

KONUM PARAMETRESİNİN BAZI SAĞLAM TAHMİN EDİCİLERİNİN ÖRNEKLEME ALANINDA KULLANILMASI VE BİR TARIM UYGULAMASI

Arzu ALTIN¹, Birdal ŞENOĞLU²

¹Eskişehir Osmangazi Üniversitesi, Fen Edebiyat Fakültesi, İstatistik
Bölümü, ESKİŞEHİR

²Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, ANKARA

ÖZET

Örnekleme çalışmalarında evren parametrelerini tahmin etmek için yaygın olarak En küçük kareler (EKK) tahmin edicileri kullanılmaktadır. Ancak EKK tahmin edicileri sadece normal dağılım varsayımı altında etkin tahmin edicilerdir. Bu çalışmada istatistikte yaygın olarak kullanılan L, M ve R sağlam tahmin edici sınıflarından 8 farklı tahmin edicinin örnekleme alanında kullanımı ele alınmıştır. Ele alınan bu tahmin ediciler Türkiye'deki patates üretim verimliliğini tahmin etmek için kullanılmıştır. Sonuç olarak evrenin uzun kuyruklu simetrik dağılıma sahip olması durumunda, konum parametresinin en etkin tahmin edicisinin MML tahmin edicisi olduğu belirlenmiştir.

Anahtar Kelimeler: En Küçük Kareler (EKK) Tahmin Edicisi, Örnekleme, Sağlam Tahmin Ediciler, Uzun Kuyruklu Simetrik Dağılım.

USE OF SOME ROBUST ESTIMATORS OF LOCATION PARAMETER IN SAMPLING AND AN AGRICULTURAL APPLICATION

ABSTRACT

Least square estimators (LSE) are commonly used to estimate the population parameters in sampling theory. However, LSEs are efficient only under the Normal distribution assumption. In this study, 8 different estimators which are from commonly used classes (L, M and R) of robust estimators in statistics were used in sampling. The estimators have been used to estimate the potato production yield in Turkey. As a result, under the Long Tailed Symmetric distributed population assumption, MML estimator has been

determined as the most efficient estimator of location parameter among all estimators.

Key Words: Least Squares (LS) Estimator, Sampling, Robust Estimators, Long Tailed Symmetric Distribution.

1. GİRİŞ

İstatistiğin bir çok alanında olduğu gibi, örnekleme alanında da En Küçük Kareler (EKK) tahmin edicisi, en çok kullanılan tahmin edicidir. Ancak EKK tahmin edicileri, sadece normal dağılım varsayımı altında etkin tahmin edicilerdir [1]. Bununla birlikte, uygulamada normal olmayan dağılımlar daha yaygındır [2-4].

Normal olmayan dağılımlar için EKK tahmin edicisinin etkinliği son derece düşüktür. Ayrıca istatistiksel veri analizinde ilgilenilen veri kümesi her zaman istenilen yapıdaki değerlerden oluşmaz. Veri setindeki bir veya birkaç değer, verinin geri kalan kısmından büyük farklılıklar gösterebilir. Aykırı değer olarak adlandırılan bu veriler dağılımın şeklini değiştirmekte ve bu dağılıma bağlı olarak yapılacak tahminlerin etkinliğini azaltmaktadır. İstatistiksel modelin dayandığı varsayımların sağlanamaması durumunda veya veri setinde aykırı değer bulunması durumunda EKK tahmin edicisinin etkinliği düşmektedir.

Bu gibi durumlarda, aykırı değerlerden ve modelden sapmalardan etkilenmeyen sağlam (robust) tahmin ediciler tercih edilir. Sağlam tahmin ediciler, verideki aykırı değerlerden ve varsayılan modelden sapmalardan etkilenmeyen tahmin edicilerdir [5].

Bu çalışmadaki amacımız, örnekleme teorisinde yaygın olarak kullanılan EKK tahmin edicisinin dayandığı model varsayımlarının sağlanamaması durumunda kullanılacak alternatif bazı sağlam kestiricileri tanıtmak ve bir uygulