rao-Blackwell tahmin edicilerinin bu versiyonu hesaplanan tahmin ediciye neredeyse eşit olduğu görülmüştür. Dolayısıyla, hem daha fazla hesaplama

gerektirmesi hem de GE üzerinde etkisiz olması nedeniyle iyi bir seçim olmadığı belirtilmiştir (Frey, 2012).

Tez çalışmasında simülasyon çalışmalarında bu tahmin edicilerin sayısal gösterimleri verilmemiştir.

3.4. Önerilen Tahmin Ediciler

Ortalama tahmin etmek için kullanılan oransal tahmin edicilerin ilişkili bir yardımcı değişken varlığında klasik tahmin edicilere alternatif ve daha etkili bir yol olarak tercih edilebileceği önceki bölümlerde açıklanmıştır. Bağ bilgisi olması durumunda ise literatürde oransal tahmin ediciler yer almadığı için, önerilen tahmin ediciler çeşitli oransal tahmin edicilerin bağ bilgisini etkili olarak kullanabileceği şekilde modifiye edilmeleriyle oluşturulmuştur.

Kesim 3.3'te olduğu gibi bu bölümde de verilen tahmin edicilerin kapalı varyans ve yan formülleri tahmin edicilerin kompleksliği nedeniyle yoktur. Bunun yerine HKO ve yan sayısal olarak hesaplanabilmektedir.

3.4.1. İzotonik Basit Oransal Tahmin Edici

SKÖ' de bağ bilgisi olması durumu için, yani SKÖ-b' de, geliştirilen ilk oransal tahmin edici Eşitlik (28)'de verilen basit oransal tahmin edici içerisinde örneklem ortalamaları yerine PAVA algoritmasıyla elde edilen izotonik tahminlerin yerleştirilmesiyle, ilk önerilen tahmin edici

$$\hat{\mu}_7 = \frac{\frac{1}{m} \sum_{i=1}^m \bar{Y}'_{[i],iso}}{\frac{1}{m} \sum_{i=1}^m \bar{X}'_{[i],iso}} \bar{X}_t = \frac{\hat{\mu}_6}{\hat{\mu}_{6[xt]}} \bar{X}_t \quad (40)$$

olarak tanımlanmıştır. Burada, $\hat{\mu}_6$ Eşitlik (39)' da verilen, Wang ve diğ. (2008) önerdiği izotonik ortalama tahmini, $\hat{\mu}_{6[xt]}$ ise $\hat{\mu}_6$ 'nın bağ bilgisi olan yardımcı değişken üzerinde uygulanmış halidir. \bar{X}_t ise bağ bilgisi olan yardımcı değişkenin kitle ortalamasıdır.

3.4.2. İzotonik Üstel Oransal Tahmin Edici

Bahl ve Tuteja (1991), BRÖ için

$$\hat{\mu}_{BRÖ} = \bar{y}_{BRÖ} \exp\left(\frac{\bar{X}_t - \bar{x}_{BRÖ}}{\bar{X}_t + \bar{x}_{BRÖ}}\right) \quad (41)$$

şeklinde tanımladığı, daha sonra Singh ve Jatwa (2011) sistematik örnelemeye,

$$\hat{\mu}_{sis} = \bar{y}_{sis} \exp\left(\frac{\bar{X}_t - \bar{x}_{sis}}{\bar{X}_t + \bar{x}_{sis}}\right) \quad (42)$$

biçiminde sistematik örneklem ortalamaları yerleştirerek adapte ettiği üstel oransal tahmin ediciden yola çıkarak ikinci önerilen tahmin edici aşağıdaki gibi tanımlanmıştır:

$$\hat{\mu}_8 = \hat{\mu}_6 \exp\left(\frac{\bar{X}_t - \hat{\mu}_{6[xt]}}{\bar{X}_t + \hat{\mu}_{6[xt]}}\right) \quad (43)$$

Önerilen tahmin edicinin BRÖ ve sistematik örneklemede varyansları elde edilmiştir. Fakat $\hat{\mu}_8$ tahmin edicisinde örneklem ortalamaları yerine algoritma tabanlı örneklem ortalamaları kullanıldığı için HKO ve yanın kapalı formülleri elde edilememektedir. Bunun yerine simülasyon çalışmalarında sırasıyla Eşitlik (9) ve (10) kullanılarak yan ve HKO sayısal olarak hesaplanabilmektedir.

3.4.3. Genelleştirilmiş İzotonik Oransal Tahmin Edici

Bahl ve Tuteja (1991)'nin önerdiği Eşitlik (41)'de verilen tahmin ediciyi Singh ve diğ. (2009),

$$\hat{\mu}_{BRÖ}^* = \bar{y}_{BRÖ} \exp \left[\frac{(\alpha \bar{X}_t + \beta) - (\alpha \bar{x}_{BRÖ} + \beta)}{(\alpha \bar{X}_t + \beta) + (\alpha \bar{x}_{BRÖ} + \beta)} \right] \quad (44)$$

biçiminde tahmin edici ailesi haline getirmişlerdir. Bu tahmin edici ailesi mantığı ile Eşitlik (43)'te verilen izotonik üstel oransal tahmin edici, üçüncü önerilen tahmin edici

$$\hat{\mu}_9 = \hat{\mu}_6 \exp \left[\frac{(\alpha \bar{X}_t + \beta) - (\alpha \hat{\mu}_{6[xt]} + \beta)}{(\alpha \bar{X}_t + \beta) + (\alpha \hat{\mu}_{6[xt]} + \beta)} \right] \quad (45)$$

biçiminde tanımlanmıştır. Burada, α ve β katsayıları 1, X_T basıklık katsayısı, değişim katsayısı ya da korelasyon katsayısı gibi kitle parametreleri olabilir. $\hat{\mu}_9$ tahmin edicisi sadece bu farklı α ve β kombinasyonları kullanılarak 16 farklı tahmin ediciye dönüşebilir. Eşitlik (45)'teki tahmin edicide $\alpha=1$ ve $\beta=0$ olarak alınırsa Eşitlik (43)'te verilen tahmin ediciye dönüştüğü görülebilir.

3.5. Sayısal Örnek

Bu kesimde, SKÖ-b örneklem şeması üzerinde tahmin edicilerin hesaplanışını gerçek bir kitle üzerinden bir örnekle açıklanmıştır. 2019 yılının sonlarına doğru Çin'de başlayarak kısa bir süre içerisinde tüm dünyayı etkisi altına alan Koronavirüs (COVID-19) bulaşma yolu olarak en hızlı ve kontrolü en zor olan, solunum yolu ile bulaşan bir virüsdür. Kitle olarak, Dünya Sağlık Örgütü'nün 11 Mart 2020 tarihinde yayınladığı Koronavirüs Durum Raporu 51'de yer alan, Çin'deki bölgelere/şehirlere ait bildirilen doğrulanmış ve şüpheli hasta sayıları kullanılmıştır (WHO, 2020). Verilen kitlede hastalıktan ölen hasta sayısını tahmin etmeye çalıştığımızı varsayalım. Son 24 saatte

hastalığa yakalanan vaka sayısı ile vefat eden hasta sayıları arasında pozitif yüksek korelasyon olduğu görülmüştür. Bu nedenle yardımcı değişken olarak son 24 saatte onaylanmış vaka sayısı kullanılmıştır. Ek 1' de verilen kitleye ait parametreler Çizelge 3.1'de verilmiştir.

Çizelge 3.1. COVID-19 verileri için kitle parametreleri

$N=34$	$\beta_{X_T}=16.77$	$\bar{X}_T = 0.9118$
$m=n=5$	$\rho_{X_T Y}=0.842$ ($p\text{-value}<0.05$)	$\bar{Y} = 93$
$\sigma_Y^2= 272275.6$, $\sigma_{X_T}^2=6.4465$	Min. $X_T= 0$, Max. $X_T=13$	Min. $Y=0$, Max. $Y=3046$

SKÖ ile $n= m= 4$ olacak şekilde bir örneklem belirlenmiştir. Bunun için öncelikle $N=34$ olan kitleden, $m^2=16$ tane yerleşim yeri BRÖ ile seçilsin. Daha sonra bu 16 birim rastgele 4 gruba ayrılır. Her bir küme içerisindeki birimler, yardımcı değişken bilgileri kullanılarak sıralanır. Çizelge 3.2'de, RSS-t şemasının bir örneği ayrıntılı olarak verilmiştir. Aralarında bağ bilgisi olan birimler "&" işaretiyle, gerçek ölçümü alınmak üzere örnekleme seçilen rastgele birim kalın olarak gösterilmiştir.

Çizelge 3.2. Örneklemdeki birimlerin sıralanışı

Küme	Sıralanmış Birimler (Şehirler/Bölgeler)	Yardımcı Değişken (X_T)	Kesin Ölçüm Değeri (Y)
1	Guangxi & Sichuan & Hunan, Hubei	0	2
2	Inner Mongolia & Anhui & Guangdong & Ningxia	0	8
3	Hebei & Yunnan, Heilongjiang , Hong Kong SAR	1	13
4	Hainan & Anhui & Liaoning, Shanghai	2	3

Çizelge 3.2'de, 1 ve 2 numaralı kümelerin içerisindeki bağların örneklem seçimini etkilediği görülebilir. 1. kümedeki Guangxi şehri birinci, ikinci ve üçünü sıra ile; 2. kümedeki Guangdong şehri tüm sıralarla bağlıdır. Bağlı olan bu birimlerden rastgele seçilen bir yerleşim yerine ait kesin ölçüm bilgileri alınmış ve bu bağ yapısı T matrisi içerisine aşağıdaki gibi kaydedilmiştir.

$$T^1 = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Örnekleme çekiminde tekrar sayısı bir olduğu için sadece bir tane T matrisi oluşacaktır. $\bar{Y}_1 = 2$, $\bar{Y}_2 = 8$, $\bar{Y}_3 = 13$ ve $\bar{Y}_4 = 3$ olduğu görülebilir. Buradan $\hat{\mu}_1 = 6.5$ olarak hesaplanır. T matrisinde kaydedilen bağılı birimleri tabakalar arasında bölünürse,

$$T_{n_{ij}}^* = \begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

olarak elde edilir ve buradan sütun toplamlarını alarak ağırlıklı örneklem büyüklüklerini $(n'_1, \dots, n'_4) = (0.58\bar{3}, 0.58\bar{3}, 1.58\bar{3}, 1.25)$ şeklinde elde edilmiş olur. Ağırlıklı örneklem ortalamaları $T_{n_{ij}}^*$ matrisi ile kesin ölçüm değerlerini içeren $[2 \ 8 \ 13 \ 3]'$ matrisi ile çarpılırsa: $\bar{Y}'_1 = 4.571$, $\bar{Y}'_2 = 4.571$, $\bar{Y}'_3 = 9.894$ ve $\bar{Y}'_4 = 4$ olarak bulunur ve örneklem ortalamalarının ortalaması alınarak $\hat{\mu}_4 = 5.75$ biçiminde hesaplanır. Örneklem ortalamalarını incelersek $\bar{Y}_1 \leq \bar{Y}_2 \leq \bar{Y}_3 \leq \bar{Y}_4$ ve $\bar{Y}'_1 \leq \bar{Y}'_2 \leq \bar{Y}'_3 \leq \bar{Y}'_4$ şeklinde azalmayan (eşit ya da artan) sırada olmadıkları görülür. Örneklem ağırlıkları kullanılarak bu ortalamalar izotonize edilir ve ortalamaların azalmayan (eşit ya da artan) sırada olması sağlanabilir. $\hat{\mu}_1$, yani SKÖ basit ortalamasında bağ bilgileri kullanılmadığı için (n_1, \dots, n_4) ağırlıkları birbirine eşit ve herbiri 1'dir. Örneklem ağırlıklarıyla, örneklem ortalamaları PAVA algoritması kullanılarak izotonize edilirse izotonik örneklem ortalamaları $\bar{Y}_{1,iso} = 2$, $\bar{Y}_{2,iso} = 8$, $\bar{Y}_{3,iso} = 8$, $\bar{Y}_{4,iso} = 8$ ve $\bar{Y}'_{1,iso} = 4.471$, $\bar{Y}'_{2,iso} = 4.571$, $\bar{Y}'_{3,iso} = 7.294$, $\bar{Y}'_{4,iso} = 7.294$ olarak bulunur. Elde edilen izotonik örneklem ortalamalarının ortalaması alınarak da $\hat{\mu}_5 = 6.5$ ve $\hat{\mu}_6 = 5.932$ olarak hesaplanır. $\hat{\mu}_2$, BRÖ kestiricisi olduğu için öncelikle $n=4$ büyüklüğünde rastgele bir örneklem çekilmelidir. Çekilen örnekleme yer alan birimler

Ningxia, Macao SAR, Hong Kong SAR ve Anhui şehirleri olsun. Buradan $\hat{\mu}_2=1.64$ olarak bulunur. $\hat{\mu}_3$ tahmin edicisini hesaplamak için bağ bilgisi kullanılmayacağından $\bar{y}_{[n]} = \hat{\mu}_1 = 6.5$ olacaktır. Buna ek olarak $\hat{\mu}_3$ oransal tahmin edici olduğu için örnekleme çekilen birimlerin yardımcı değişkenine ait örneklem ortalamasına da ihtiyaç duyulmaktadır ve bu ortalama $\bar{x}_{[n]} = 0.65$ olarak hesaplanabilir. Buradan $\hat{\mu}_3 = 7.9$ şeklinde hesaplanır. Önerilen tahmin edicilerin hesaplanabilmesi için yardımcı değişkenin izotonik örneklem ortalamaları gereklidir. Bu $\hat{\mu}_6$ tahmin edicisi hesaplanırken izlenen yolun, bu kez yardımcı değişken için uygulanması demektir. Fakat çekilen örneklemede, Çizelge 3.2' den, $\bar{X}'_1 \leq \bar{X}'_2 \leq \bar{X}'_3 \leq \bar{X}'_4$ olduğu için izotonik örneklem ortalamaları, basit örneklem ortalamalarına yani $\hat{\mu}_{4[x]}$ 'ye eşittir. $\hat{\mu}_{4[x]}$ 'de $\hat{\mu}_4$ tahmin edicisinin yardımcı değişkenin verisiyle hesaplanmış halidir. Böylece $\hat{\mu}_{6[x]} = \hat{\mu}_{4[x]} = 0.558$ olarak hesaplanır. $\hat{\mu}_6$ ve $\bar{X}_T = 0.9118$ bilgileri de kullanılarak $\hat{\mu}_7 = 9.693$ ve $\hat{\mu}_8 = 7.54$ olarak bulunur. α ve β sırasıyla yardımcı değişkenin çarpıklık katsayısı (β_{XT}) ve korelasyon katsayısı (ρ_{XTY}) olarak alınmış olsun buradan $\hat{\mu}_9 = 7.43$ şeklinde hesaplanır. Rao-Blackwell tahmin edicileri için bağların olası tüm durumlar için değerlendirilmesi ile oluşacak tahminlerin ortalaması alınmalıdır. İlk kümede yer alan Guangxi şehri 1, 2 ve 3 ranklarına bağlıdır. İkinci kümede ise Guangdong şehri 1,2,3 ve 4 ranklarına bağlıdır. Bu durumda bağ 12 farklı durumda Çizelge 3.3'teki gibi kırılabilir.

Çizelge 3.3. Örneklemdaki birimlerin sıralanışı çizelgesindeki bağların kırılabilceği tüm farklı durumlar

Guangxi şehri	Guangdong şehri	$(n_1, n_2, n_3, n_4) = (1, 1, 1, 1)$
1	1	(2, 0, 1, 1)
1	2	(1, 1, 1, 1)
1	3	(1, 0, 2, 1)
1	4	(1, 0, 1, 2)
2	1	(1, 1, 1, 1)
2	2	(0, 2, 1, 1)
2	3	(0, 1, 2, 1)
2	4	(0, 1, 1, 4)
3	1	(1, 0, 2, 1)
3	2	(0, 1, 2, 1)
3	3	(0, 0, 3, 1)
3	4	(0, 0, 2, 2)

Çizelge 3.3'deki 12 farklı durum içerisinde sadece ikisi örneklem seçim yöntemine uygunluk göstermektedir ve koyu olarak işaretlenmiştir. Guangxi şehrinin 1, Guangdong şehrinin 2 olduğu durum $\hat{\mu}_1$ tahmin edicisinin aynıdır ve değeri 6.5'tir. Guangxi şehrinin 2, Guangdong şehrinin 1 olduğu durum için ise tahmin edici yine aynı şekilde 6.5 olarak bulunacaktır. Bu iki tahmin birbirine eşit olduğundan ortalamaları da kendisine eşit olur ve $\hat{\mu}_{1,Rao} = \hat{\mu}_1 = 6.5$ olarak bulunmuş olur. Benzer hesaplamalar $\hat{\mu}_5$ tahmin edicisi ile yapılırsa yine aynı şekilde $\hat{\mu}_{5,Rao} = \hat{\mu}_5 = 6.5$ olarak bulunabilir ve böylece $\hat{\mu}_{1,Rao} = \hat{\mu}_1 = \hat{\mu}_{5,Rao} = \hat{\mu}_5$ olarak bulunur. Kesim 3.3.3'te bahsedildiği gibi Rao-Blackwell versiyonu ile hesaplanan tahmin edicilerin, tahmin edicilerin kendisine eşit olduğu bu örnek üzerinde de görülmüştür.

4. SİMÜLASYON

Bu bölümde, kitle üzerinde bağ bilgisinin nasıl oluşturulduğu anlatıldıktan sonra, bağ bilgisini içeren büyük (N=1000) ve küçük (N=50) olmak üzere iki kitle üzerinden simülasyon ile örneklemeler çekilip tahmin ediciler hesaplandıktan sonra HKO, GE ve yan değerleri elde edilerek yorumlanmıştır.

4.1. Simülasyonla Veri Setinde Bağ Üretme Yöntemleri

Her i -inci birimin kesin ölçüm değeri Y_i 'ye sahip olduğu bir set düşünelim. Her birime ait $(X_1, Y_1), \dots, (X_m, Y_m)$ ikililerinin bağımsız ve iki değişkenli bir dağılımdan çekildiğini varsayalım. Dell ve Clutter 1972'deki önerisine göre, rastgele değişkenler X_1, X_2, \dots, X_m ve Y_1, Y_2, \dots, Y_m ,

$$X_i = Y_i + \varepsilon_i \quad (46)$$

modeline uyar. Burada, Y_1, Y_2, \dots, Y_m birbirinden bağımsız ve herhangi bir dağılımdan gelen, $\varepsilon_1, \dots, \varepsilon_m$ bağımsız ve ortalaması 0 varyansı σ^2 olan normal dağılımlıdır. Sıralama X_1, X_2, \dots, X_m değişkenleriyle yapılır ve kesin ölçüm değerleri Y değişkenine karşılık gelir. Modelde $\sigma^2=0$ olması durumu mükemmel sıralamayı gösterirken σ^2 büyüdükçe sıralama rastgeleleşir ve sıralama hatası artar.

Simülasyonla türetilen verilerde sıralamada bağ oluşturabilmek için tanımlanmış iki yöntem vardır. Birincisi, Flingner ve MacEachern (2006) tarafından geliştirilmiş olan Kesikli Seçilen Boyut (Discrete perceived size, DPS) ve Dell ve Clutter (1972) tarafından önerilen Yakınsa Bağlı (Tied-if-close, TIC) yöntemleridir.

DPS yönteminde X değişkeni, X/e işleminin aşağıya yuvarlanmasıyla kesikli hale getirilir. $X_T = \lfloor X/e \rfloor$ olarak yeniden elde edilir ve burada $e > 0$ olacak şekilde seçilmesi gereken bir parametredir. $X_{T_i} = X_{T_j}$ olduğu durumda i -inci ve j -inci birimler bağlıdır denir.

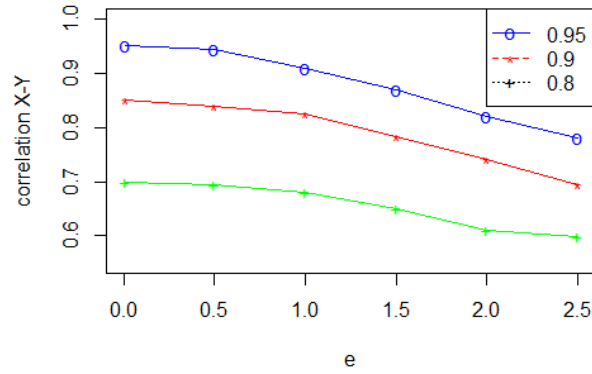
TIC yönteminde ise $|X_i - X_j| < e$ ise bu durumda i -inci ve j -inci birimler bağlıdır denir. Burada, e aynı şekilde sıfırdan büyük olacak şekilde seçilmesi gereken bir parametredir. $|X_i - X_j| \geq e$ olduğu durumlarda birimlerin birbirinden yeteri kadar ayrık olduğu düşünülerek bağlı kabul edilemez. Eğer X sürekli ve e çok küçükse oluşan bağların sayısı az, e büyüdükçe bağların sayısı daha çok olacaktır.

Ancak bu iki yöntemde de e ve m parametreleri değiştirildiğinde dağılımlar istenmeyen davranışlar gösterebilir. Bu durum genellikle e değeri arttığında ortaya çıkar. Artan e değerinin daha fazla bağa yol açması beklenir fakat durum her zaman istenilen sonucu vermeyebilir. Bu durum Frey' in 2012'deki teoremiyle de gösterilmiştir. Sabit bir e ve aynı dağılımlı X ler için TIC modelinde DPS modelinden daha çok sayıda bağ oluştuğu Frey (2012) ve Zamanzade (2018) çalışmalarında belirtilmiştir. Herhangi bir yöntemde, e 'nin yeterli büyüklükte seçilmesiyle kümelerde bağ oluşma olasılığı sıfırdan farklı ve büyük olacaktır. Ancak, iki yöntemde de modeller e ve m parametreleri değiştirildiğinde istenmeyen davranışlar gösterebilir. TIC modelinde veri setine yeni ek bir birim eklendiğinde bağların sayısı aniden artabilir. Bu durum eklenen birimin belirlenen e büyüklüğündeki boşluğu doldurmasıyla ilgilidir. Örneğin, küme büyüklüğü 2 ve $e=2$ olsun. Kümedeki birimler $X_1 = 1.2$ ve $X_2 = 3.8$ iken TIC yöntemine göre bu sette hiç bağ oluşmaz. Fakat, üçüncü bir birim olarak $X_3 = 2.6$ eklenirse bu sette $|X_1 - X_3| < 2$ ve $|X_2 - X_3| < 2$ olduğu için bağlar oluşacaktır. DPS modelinde istenmeyen durumlar e değeri arttığında ortaya çıkar. Artan e değerinin daha fazla bağa yol açması beklenir fakat durum her zaman istenilen sonucu vermeyebilir. Örneğin yardımcı değişken $X_i = \{\dots, -2, -1, 0, 1, 2, \dots\}$ olduğunda, $e=2$ için $X_i^T = \{\dots, -1, -0.5, 0, 0.5, 1, \dots\}$ biçiminde birbirinden farklı şekilde ve hiç bağ üretmeden de oluşabilir.

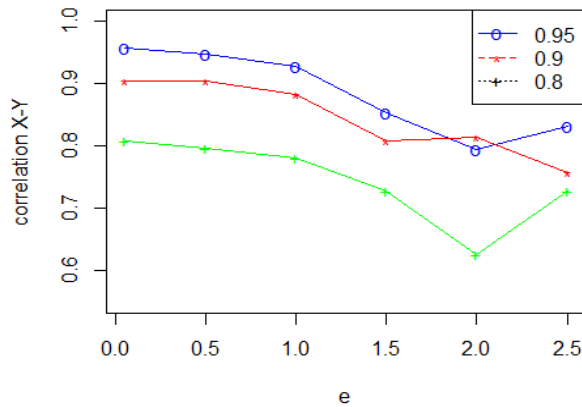
4.2. Simülasyon Çalışması

Oransal tahmin edicilerde yardımcı değişkenin kitle ortalaması bilgisine ihtiyacımız olduğu için ve TIC yöntemiyle kitle üzerinde bağ üretilmediğinden, simülasyonda DPS yöntemi kullanılmıştır.

e değeri büyüdükçe, X_T yardımcı değişkeni gerçek değerlerinden uzaklaştığı için, X_T ve Y arasındaki korelasyonunun azalması beklenen bir durumdur. Simülasyon çalışmasında e değerini artırarak bağ sayısını çoğalttığımızda, Şekil 4.1 ve 4.2'den, başlangıçtaki korelasyon değerinin küçük kitlelerde daha hızlı azaldığını ve daha fazla bozulduğunu görüyoruz. Bunun nedeni kitle birimlerinin örnekleme seçilme olasılıklarının büyük ve küçük kitlelerde aynı olmayışdır. Frey (2011) çalışmasında, SKÖ'de kitledeki birimlerin örnekleme dahil olma olasılıklarını araştırmış ve bu olasılıkların büyük ve küçük kitlelerde farklı olduğunu göstermiştir.

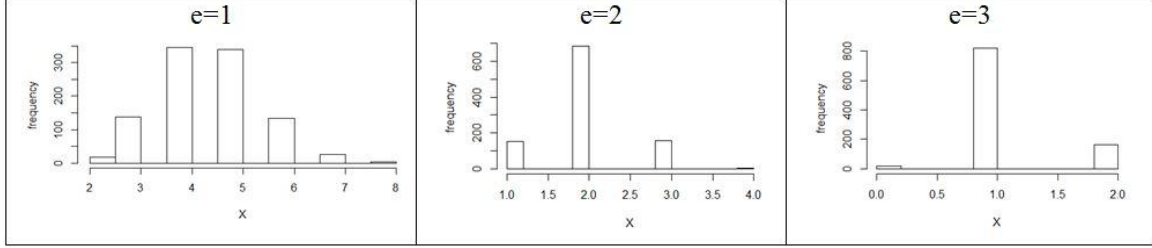


Şekil 4.1. Büyük bir kitlede korelasyon katsayısının e 'ye göre değişimleri (N=1000)



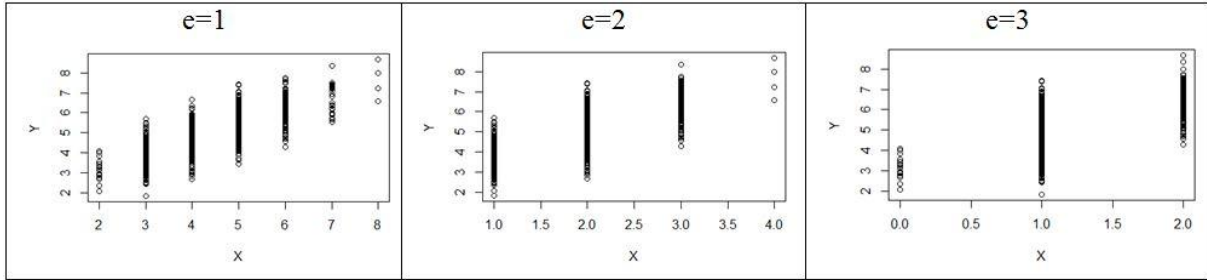
Şekil 4.2. Küçük bir kitlede korelasyon katsayısının e 'ye göre değişimleri (N=50)

Farklı e değerleri için normal dağılım altında yardımcı değişkene ait histogram grafikleri Şekil 4.3'te gösterilmiştir.

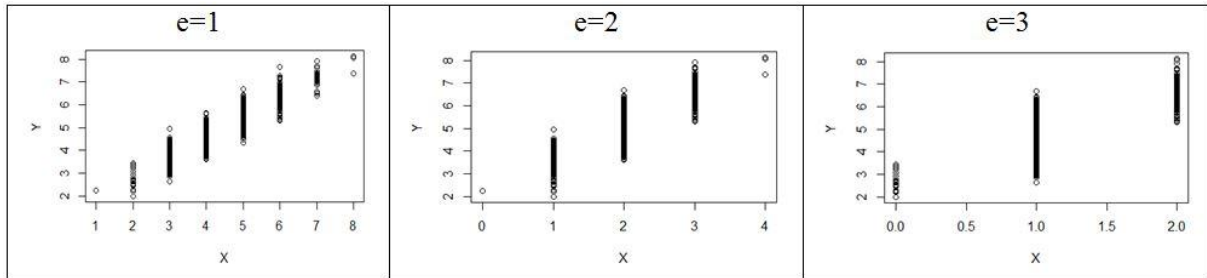


Şekil 4.3. $N(5,1)$ ve $\rho=0.75$ iken farklı e değerleri için X histogramları

Farklı e değerleri için normal dağılım altında X - Y saçılım grafikleri Şekil 4.4 ve 4.5'te gösterilmiştir.



Şekil 4.4. $N(5,1)$ ve $\rho=0.75$ iken farklı e değerleri için X - Y saçılım grafikleri



Şekil 4.5. $N(5,1)$ ve $\rho=0.9$ iken farklı e değerleri için X - Y saçılım grafikleri

Oransal tahmin edicilerin kullanılabilmesi için ilgilenilen deęişken ile yardımcı deęişken arasında pozitif yönlü yüksek bir ilişki olması beklenir. Bu nedenle DPS modeli için gereken e parametresi X ile Y arasındaki korelasyon katsayısını düşürmeyecek şekilde her dağılıma özel olarak belirlenmiştir.

Oransal tahmin edicilerde yardımcı deęişkenin kitle ortalaması bilgisine ihtiyaç olduğu için ve TIC yöntemiyle kitle üzerinde bağ üretilemediğinden, simülasyon çalışmasında DPS yöntemi kullanılmıştır. R programı kullanılarak yazılan simülasyonda, 1000 büyüklüğündeki ve 50 büyüklüğündeki kitleler, iki deęişkenli normal dağılımlı olacak şekilde türetilmiştir.

Parametreler $\rho = 0.75, 0.85$ ve 0.95 ; $m = 3, 4$ ve 5 olarak belirlenmiş verilen ortalama tahmin edicileri dengeli SKÖ yöntemi ile yerine konmadan ve simülasyon sayısı 100000 için elde edilmiştir. Daha sonra Eşitlik (10) kullanılarak HKO' ları, Eşitlik (9) kullanılarak yan deęerleri elde edilmiştir. Referans tahmin edicisi $\hat{\mu}_1$ alınarak, GE deęerleri Eşitlik (14) ile verilen formül kullanılarak Eşitlik (47)' de verildiği gibi hesaplanmıştır. $\hat{\mu}_9$ tahmin edicisinin katsayıları rastgele seçilip α deęeri X_T yardımcı deęişkeninin basıklık katsayısı, β deęeri X_T ile Y arasındaki kitle korelasyon deęeri olarak alınmıştır. $\hat{\mu}_1$ tahmin edicisinin izotonik versiyonu olan $\hat{\mu}_5$ tahmin edicisi her zaman $\hat{\mu}_1$ 'e eşit olduğu için GE deęeri bire, yanı ise $\hat{\mu}_1$ 'in yanına eşittir. Dolayısıyla, $\hat{\mu}_5$ tahmin edicisi Çizelge 4.2, 4.3, 4.4 ve 4.5'te yer almamaktadır.

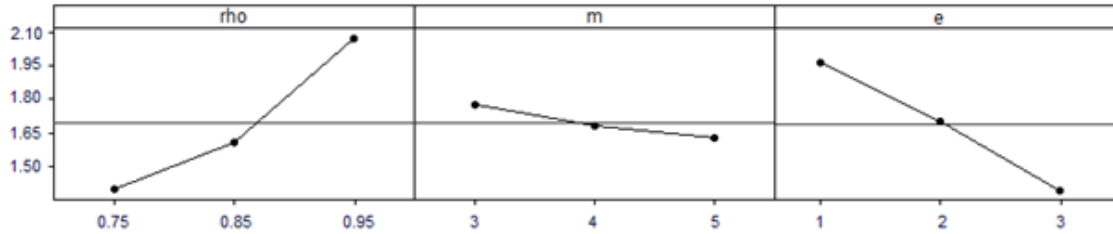
$$GE_i = \frac{HKO(\hat{\mu}_1)}{HKO(\hat{\mu}_i)}, \quad i = 0, 2, 3, 4, 6, 7, 8, 9 \quad (47)$$

GE deęerlerinden elde edilebilecek bilgiler Çizelge 4.1'de özetlenmiştir.

Çizelge 4.1. Görelî etkinlik bilgileri

	Örnekleme Türü	Bağ Bilgisi	Yardımcı Değişken Bilgisi	İzotonik
GE_0	BRÖ	-	-	-
GE_2	BRÖ	-	+	-
GE_3	SKÖ	-	+	-
GE_4	SKÖ	+	-	-
GE_6	SKÖ	+	-	+
GE_7	SKÖ	+	+	+
GE_8	SKÖ	+	+	+
GE_9	SKÖ	+	+	+

SKÖ'de bağ bilgisi olan yardımcı değişken olması durumunda, kitle simetrik dağılımlı ve sıfır ortalamalı ise \bar{X}_t ya da önerilen tahmin edicide yer alan $\hat{\mu}_{6[x]}$ SKÖ örneklem ortalama değeri sıfır olabilmektedir. Bu nedenle normal dağılım için parametreler (5,1) olarak rastgele seçilmiştir. GE sonuçları Çizelge 4.2 ve 4.3'te, yan değerleri ise Çizelge 4.4 ve 4.5'te verilmiştir. Çizelgelerde verilen başlangıç korelasyon değerlerinin Şekil 4.1 ve 4.2'deki gibi e değeri arttıkça değiştiğine dikkat edilmelidir. Şekil 4.6'da simülasyon sonuçlarından elde edilen GE değerlerinin küme büyüklüğü ve kitledeki bağ sayısı arttıkça azaldığı, ilişki katsayısı arttıkça arttığı görülebilir.



Şekil 4.6. Görelî etkinliğin korelasyon katsayısına, küme büyüklüğüne ve e değerine göre değişimleri

Çizelge 4.2. $\mu=5, \sigma^2=1$ parametrelili ve $N=1000$ olan Normal dağılımlı kitle için göreceli etkinlik değerleri

DPS $x_i = \lfloor X/e \rfloor$	ρ_{XY} m	0.75			0.85			0.95		
		3	4	5	3	4	5	3	4	5
$e=1$	GE_0	0.7748	0.7094	0.6805	0.6998	0.6367	0.5921	0.6322	0.5571	0.4973
	GE_2	1.125	1.0635	1.0331	1.5471	1.4597	1.3639	2.57	2.3116	2.0921
	GE_3	1.3412	1.2897	1.2882	1.116	1.7239	1.6376	3.0204	2.7691	2.5933
	GE_4	1.0236	1.0235	1.0291	1.0434	1.0517	1.065	1.0892	1.1201	1.1403
	GE_6	1.0673	1.0689	1.0748	1.0868	1.0989	1.1107	1.1301	1.1632	1.1842
	GE_7	1.3462	1.3247	1.2407	1.7551	1.66	1.5917	3.0011	2.7144	2.5555
	GE_8	1.5873	1.4779	1.2437	1.9365	1.7964	1.6914	2.8058	2.5872	2.428
	GE_9	1.5842	1.4757	1.4217	1.9632	1.8167	1.7069	2.7975	2.5809	2.423
$e=2$	GE_0	0.855	0.8263	0.7866	0.814	0.7759	0.7538	0.7747	0.7197	0.6835
	GE_2	0.6276	0.6396	0.6317	0.7661	0.7708	0.7649	0.9299	0.9125	0.9014
	GE_3	0.8278	0.8682	0.8775	1.007	1.0274	1.0468	1.2404	1.2694	1.2776
	GE_4	1.0794	1.0819	1.0843	1.1009	1.1088	1.1132	1.1682	1.1822	1.1829
	GE_6	1.1197	1.1218	1.1224	1.1374	1.147	1.1504	1.1997	1.2173	1.2206
	GE_7	0.9867	1.0201	1.028	1.1318	1.1382	1.1521	1.5698	1.5629	1.5105
	GE_8	1.4635	1.4187	1.3837	1.688	1.6228	1.5753	2.1764	2.0579	1.9727
	GE_9	1.4744	1.4264	1.3913	1.6886	1.6229	1.5754	2.1736	2.0555	1.9706
$e=3$	GE_0	0.9265	0.8987	0.8901	0.8942	0.8769	0.8619	0.8798	0.8517	0.8288
	GE_2	0.3979	0.4107	0.4118	0.0001	0.4029	0.4058	0.0001	0.4573	0.4537
	GE_3	0.471	0.5041	0.5293	0.4633	0.5019	0.5321	0.5386	0.5695	0.602
	GE_4	1.0664	1.0669	1.0718	1.0911	1.0949	1.0994	1.1316	1.1376	1.139
	GE_6	1.0985	1.0971	1.099	1.1216	1.1234	1.1266	1.1636	1.1698	1.1698
	GE_7	0.5419	0.6674	0.7771	0.6707	0.8078	0.9121	0.6479	0.799	0.9255
	GE_8	1.1717	1.1609	1.1585	1.2547	1.2368	1.2381	1.5019	1.4625	1.4488
	GE_9	1.298	1.2751	1.2635	1.3831	1.3551	1.3422	1.5971	1.5527	1.527

Çizelge 4.3. $\mu=5, \sigma^2=1$ parametrel ve $N=50$ olan Normal dağılımlı kitle için görel etkinlik değerleri

DPS $x_i = \lfloor x / e \rfloor$	ρ_{XY} m	0.75			0.85			0.95		
		3	4	5	3	4	5	3	4	5
e=1	GE_0	0.7896	0.7347	0.6927	0.6323	0.5499	0.4948	0.6099	0.5213	0.4569
	GE_2	0.8321	0.7906	0.7604	1.9353	1.7274	1.5515	3.7192	3.2281	2.8789
	GE_3	1.0675	1.076	1.099	2.3207	2.1148	1.9114	3.9967	3.4852	3.0564
	GE_4	0.9925	0.9895	0.9895	1.0839	1.1059	1.1253	1.1432	1.2075	1.2689
	GE_6	1.0464	1.0525	1.0576	1.1258	1.1535	1.1804	1.1867	1.2572	1.3214
	GE_7	1.013	1.0226	1.0549	2.2245	2.0051	1.8023	3.7329	3.1975	2.7894
	GE_8	1.3321	1.2461	1.1958	2.2294	1.9867	1.7871	2.6274	2.4467	2.2815
	GE_9	1.3292	1.2433	1.1921	2.1515	1.9339	1.7518	2.492	2.3465	2.2083
e=2	GE_0	0.8121	0.7766	0.743	0.7717	0.724	0.6792	0.7972	0.7522	0.7092
	GE_2	0.4535	0.4653	0.4656	0.8686	0.8548	0.106	1.1297	1.1159	1.0981
	GE_3	0.6692	0.7229	0.753	1.1403	1.1682	1.1706	1.3729	1.3945	1.3803
	GE_4	1.0707	1.0815	1.092	1.1663	1.1772	1.1847	1.1944	1.2029	1.2118
	GE_6	1.119	1.1342	1.1418	1.2002	1.2194	1.2335	1.2295	1.2437	1.2579
	GE_7	0.7996	0.632	0.9111	1.3058	1.3218	1.3167	1.5049	1.5084	1.4881
	GE_8	1.4274	1.3767	1.3271	1.9716	1.8463	1.7488	1.8689	1.7862	1.7109
	GE_9	1.4561	1.3961	1.341	1.9263	1.8182	1.7265	1.8304	1.7532	1.6837
e=3	GE_0	0.9206	0.8937	0.8677	0.9055	0.8841	0.8566	0.9055	0.8919	0.8647
	GE_2	0.376	0.3557	0.3428	0.4373	0.4419	0.4328	0.7206	0.7152	0.695
	GE_3	0.4303	0.4522	0.4845	0.4965	0.5205	0.5444	0.7756	0.8071	0.8227
	GE_4	0.9905	1.001	1.0101	1.1169	1.1196	1.1181	1.2091	1.2197	1.2171
	GE_6	1.0381	1.0455	1.0505	1.1481	1.1506	1.1507	1.2315	1.2449	1.2437
	GE_7	0.4562	0.4789	0.5235	0.663	0.6718	0.690	0.9455	0.9656	0.9648
	GE_8	0.998	0.999	1.0142	1.3197	1.2938	1.279	1.5641	1.5579	1.5184
	GE_9	1.1003	1.0916	1.093	1.38	1.3499	1.3296	1.5862	1.5579	1.5376

Çizelge 4.4. $\mu=5, \sigma^2=1$ parametrelili ve $N=50$ olan Normal dağılımlı kitle için yan değerleri

DPS $X_i = \lfloor X / e \rfloor$	ρ_{XY}	0.75			0.85			0.95		
	m	3	4	5	3	4	5	3	4	5
$e=1$	B_0	-0.0007	-0.0004	-0.0012	-0.0023	-0.0021	-0.0005	-0.0045	-0.0016	0.0007
	$B_{1,5}$	0.0008	0.0001	-0.0001	0.0018	-0.0011	0.0005	-0.0011	-0.0008	-0.0017
	B_2	0.0405	0.0324	0.0202	0.0224	0.0161	0.0117	0.0078	0.0071	0.0046
	B_3	0.0198	0.013	0.0076	0.0114	0.006	0.0058	0.0046	0.0011	0.0012
	B_4	-0.0015	-0.0002	0.0001	0.0084	0.0042	0.0044	-0.0037	-0.0032	-0.0045
	B_6	0.0007	0.0009	0.0006	0.0065	0.0023	0.0029	-0.0032	-0.0028	-0.0035
	B_7	0.0321	0.0242	0.0176	0.009	0.0047	0.0052	0.003	0.0003	0.0006
	B_8	0.0104	0.0094	0.0074	0.0043	0.0021	0.0034	-0.0046	-0.0036	-0.0031
	B_9	0.0083	0.0079	0.0064	0.0037	0.0018	0.0032	-0.0053	-0.004	-0.0034
$e=2$	B_0	-0.0003	-0.0019	-0.0008	0.0009	0.0007	0.0008	0.0009	0.0015	0.0002
	$B_{1,5}$	-0.0004	0.0006	-0.0009	0.0017	0.0004	0.0007	-0.0003	0.00002	-0.0011
	B_2	0.0883	0.0629	0.0474	0.0578	0.0412	0.0338	0.0494	0.0342	0.0244
	B_3	0.055	0.0305	0.0223	0.0387	0.0257	0.0168	0.0314	0.0188	0.0137
	B_4	0.0029	0.0045	0.0039	0.0074	0.0055	0.0043	0.0017	-0.0011	-0.0034
	B_6	0.0032	0.0044	0.0037	0.0066	0.0049	0.0039	0.0003	-0.0021	-0.0039
	B_7	0.0432	0.0223	0.0145	0.0258	0.0185	0.0135	0.0111	0.0039	0.0023
	B_8	0.0152	0.0089	0.0065	0.0097	0.008	0.0063	0.0002	-0.0022	-0.0029
	B_9	0.0104	0.0065	0.0051	0.0071	0.0063	0.0052	-0.0012	-0.0031	-0.0036
$e=3$	B_0	-0.002	0.0016	0.0004	-0.0006	0.00006	-0.0015	0.0023	0.0002	-0.0012
	$B_{1,5}$	-0.0022	0.001	0.0009	0.0012	0.00004	0.0004	0.00001	0.0029	0.0004
	B_2	0.1196	0.0875	0.0705	0.1209	0.0859	0.0693	0.0751	0.0686	0.0523
	B_3	0.0912	0.0583	0.0383	0.0936	0.0668	0.0487	-0.038	0.0508	0.0364
	B_4	-0.0465	-0.0392	-0.0341	-0.0377	-0.0347	-0.0302	-0.0293	-0.0327	-0.0304
	B_6	-0.0332	-0.0283	-0.0255	-0.0279	-0.0265	-0.0231	0.1123	-0.025	-0.0241
	B_7	0.1826	0.1468	0.1157	0.1345	0.1156	0.0974	0.0273	0.0964	0.0835
	B_8	0.0546	0.0451	0.0351	0.0369	0.0327	0.0281	0.0161	0.0252	0.0217
	B_9	0.0278	0.024	0.0184	0.0224	0.0205	0.0181	0.0191	0.0159	0.0138

Çizelge 4.5. $\mu=5, \sigma^2=1$ parametrelili ve $N=1000$ olan Normal dağılımlı kitle için yan değerleri

DPS $X_i = \lfloor X / e \rfloor$	ρ_{XY}	0.75			0.85			0.95		
	m	3	4	5	3	4	5	3	4	5
$e=1$	B_0	0.0022	0.00197	0.0007	-0.0007	-0.0022	0.0004	0.0025	0.0028	0.0019
	$B_{1,5}$	0.0017	-0.0027	0.00053	0.0021	-0.0011	-0.0005	0.0002	-0.0002	0.00005
	B_2	0.0022	0.00197	0.0007	-0.0007	-0.0022	0.0004	0.0025	0.0028	0.0019
	B_3	0.0022	0.00197	0.0007	-0.0007	-0.0022	0.0004	0.0025	0.0028	0.0019
	B_4	0.0032	0.000087	0.003	0.0043	0.0013	0.0027	-0.0004	-0.0006	0.0008
	B_6	0.0022	0.00197	0.0007	-0.0007	-0.0022	0.0004	0.0025	0.0028	0.0019
	B_7	0.0022	0.00197	0.0007	-0.0007	-0.0022	0.0004	0.0025	0.0028	0.0019
	B_8	0.0022	0.00197	0.0007	-0.0007	-0.0022	0.0004	0.0025	0.0028	0.0019
	B_9	0.0022	0.00197	0.0007	-0.0007	-0.0022	0.0004	0.0025	0.0028	0.0019
$e=2$	B_0	0.0024	0.0028	0.000494	-0.0013	0.0004	-0.0023	0.00027	0.0014	-0.0007
	$B_{1,5}$	-0.0012	-0.0012	0.0028	0.0015	-0.0017	-0.0009	0.0007	0.0035	-0.0006
	B_2	0.0024	0.0028	0.000494	-0.0013	0.0004	-0.0023	0.00027	0.0014	-0.0007
	B_3	0.0024	0.0028	0.000494	-0.0013	0.0004	-0.0023	0.00027	0.0014	-0.0007
	B_4	-0.0021	-0.00065	0.00492	0.0031	0.0004	0.0009	-0.00022	0.0438	-0.0008
	B_6	0.0024	0.0028	0.000494	-0.0013	0.0004	-0.0023	0.0008	0.0014	-0.0007
	B_7	0.0024	0.0028	0.000494	-0.0013	0.0004	-0.0023	0.00027	0.0014	-0.0007
	B_8	0.0024	0.0028	0.000494	-0.0013	0.0004	-0.0023	0.00027	0.0014	-0.0007
	B_9	0.0024	0.0028	0.000494	-0.0013	0.0004	-0.0023	0.00027	0.0014	-0.0007
$e=3$	B_0	-0.0006	0.0011	0.0001	0.001	-0.0014	-0.0023	0.0023	0.0007	0.0004
	$B_{1,5}$	-0.00083	0.000038	0.0011	0.0034	0.00006	0.0015	0.0008	0.0004	-0.0015
	B_2	-0.0006	0.0011	0.0001	0.001	-0.0014	-0.0023	0.0023	0.0007	0.0004
	B_3	-0.0006	0.0011	0.0001	0.001	-0.0014	-0.0023	0.0023	0.0007	0.0004
	B_4	0.0345	0.03071	0.0277	-0.0302	-0.0299	-0.0234	0.0416	0.0361	0.028
	B_6	-0.0006	0.0011	0.0001	0.001	-0.0014	-0.0023	0.0023	0.0007	0.0004
	B_7	-0.0006	0.0011	0.0001	0.001	-0.0014	-0.0023	0.0023	0.0007	0.0004
	B_8	-0.0006	0.0011	0.0001	0.001	-0.0014	-0.0023	0.0023	0.0007	0.0004
	B_9	-0.0006	0.0011	0.0001	0.001	-0.0014	-0.0023	0.0023	0.0007	0.0004

GE₀, BRÖ' nün basit ortalama tahmininin ($\hat{\mu}_0$) SKÖ' nün klasik ortalama tahminine ($\hat{\mu}_1$) karşı etkinliğini gösterir. Çizelge 4.2 ve Çizelge 4.3'ten, büyük ve küçük kitle simülasyon çalışmalarında her farklı korelasyon, küme büyüklüğü ve bağ durumunda değeri birden küçüktür.

BRÖ' nün basit oransal tahmin edicisinin ($\hat{\mu}_2$), SKÖ' nün klasik ortalama tahmin edicisine ($\hat{\mu}_1$) karşı etkinliği incelendiğinde; Çizelge 4.3'e göre, $N=50$ olan simülasyon çalışmasında, $e=1$ için korelasyon değeri 0.85 ve 0.95 iken, $e=2$ için korelasyon 0.95 iken yani yüksek korelasyon ile düşük ve orta düzeydeki bir bağ sayısında $\hat{\mu}_2$ 'nin $\hat{\mu}_1$ 'e göre daha etkin olduğu görülmüştür. Çizelge 4.2'de $N=1000$ olan kitle üzerinden yapılan simülasyon çalışmasında ise, $e=1$ olan tüm korelasyon ve küme büyüklüğü değerlerinde de $\hat{\mu}_2$ tahmin edicisi daha etkindir.

SKÖ' de oransal tahmin edicinin ($\hat{\mu}_3$) klasik ortalama tahmin edicisine göre etkinliğine (GE₃) bakıldığında, büyük ve küçük kitle simülasyon çalışmalarına göre, Çizelge 4.2 ve Çizelge 4.3'de $e=1$ olduğu durumlarda GE değerinin birden büyük olduğu görülür. SKÖ' nün basit oransal tahmin edicisi yüksek korelasyon ve düşük bağ sayısında iyi sonuçlar vermektedir. Fakat bağ sayısı arttığında, $e=3$ için Çizelge 4.2 ve Çizelge 4.3'de, etkinliği ciddi şekilde azalmaktadır.

Bağ matrisini kullanan ve SKÖ-b' de en basit tahmin edici olan $\hat{\mu}_4$ 'ün etkinliği incelendiğinde, büyük kitle simülasyon çalışmasında, tüm durumlarda SKÖ' de bağ bilgisi kullanmayan klasik ortalama tahmin edicisinden daha etkin olduğu görülmüştür. Küçük kitle üzerinden yürütülen simülasyon çalışmasında ise, Çizelge 4.3'ten, düşük korelasyon altındaki bazı durumlarda GE₄ değerinin yaklaşık bir, diğer tüm durumlarda birden büyük olduğu görülür.

Bağ bilgisini kullanan $\hat{\mu}_4$ tahmin edicisinin izotonik versiyonu $\hat{\mu}_6$ 'nın etkinliğine bakıldığında, iki simülasyon çalışmasında da SKÖ basit ortalama tahmininden daha etkin bir tahmin edici olduğu görülür. GE₆ değeri her zaman birden büyüktür. Fakat $\hat{\mu}_6$ tahmin edicisi, hiçbir düşük/yüksek korelasyon ve bağ kombinasyonunda en etkin tahmin edici olarak bulunmamıştır.

Önerilen ilk izotonik oransal tahmin edici $\hat{\mu}_7$ ' nin etkinliđi GE_7 , büyük ve küçük kitle simülasyon çalışma sonuçlarına göre, korelasyonun düşük ve bađın yüksek olduđu sütunlarda ciddi derecede azalmaktadır. Önerilen tahmin edicinin en küçük etkinlik deđeri Çizelge 4.3'teki küçük kitle simülasyon çalışmasında, küme büyüklüğü 3, korelasyon 0.75 ve $e=3$ olduđunda yani düşük korelasyon ve küme büyüklüğü ile yüksek bađ sayısında 0.4562 olarak hesaplanmıřtır. Tahmin edicinin en etkin olduđu durum ise, yine Çizelge 4.3'teki küçük kitle simülasyon çalışmasında korelasyonun 0.95, küme büyüklüğünün 3 ve $e=1$ olduđundadır ve 3.7329 olarak bulunmuřtur.

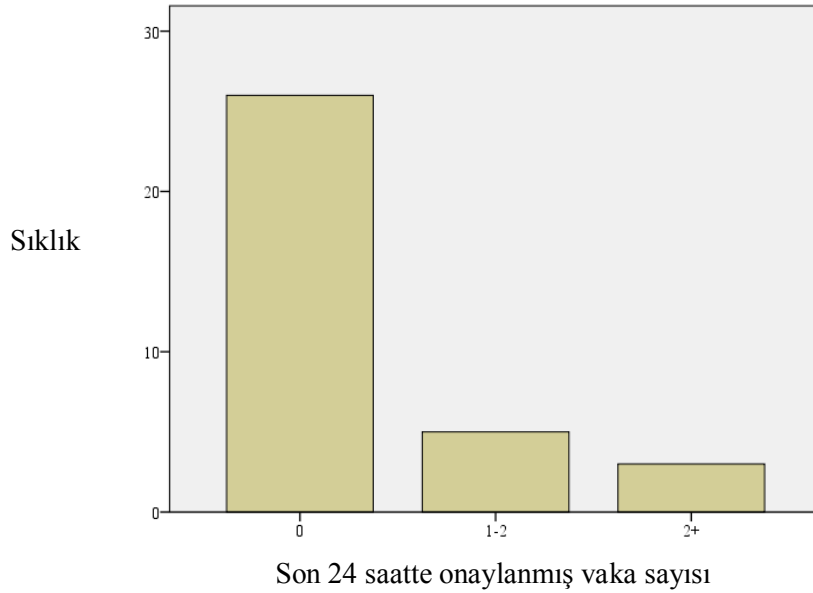
Önerilen ikinci üstel oransal tahmin edicisi $\hat{\mu}_8$, büyük ve küçük kitle simülasyon çalışmasında, Çizelge 4.2 ve Çizelge 4.3'de her zaman SKÖ'nün ortalama tahmininden daha etkindir. $\hat{\mu}_8$ tahmin edicisi düşük korelasyon, düşük bađ ve küçük küme büyüklüğü kořullarındaki en iyi tahmin edicidir. Simülasyon çalışmalarında bu tahmin edici, en yüksek GE deđerini Çizelge 4.2'deki büyük kitle simülasyon çalışmasında, yüksek korelasyon, küme büyüklüğü 3 ve $e=1$ kořullarında 2.8 olarak almıřtır. En az etkin olduđu durum ise Çizelge 4.3'teki küçük kitle simülasyon çalışmasındaki düşük korelasyon, küme büyüklüğü 3 ve $e=3$ kořullarıdır.

$\hat{\mu}_9$ tahmin edicisi özellikle bađ sayısının en yüksek olduđu durumlarda en etkin tahmin edici olarak görölmüřtür. Korelasyonun en yüksek ve bađ sayısının en az olduđu durum haricinde ve tahmin edici içerisinde $\alpha=1$, $\beta=0$ olarak alındıđında $\hat{\mu}_8$ tahmin edicisine de dönüřebildiđi düşünöldüđünde, Çizelge 4.2 ve Çizelge 4.3'den büyük ve küçük kitle simülasyon çalışmalarındaki en etkili tahmin edici olduđu söylenebilir.

Kitlenin büyüklüğünden ya da bađ sayısı, korelasyon katsayısı ve küme büyüklüğü gibi faktörlerden bađımsız olarak tüm tahmin edicilerin yanları yaklaşık olarak sıfır bulunmuřtur.

5. GERÇEK VERİ UYGULAMASI

Kesim 3.5'te tahmin edicilerin, Çin'in şehirlerine/bölgelerine ait COVID-19 verilerini kullanarak, bir örneklem çekimi üzerinden hesaplamaları yapılmıştır. Bu bölümde, Kesim 3.5'te kullanılan değişkenler ve aynı kitle ($N=34$) tekrar ele alınmıştır. Son 24 saatte hastalığı onaylanan vaka sayısının yardımcı değişken, ülkede hastalıktan ölen sayısı ilgilenilen değişken olmak üzere, kitlede yardımcı değişken üzerinde bağlar olduğu Şekil 5.1'deki bar grafiğinde görülebilir. Verilen grafiğe göre, örneğin Çin'de son 24 saatte onaylanmış vakası olmayan 26 adet şehir vardır. Değişkenlerimiz hasta sayıları olduğu için, hiç hasta olmayabileceğini yani değişkenlerin sıfır değerini alabileceği açıktır.



Şekil 5.1. Çin'e ait COVID-19 verisindeki yardımcı değişken histogram grafiği

R programı kullanılarak yazılan simülasyonda, parametreler $\rho = 0.75, 0.85$ ve 0.95 ; $m = 3, 4$ ve 5 olarak belirlenmiş verilen ortalama tahmin edicileri $N=34$ olan kitlede yerine konmadan simülasyon sayısı 100000 için dengeli SKÖ tasarımı altında elde edilmiştir. $\hat{\mu}_0$ tahmin edicisinin katsayıları değiştirildiğinde GE değerinde önemli bir değişiklik

olmadığı görülmüş ve bu nedenle katsayılar rastgele seçilip, α değeri X_T yardımcı değişkeninin basıklık katsayısı, β değeri 1 olarak alınmıştır. Referans tahmin edicisi $\hat{\mu}_1$ olarak alınarak, GE değerleri Eşitlik (47) ile hesaplanmış ve sonuçları Çizelge 5.1'de, yan değerleri ise Eşitlik (9) kullanılarak hesaplanmış ve sonuçları Çizelge 5.2'de verilmiştir. $\hat{\mu}_5$ tahmin edicisi simülasyon çalışmasında olduğu gibi $\hat{\mu}_1$ 'e eşit olduğundan Çizelge 5.1 ve 5.2'de yer almamaktadır.

Çizelge 5.1. Çin'e ait COVID-19 verisi için göreceli etkinlik değerleri (N=34, $\rho=0.84$)

	<i>m=3</i>	<i>m=4</i>	<i>m=5</i>
<i>GE</i> ₀	1.19	1.14	1.068
<i>GE</i> ₂	-	-	-
<i>GE</i> ₃	-	-	-
<i>GE</i> ₄	1.67	1.57	1.47
<i>GE</i> ₆	1.67	1.58	1.47
<i>GE</i> ₇	2.52	1.58	1.46
<i>GE</i> ₈	3.60	2.87	2.54
<i>GE</i> ₉	3.63	2.89	2.55

Çizelge 5.2. Çin'e ait COVID-19 verisi için yan değerleri (N=34, $\rho=0.84$)

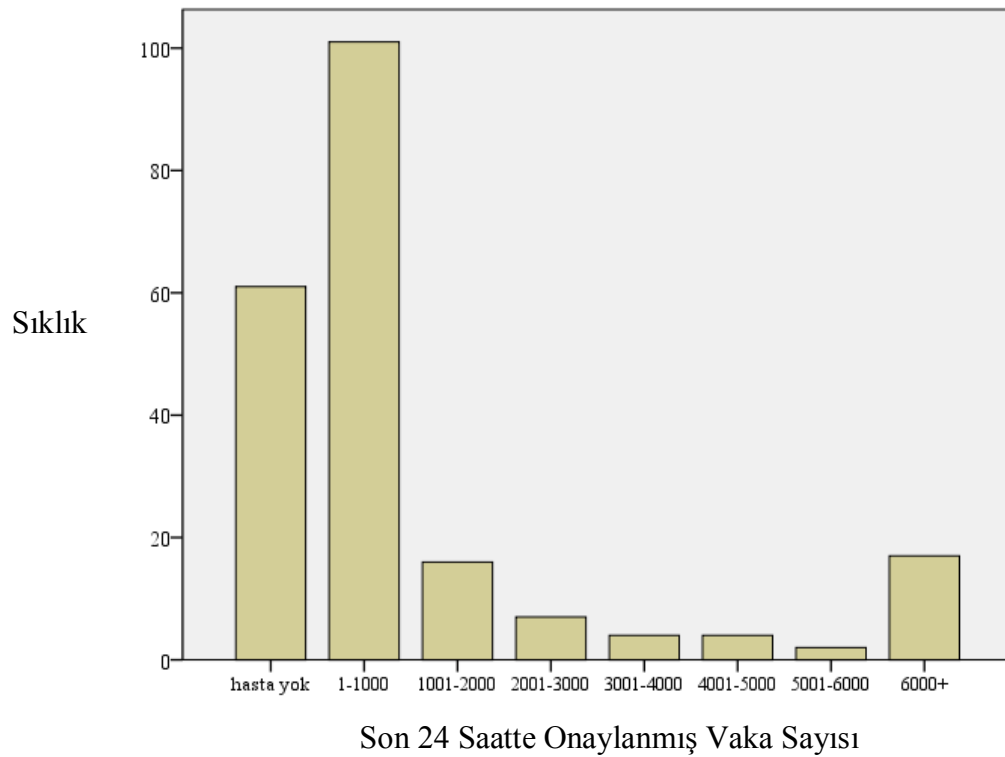
	<i>m=3</i>	<i>m=4</i>	<i>m=5</i>
<i>B</i> ₀	1.35	-0.25	1.55
<i>B</i> _{1 ve 5}	20.83	14.97	9.19
<i>B</i> ₂	-	-	-
<i>B</i> ₃	-	-	-
<i>B</i> ₄	-5.71	-7.58	-9.14
<i>B</i> ₆	-5.71	-7.56	-9.11
<i>B</i> ₇	-50.18	-40.84	-39.41
<i>B</i> ₈	-36.01	-32.22	-31.14
<i>B</i> ₉	-35.7	-32.01	-30.83

İkinci gerçek kitle olarak, AB Açık Veri Portalı (EU Open Data Portal) tarafından yayınlanan, dünya üzerindeki tüm ülkelere ait 14 Aralık 2020 gününe ait bildirilen doğrulanmış ve şüpheli hasta sayıları kullanılmıştır (Data Europa, 2020). Verilen kitlede yine hastalıktan ölen hasta sayısını tahmin etmeye çalıştığımızı varsayalım. Hastalığa yakalanan vaka sayısı ile vefat eden hasta sayıları arasında pozitif yüksek korelasyon olduğu görülmüştür. Bu nedenle, yardımcı değişken olarak son 24 saatte onaylanmış vaka sayısı kullanılmıştır. Kitlede yardımcı değişken üzerinde bağlar olduğu Şekil 5.2'deki bar grafiğinde görülebilir. Örneğin, verilen grafiğe göre, hasta yok sütunun

sıklık değerinden, Dünya üzerinde son 24 saatte onaylanmış vakası olmayan 60 ülke vardır. Kitleye ait parametreler Çizelge 5.3'te verilmiştir.

Çizelge 5.3. Dünya geneli COVID-19 verileri için kitle parametreleri

$N=212$	$\beta_{XT}= 166.6721$	$\bar{X}_T = 2550.278$
$\sigma_Y^2=13213.4,$	$\rho_{XTY}=0.923$ ($p\text{-value}<0.05$)	$\bar{Y} = 33.4198$
$\sigma_{XT}^2= 187381816$	Min. $X_T=0$, Max. $X_T = 189723$	Min. $Y=0$, Max. $Y=1340$



Şekil 5.2. Dünya geneli COVID-19 verisindeki yardımcı değişken histogram grafiği

Tüm hesaplamalar Çin'e ait COVID-19 verilerine yapıldığı şekilde yapılmış ve GE değeri sonuçları Çizelge 5.4'te, yan değerleri ise Çizelge 5.5'te verilmiştir.

Çizelge 5.4. Dünya geneli COVID-19 veri seti için görelî etkinlik deęerleri (N=212, $\rho=0.92$)

	$m=5$	$m=4$	$m=3$
GE_0	0.8202	0.8516	0.8896
GE_2	-	-	-
GE_3	1.9067	-	-
GE_4	1.0055	1.0103	1.0265
GE_6	1.0052	1.01	1.0262
GE_7	1.9471	-	-
GE_8	3.6518	3.678	3.8617
GE_9	3.6518	3.678	3.8617

Çizelge 5.5. Dünya geneli COVID-19 veri seti için yan deęerleri (N=212, $\rho=0.92$)

	$m=5$	$m=4$	$m=3$
B_0	0.1971	0.012	0.0432
$B_{1\ ve\ 5}$	-0.0298	0.0053	0.0646
B_2	-	-	-
B_3	12.9492	-	-
B_4	-0.3377	-0.4308	-0.6132
B_6	-0.229	-0.3088	-0.4951
B_7	12.5865	-	-
B_8	-1.4165	-2.4914	-4.1888
B_9	-1.4165	-2.4914	-4.1888

GE_0 deęeri sadece Çizelge 5.1'deki Çin'e ait COVID-19 gerek veri setinde birden büyük ıkmıřtır. Büyük ve küçük kitle simülasyon alıřmalarında olduęu gibi ve dünya geneli COVID-19 verilerindeki her farklı korelasyon, küme büyüklüęü ve baę durumunda deęeri birden küçüktür. Bu nedenle, $\hat{\mu}_0$ 'ın etkili bir tahmin edici olmadıęı söylenebilir.

GE_2 deęeri simülasyon alıřmalarında elde edilebilmiřken gerek veri setlerinde $\hat{\mu}_2$ oransal tahmin edicisi hesaplanamadıęı için GE_2 deęeri de bulunamamıřtır. Bunun nedeni simülasyon alıřmalarında kullanılan kitlenin $N(5,1)$ olmasıdır. Oransal tahmin edicilerin payda kısmındaki örneklem ortalaması sıfır deęerini aldıęında tahmin ediciler hesaplanamamaktadır. Kesim 4.2'de bahsedildięi üzere oransal tahmin edicilerin hesaplanabilmesi için bu parametreler seilmiřtir.

GE_3 deęerlerine baktıęımızda ise, BRÖ oransal tahmin edicisi $\hat{\mu}_2$ 'de olduęu gibi $\hat{\mu}_3$ oransal tahmin edicisinin de hesaplanamadıęı durumlar görölmüřtür.

Bağ matrisini kullanan ve SKÖ-b' de en basit tahmin edici olan $\hat{\mu}_4$ her iki gerçek veri çalışmasında da, $\hat{\mu}_1$ 'den daha etkindir.

$\hat{\mu}_6$ 'nın etkinliğine bakıldığında, SKÖ basit ortalama tahmininden daha etkin bir tahmin edici olduğu görülür. GE_6 değeri, simülasyon çalışmasında olduğu gibi, her zaman birden büyüktür. Fakat $\hat{\mu}_6$ tahmin edicisi, hiçbir düşük/yüksek korelasyon ve bağ kombinasyonunda en etkin tahmin edici olarak bulunmamıştır.

$\hat{\mu}_7$ tahmin edicisi SKÖ basit ortalama tahmininden daha etkindir. Fakat Çizelge 5.4'e göre diğer oransal tahmin edicilerde olduğu gibi hesaplanamadığı görülmüştür. $\hat{\mu}_7$ tahmin edicisi genel olarak diğer oransal tahmin ediciler gibi düşük korelasyon altında etkin değildir ve sıfır değerini alabilen veriler üzerinde uygulandığında hesaplanamayacağı durumlarla karşılaşılabilir.

Önerilen ikinci üstel oransal tahmin edicisi $\hat{\mu}_8$, Çin'e ait COVID-19 verisinde ikinci en iyi, dünya genelindeki COVID-19 verisinde en iyi tahmin edici olarak bulunmuştur. İki gerçek veri çalışmasında da en etkin olan tahmin edicinin $\hat{\mu}_9$ olduğu görülmüştür.

Yan değerleri incelendiğinde, simülasyon çalışmasının aksine gerçek veri çalışmalarında önerilen tahmin ediciler yanlıdır. Özellikle $N=34$ olan kitle üzerinde hesaplanan yanların, $N=212$ olan kitleden daha büyük olduğu görülmüştür. Bu durum yanın kitle büyüklüğünden etkilendiğini gösterir. Gerçek verilerle yapılan çalışmalarda en etkin tahmin edici olmasına rağmen önerilen $\hat{\mu}_9$ tahmin edicisi, önerilen $\hat{\mu}_7$ ve $\hat{\mu}_8$ tahmin edicileri gibi yanlıdır.

6. SONUÇ VE TARTIŞMA

Bu çalışmada SKÖ-b durumu için $\hat{\mu}_7$, $\hat{\mu}_8$ ve $\hat{\mu}_9$ olmak üzere üç yeni tahmin edici önerilmiştir. Önerilen üçüncü tahmin edici $\hat{\mu}_9$, detayları önceki bölümlerde verilen tüm simülasyon ve gerçek veri çalışmalarında en etkin tahmin edici olarak görülmüştür. $\hat{\mu}_9$ tahmin edicisi oransal tahmin edici olmasına rağmen düşük korelasyonlarda da etkili sonuçlar vermektedir. $\hat{\mu}_8$ tahmin edicisine göre avantajı bağ sayısının yüksek olduğu yerlerde $\hat{\mu}_9$ tahmin edicisinin daha etkin olmasıdır. Diğer tahmin edicilerin etkinlikleri korelasyona, küme büyüklüğüne ya da bağ sayısına göre tutarsızlaşırken $\hat{\mu}_9$ tahmin edicisinin etkinliğinde ani düşüş ve yükselişler görülmemiştir. Gerçek veri çalışmalarında en etkin tahmin edici olarak hesaplanmıştır. Bunun nedeni gerçek verilerdeki bağ sayısının çok yüksek olmasıdır.

Kesim 4.2'de bahsedildiği üzere, küçük kitleler üzerinde uygulanan SKÖ'de birimlerinin örnekleme seçilme olasılıkları aynı olmadığı için küçük ve büyük kitlelerde yapılan hesaplamalarının farklı sonuçlar verdiği bilgisi bu tez çalışması sonuçlarında da gözlemlenmiştir. Küçük kitlede e değeri değiştikçe, korelasyon katsayısı büyük kitledekinden daha fazla değişir. Korelasyon katsayısındaki bu tutarsızlığa rağmen, önerilen $\hat{\mu}_9$ tahmin edicisinin göreceli etkinliği etkilenmemiştir. Küçük kitlede korelasyon azaldıkça diğer oran tahmin edicilerinin ve önerilen ilk izotonik oransal tahmin edicinin bozulduğu görülmektedir.

Simülasyon çalışmalarından elde edilen sonuçlarda, tüm tahmin edicilerin yan değeri yaklaşık olarak sıfır bulunmuştur. Fakat, gerçek verilerden elde edilen sonuçlarda, bazı yan değerleri sıfırdan farklıdır. Bunun nedeni, gerçek verilerin herhangi bir dağılıma uymaması ve aşırı çarpık olması, kitlelelerin yeterli ölçüde büyük olmaması, kitlelerdeki bağ sayısının çok yüksek olması ve aykırı değerler içerebilmesidir. Bu sonuçlara bağlı olarak, önerilen tahmin edicilerin küçük kitleler üzerinde yanlı olduğu söylenebilir. Tahmin edicilerin yanlılığını azaltabilmek için küçük kitlelerde düzeltme katsayısı uygulanması düşünülebilir.

Sonuç olarak, simülasyon ve gerçek veri çalışmasının sonuçlarına göre önerilen $\hat{\mu}_8$ ve $\hat{\mu}_9$ tahmin edicilerinin, literatürde bağ durumu için önerilen tahmin edicilerden daha iyi olduğu görülmektedir. Beklenildiği üzere, bağ sayısı artıp korelasyon düştüğünde oransal tahmin ediciler kötü performans göstermiştir. Ancak önerilen tahmin ediciler, oransal tahmin edici olmalarına rağmen düşük korelasyonlarda en iyi sonuçları vermektedir. Yüksek bağ sayısı da tahmin edicileri önemli ölçüde etkilemektedir. Diğer tahmin edicilerin etkinliklerinin ciddi şekilde azaldığı bu koşullarda, önerilen $\hat{\mu}_9$ tahmin edicisi daha kararlıdır. Oransal tahmin edicilerin hesaplanamadığı, örneklem ortalamasının sıfır olduğu değişkenlerle çalışmak gibi veya COVID-19 veri setlerinde olduğu gibi yardımcı değişkenin sıfır değerini alabileceği durumlarda önerilen bu tahmin ediciler gerçekçi sonuçlar vermektedir.

7. KAYNAKLAR

- Bahl, S., Tuteja, R., Ratio and product type exponential estimators. *Journal of information and optimization sciences*, 12(1), 159-164, 1991.
- Beale, E. M. L., Some use of computers in operational research, *Industrielle Organisation*, 31, 27-28, 1962.
- Chen, H., Stasny, E. A., Wolfe, D. A., Ranked set sampling for efficient estimation of a population proportion, *Statistics in Medicine*, 24:3319–3329, 2005.
- Cingi, H., Kadilar, C., *Advances in sampling theory-ratio method of estimation*. Bentham Science Publishers, 2009.
- Cobby, J. U., Ridout, M. S., Bassett P. J. and Large, R. V., An Investigation into Use of Ranked Set Sampling on Grass-Clover Swards, *Grass and Forage Science*, Vol.40, 257-263, 1985.
- Cochran, W. G., *Sampling Techniques*, John Wiley and Sons, New-York, 1977.
- Çingil, H. *Örnekleme Kuramı*, HÜ Fen Fakültesi Basımevi. 2009.
- Dell, T. R., Clutter, J. L., Ranked-set sampling theory with order statistics background, *Biometrics*, 28:545–555, 1972.
- Dell, T. R., Clutter J. L., Ranked-set sampling theory with order statistics background, *Biometrics*, 28 (2):545–55, 1972.
- Eu Open Data Portal, <https://data.europa.eu/euodp/data/dataset/covid-19-coronavirus-data>, Doi: 10.2906/101099100099/1 (Erişim Tarihi: 29 Aralık 2020)
- Fligner, M. A., MacEachern, S. N., Nonparametric two-sample methods for ranked-set sample data, *Journal of the American Statistical Association*, 101:1107–1118, 2006.
- Frey, J. , Nonparametric mean estimation using partially ordered sets, *Environmental and Ecological Statistics*, 19 (3):309–326, 2012.
- Frey, J., Feeman, T. G., An improved mean estimator for judgment post-stratification. *Computational Statistics & Data Analysis* 56 (2):418–26, 2012.

- Frey, J., Feeman, T. G., Variance estimation using judgment post-stratification. *Annals of the Institute of Statistical Mathematics* 65 (3):551–69, 2013.
- Frey, J., Recursive computation of inclusion probabilities in ranked-set sampling, *Journal of Statistical Planning and Inference* 141, pp.3632-3639, 2011.
- Haki, D., İstatistiksel tahminlemede sıralı küme örnekleme ve uygulamaları, Yüksek Lisans Tezi, Marmara Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 2012.
- Halls, L. K., Dell, T. R., Trial of ranked-set sampling for forage yields, *Forest Science*, 12:22–26, 1966.
- İnal, C., Günay, S., Olasılık ve Matematiksel İstatistik, Hacettepe Üniversitesi Fen Fakültesi Yayınları, Ankara, 2009.
- Kadilar, C., Cingi, H., Ratio estimators in stratified random sampling. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 45(2), 218-225, 2003.
- Karakulah, Ü. H., Basit Rastgele Örnekleme Yönteminde Oransal Tahmin Ediciler, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 2006.
- Kvam, P. H., Ranked set sampling based on binary water quality data with covariates. *Journal of Agricultural, Biological, and Environmental Statistics* 8:271–279, 2003.
- MacEachern, S. N., E. A. Stasny, D. A. Wolfe., Judgement post-stratification with imprecise rankings, *Biometrics*, 60 (1):207–15, 2004.
- Mahdizadeh, M., Zamanzade, E., Efficient body fat estimation using multistage pair ranked set sampling, *Statistical Methods in Medical Research*, 28 (1):223-234, 2019.
- Mair, P., Hornik, K., Leeuw, J., Isotone optimization in R: pool-adjacent-violators algorithm (PAVA) and active set methods, *Journal of statistical software*, 32(5),1-24, 2009.
- McIntyre G. A., A method for unbiased selective sampling, using ranked sets, *Australian Journal of Agricultural Research*, 3:385–390, 1952.

- Nussbaum, B. D., Sinha, B. K., Cost effective gasoline sampling using ranked set sampling, *ASA Proceedings of the Section on Statistics and the Environment*, 83-87, 1997.
- Ozturk, O., Statistical inference under a stochastic ordering constraint in ranked set sampling, *J Nonparametr Stat*, 19:131–144, 2007.
- Ozturk, O., Sampling from partially rank-ordered sets. *Environmental and Ecological Statistics*, 18 (4):757–779, 2011.
- Ozturk, O., Bilgin, O. C., Wolfe, D. A., Estimation of population mean and variance in flock management: a ranked set sampling approach in a finite population setting. *Journal of Statistical Computation and Simulation*, 75(11), 905-919, 2005.
- Özdemir, Y. A., Sıralı küme örneklemesiyle doğrusal regresyon modelinde parametre tahminlerinin incelenmesi, *Doktora Tezi, Gazi Üniversitesi, Ankara*, 2005.
- Patil, G. P., Sinha, A. K., Taillie, C., Ranked set sampling, *Handbook of Statistics*, 12, 167-200, 1994.
- Patil, G. P., Editorial: ranked set sampling, *Environmental and Ecological Statistics*, 2:271–285, 1995.
- Prasad, B., Some improved ratio type estimators of population mean and ratio in finite population sample surveys, *Communications in Statistics-Theory and Methods*, 18:379–392, 1989.
- Quenouille, M. H., Notes on bias in estimation, *Biometrika*, 43(3/4), 353-360, 1956.
- Ray, S. K., Singh, R. K., Difference cum-ratio type estimators, *Journal of Indian Statistical Association*, 19, 147-151, 1981.
- Robertson, T., Order restricted statistical inference (No. 04; QA278. 7, R6.), 1988.
- Samawi, H. M., Al-Sagheer, O. A. M., On the estimation of the distribution function using extreme and median ranked set sampling, *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 43 (3):357–373, 2001.
- Samawi, H. M., Muttlak H. A., Estimation of ratio using rank set sampling, *Biometrical Journal*, 38:753–764, 1996.

- Sampath, S., Durairajan, T.M., An efficient ratio type estimator, *Journal of the Indian Statistical Association*, 26, 67-71, 1988.
- Semiz, M., *Örnekleme Yöntemleri*, Dizgi Ofset Matbaacılık, Konya, 2007.
- Shabbir, J., Yaab, M. Z., Improvement over transformed auxiliary variable in estimating the finite population mean, *Biometrical Journal*, 45, 723-729, 2003.
- Shukla, N. D., Systematic sampling and product method of estimation, In *Proceedings of all India Seminar on Demography and Statistics*, B.H.U. Varanasi: India, 1971.
- Singh, H. P., Tailor, R., Jatwa, N. K., Modified ratio and product estimators for population mean in systematic sampling, *Journal of Modern Applied Statistical Methods*, 10 (2):4, 2011.
- Singh, H. P., Tailor, R., Use of known correlation coefficient in estimating the finite population mean, *Statistics in Transition*, 6, 4, 555-560, 2003.
- Singh, R., Chauhan, P., Sawan, N., Smarandache, F., Improvement in estimating the population mean using exponential estimator in simple random sampling, *Bulletin of Statistics&Economics*, 3(A09), 13-18, 2009.
- Stokes, S. L., Estimation of variance using judgment ordered ranked set samples, *Biometrics*, 35-42, 1980.
- Takahasi, K. and Wakimoto, K., On unbiased estimates of the population mean based on the sample stratified by means of ordering, *Annals of the Institute of Statistical Mathematics*, 20, 1-31, 1968.
- Tin, M., Comparison of some estimators, *Journal of American Statistical Association*, 60, 294-307, 1965.
- Upadhyaya, L. N., Singh, H. P., Use of transformed auxiliary variable in estimating the finite population mean, *Biometrical Journal*, 41, 5, 627-636, 1999.
- Yamane, T., Esin, A., *Temel örnekleme yöntemleri*, Literatür Yayıncılık 2001.
- Wang, X., Lim, J, Stokes, S. L., A nonparametric mean estimator for judgment post-stratified data, *Biometrics* 64:355–363, 2008.

World Health Organization, Coronavirus disease 2019 (COVID-19) situation report–51. Geneva, Switzerland: World Health Organization; 2020. Accessed August 16, 2020.

Zamanzade, E., Wang, X., Improved Nonparametric Estimation Using Partially Ordered Sets, *Statistical Methods and Applications in Forestry and Environmental Sciences*, 57-77, 2020(a).

Zamanzade, E., Wang, X., Estimating the area under a receiver operating characteristic curve using partially ordered sets, To appear in *International Journal of Biostatistics*, 2020(b).

Zamanzade, E., X. Wang., Proportion estimation in ranked set sampling in the presence of tie information, *Computational Statistics*, 33 (3):1349–66, 2018.

EKLER

EK 1 – 11 Mart 2020 tarihli Çin'e ait COVID-19 veri seti

NO	Bölge/ Şehir	Onaylanmış Vaka	Toplam Ölüm
1	Hubei	13	3046
2	Guangdong	0	8
3	Henan	0	22
4	Zhejiang	0	1
5	Hunan	0	4
6	Anhui	0	6
7	Jiangxi	0	1
8	Shandong	1	6
9	Jiangsu	0	0
10	Chongqing	0	6
11	Sichuan	0	3
12	Heilongjiang	1	13
13	Beijing	6	8
14	Shanghai	2	3
15	Hebei	0	6
16	Fujian	0	1
17	Guangxi	0	2
18	Shaanxi	0	1
19	Yunnan	0	2
20	Hainan	0	6
21	Guizhou	0	2
22	Tianjin	0	3
23	Shanxi	0	0
24	Liaoning	0	1
25	Gansu	1	2
26	Hong Kong SAR	5	3
27	Jilin	0	1
28	Xinjiang	0	3
29	Ningxia	0	0
30	Inner Mongolia	0	1
31	Taipei and environs	2	1
32	Qinghai	0	0
33	Macao SAR	0	0
34	Xizang	0	0

EK 2 – 14 Aralık 2020 tarihli dünya geneli COVID-19 veri seti

Ülke/Bölge	Vaka Sayısı	Ölü Sayısı
Afghanistan	746	6
Albania	788	14
Algeria	464	12
Andorra	50	1
Angola	27	5
Anguilla	0	0
Antigua_and_Barbuda	0	0
Argentina	3558	98
Armenia	357	17
Aruba	5	0
Australia	6	0
Austria	3005	45
Azerbaijan	4451	39
Bahamas	15	0
Bahrain	178	0
Bangladesh	1355	32
Barbados	0	0
Belarus	1961	9
Belgium	2	6
Belize	122	3
Benin	0	0
Bermuda	40	0
Bhutan	1	0
Bolivia	120	4
Bonaire, Saint Eustatius and Saba	0	0
Bosnia_and_Herzegovina	811	38
Botswana	0	0
Brazil	21825	279

Ülke/Bölge	Vaka Sayısı	Ölü Sayısı
British_Virgin_Islands	0	0
Brunei_Darussalam	0	0
Bulgaria	497	62
Burkina_Faso	136	0
Burundi	1	0
Cambodia	0	0
Cameroon	0	0
Canada	5891	81
Cape_Verde	55	0
Cayman_Islands	0	0
Central_African_Republic	0	0
Chad	19	0
Chile	2138	40
China	12	0
Colombia	8702	187
Comoros	0	0
Congo	0	0
Costa_Rica	0	0
Cote_dIvoire	41	0
Croatia	3363	78
Cuba	69	0
Curaçao	48	1
Cyprus	301	1
Czechia	2000	74
Democratic_Republic_of_the_Congo	52	4
Denmark	2642	6
Djibouti	3	0
Dominica	0	0
Dominican_Republic	1107	1

Ülke/Bölge	Vaka Sayısı	Ölü Sayısı
Ecuador	586	1
Egypt	486	22
El_Salvador	486	5
Equatorial_Guinea	0	0
Eritrea	0	0
Estonia	342	1
Eswatini	54	0
Ethiopia	472	3
Falkland_Islands_(Malvinas)	0	0
Faroe_Islands	3	0
Fiji	2	0
Finland	360	0
France	11533	150
French_Polynesia	0	0
Gabon	0	0
Gambia	2	0
Georgia	1337	49
Germany	16362	188
Ghana	81	0
Gibraltar	3	0
Greece	692	85
Greenland	0	0
Grenada	26	0
Guam	16	1
Guatemala	183	18
Guernsey	0	0
Guinea	0	0
Guinea_Bissau	0	0
Guyana	41	1
Haiti	74	1
Holy_See	0	0

Ülke/Bölge	Vaka Sayısı	Ölü Sayısı
Honduras	316	1
Hungary	3470	165
Iceland	5	0
India	27071	336
Indonesia	6189	166
Iran	7451	247
Iraq	1012	14
Ireland	429	1
Isle_of_Man	0	0
Israel	1844	16
Italy	17937	484
Jamaica	102	2
Japan	2366	23
Jersey	91	0
Jordan	2339	30
Kazakhstan	764	0
Kenya	366	1
Kosovo	462	12
Kuwait	174	0
Kyrgyzstan	318	5
Laos	0	0
Latvia	629	25
Lebanon	1275	10
Lesotho	0	0
Liberia	0	0
Libya	899	21
Liechtenstein	12	0
Lithuania	1920	10
Luxembourg	517	4
Madagascar	0	0
Malawi	3	1

Ülke/Bölge	Vaka Sayısı	Ölü Sayısı
Malaysia	1229	4
Maldives	20	0
Mali	93	5
Malta	0	0
Marshall_Islands	0	0
Mauritania	0	0
Mauritius	0	0
Mexico	8608	249
Moldova	795	25
Monaco	0	0
Mongolia	5	0
Montenegro	478	4
Montserrat	0	0
Morocco	2012	35
Mozambique	142	2
Myanmar	1127	23
Namibia	267	0
Nepal	830	9
Netherlands	9868	29
New_Caledonia	0	0
New_Zealand	0	0
Nicaragua	0	0
Niger	59	0
Nigeria	418	3
North_Macedonia	613	25
Northern_Mariana_Islands	0	0
Norway	0	0
Oman	571	8
Pakistan	2362	36
Palestine	1639	17
Panama	2422	25

Ülke/Bölge	Vaka Sayısı	Ölü Sayısı
Papua_New_Guinea	5	0
Paraguay	669	16
Peru	1928	68
Philippines	1069	3
Poland	8976	188
Portugal	4044	98
Puerto_Rico	1104	6
Qatar	0	0
Romania	4435	121
Russia	28080	488
Rwanda	131	0
Saint_Kitts_and_Nevis	0	0
Saint_Lucia	1	0
Saint_Vincent_and_the_Grenadines	4	0
San_Marino	15	0
Sao_Tome_and_Principe	0	0
Saudi_Arabia	139	12
Senegal	78	0
Serbia	4995	56
Seychelles	0	0
Sierra_Leone	0	0
Singapore	7	0
Sint_Maarten	0	0
Slovakia	5897	53
Slovenia	833	11
Solomon_Islands	0	0
Somalia	0	0
South_Africa	7999	170
South_Korea	718	7
South_Sudan	0	0
Sri_Lanka	655	3

Ülke/Bölge	Vaka Sayısı	Ölü Sayısı
Sudan	239	3
Suriname	4	0
Sweden	0	0
Switzerland	0	0
Syria	125	12
Taiwan	4	0
Tajikistan	38	0
Thailand	28	0
Timor_Leste	0	0
Togo	30	0
Trinidad_and_Tobago	15	0
Tunisia	968	58
Turkey	29136	222
Turks_and_Caicos_islands	0	0
Uganda	461	1
Ukraine	15627	249
United_Arab_Emirates	1194	8
United_Kingdom	18447	144
United_Republic_of_Tanzania	0	0
United_States_of_America	189723	1340
United_States_Virgin_Islands	0	0
Uruguay	528	1
Uzbekistan	138	0
Vanuatu	0	0
Venezuela	462	5
Vietnam	2	0
Wallis_and_Futuna	0	0
Western_Sahara	0	0
Yemen	0	0
Zambia	57	1
Zimbabwe	27	0

EK 4 - Simülasyonda Kullanılan R Program Kodları

```
setwd("C:/Users/... ")
pop <- read.csv(file="...", header=..., sep = "...")
colnames(pop)=c("xt","y")

N= nrow(pop)

set<- c(3)   #3, 4 ve 5 girilerek ayrı ayrı çalıştırılacak

cycle<- c(1)

m1<-mean(pop$y)

##### Gerçek kitle kullanılıyorsa bu kısım çalıştırılmaz #####

e<-1   #1,2 ve 3 girilerek ayrı ayrı çalıştırılacak

pop$xt<-floor(pop$x/e)   #DPS yöntemi

#####

Sim<-100000   #Simülasyon sayısı

Tlist2<-list()

yoslist<-list()

sayac<-1

mu1<-numeric(Sim)   #Tahmin edicileri tanımlama

mu2<-numeric(Sim)

mu3<-numeric(Sim)

mu4<-numeric(Sim)

mu5<-numeric(Sim)

xos1iso<-numeric(Sim)

mu6<-numeric(Sim)

mu7<-numeric(Sim)

mu8<-numeric(Sim)

mu9<-numeric(Sim)
```

```

mu99<-numeric(Sim)
rss0<-numeric(Sim)  #Hata kareler için deęişken tanımlama
rss1<-numeric(Sim)
rss2<-numeric(Sim)
rss3<-numeric(Sim)
rss4<-numeric(Sim)
rss5<-numeric(Sim)
rss6<-numeric(Sim)
rss7<-numeric(Sim)
rss8<-numeric(Sim)
rss9<-numeric(Sim)
rss99<-numeric(Sim)
b0<-numeric(Sim)  #Tahmin edicilerin yanları için deęişken tanımlama
b1<-numeric(Sim)
b2<-numeric(Sim)
b3<-numeric(Sim)
b4<-numeric(Sim)
b5<-numeric(Sim)
b6<-numeric(Sim)
b7<-numeric(Sim)
b8<-numeric(Sim)
b9<-numeric(Sim)
b99<-numeric(Sim)
for (c in cycle) {  #Tekrar döngüsü
  for (s in set) {  #Küme döngüsü
    yos1<- matrix(nrow = c, ncol = s)

```

```

xos1<- matrix(nrow = c, ncol = s)

TList<-list()

xsample<-matrix(nrow=s,ncol=s)

ysample<-matrix(nrow=s,ncol=s)

for (si in 1:Sim) {      #Simülasyon döngüsü

for (j in 1:c) {

Tj<-matrix(nrow=s,ncol=s)

Cy <- pop[sample(c(1:nrow(pop)),size = s^2,replace = F),]

Cy2<-pop[sample(c(1:nrow(pop)),size = s,replace = F),]

for (i in 1:s) {

a<- sample(c(1:nrow(Cy)),size = s^2) # örneklem index değerleri

xsample<-matrix(Cy[a,]$xt,nrow=s,ncol=s)

ysample<-matrix(Cy[a,]$y,nrow=s,ncol=s)

xos<-t(apply(xsample, 1, sort))

xos1<-diag(xos)

}

yo<-t(sapply(1:nrow(xsample), function(i) ysample[i,][order(xsample[i,])]))

for (r in 1:nrow(xos)){

if (sum(xos[r,]==xos[r,r])>1){

Tj[r,] <- (xos[r,]==xos[r,r])

}

else {

Tj[r,r] <- 1

Tj[r,-r]<-0

}

}

}

```

```

Tc<-matrix(as.integer(Tj),ncol=s)

TList[[j]]<-Tc

yoslist[[j]]<-Tc * yo

yos1[j,]<- diag(yo)

}

SAYAC<-1

Te<- matrix(nrow = s*c,ncol = s)

for(ii in 1:c){
  for (jj in 1:s){
    for(ij in 1:s){
      Te[SAYAC,ij]<-TList[[ii]][jj,ij]/sum(TList[[ii]][jj,])
    }
    SAYAC<-SAYAC+1
  }
}

ni=numeric()

for(i in 1:s){
  ni[i]<-sum(Te[,i])
}

mu0<-mean(Cy2$y)
mu1<-mean(yos1)
mu2<-((mean(Cy2$y)*mean(pop$xt))/mean(Cy2$xt))
mu3<-mu1*mean(pop$xt)/mean(diag(xos))
mu4<-mean((as.vector(t(yos1))%*%Te)/ni)

Tlist2[[sayac]]<-TList

sayac<-sayac+1

```



```

library(logcondens) #PAVA için kütüphane tanımlama
mu5<-mean(isoMean(as.vector(t(yos1)),rep(c,times=s)))
mu6<-mean(isoMean((as.vector(t(yos1))%*%Te)/ni,ni))
xos1iso<-mean(isoMean((as.vector(t(xos1))%*%Te)/ni,ni))
mu7<-mu4*mean(pop$xt)/xos1iso
mu8<-mu4*exp((mean(pop$xt)-xos1iso)/(mean(pop$xt)+xos1iso))
library(moments) #Çarpıklık basıklık katsayıları için kütüphane tanımlama
mu9<-mu4*exp((((kurtosis(pop$xt)*mean(pop$xt))+1)-
((kurtosis(pop$xt)*xos1iso)+1))/((((kurtosis(pop$xt)*mean(pop$xt))+1)+((kurtosis(pop$
xt)*xos1iso)+1)))
mu99<-mu4*exp((((kurtosis(pop$xt)*mean(pop$xt))+cor(pop$xt,pop$y))-
((kurtosis(pop$xt)*xos1iso)+cor(pop$xt,pop$y)))/((((kurtosis(pop$xt)*mean(pop$xt))+c
or(pop$xt,pop$y))+((kurtosis(pop$xt)*xos1iso)+cor(pop$xt,pop$y))))
rss0[si]<-((mu0-m1)^2) #Hata Kareler Toplamlarının oluşması
rss1[si]<-((mu1-m1)^2)
rss2[si]<-((mu2-m1)^2)
rss3[si]<-((mu3-m1)^2)
rss4[si]<-((mu4-m1)^2)
rss5[si]<-((mu5-m1)^2)
rss6[si]<-((mu6-m1)^2)
rss7[si]<-((mu7-m1)^2)
rss8[si]<-((mu8-m1)^2)
rss9[si]<-((mu9-m1)^2)
rss99[si]<-((mu99-m1)^2)
b0[si]<-((mu0-m1) #Hataların oluşması
b1[si]<-((mu1-m1)
b2[si]<-((mu2-m1)

```

```

b3[si]<-(mu3-m1)
b4[si]<-(mu4-m1)
b5[si]<-(mu5-m1)
b6[si]<-(mu6-m1)
b7[si]<-(mu7-m1)
b8[si]<-(mu8-m1)
b9[si]<-(mu9-m1)
b99[si]<-(mu99-m1)
}
}
}
re0<-mean(rss1)/mean(rss0) #HKO hesaplanması
re2<-mean(rss1)/mean(rss2)
re3<-mean(rss1)/mean(rss3)
re4<-mean(rss1)/mean(rss4)
re5<-mean(rss1)/mean(rss5)
re6<-mean(rss1)/mean(rss6)
re7<-mean(rss1)/mean(rss7)
re8<-mean(rss1)/mean(rss8)
re9<-mean(rss1)/mean(rss9)
re99<-mean(rss1)/mean(rss99)
mean(b0) #Yanların hesaplanması
mean(b1)
mean(b2)
mean(b3)
mean(b4)

```

mean(b5)

mean(b6)

mean(b7)

mean(b8)

mean(b9)

mean(b99)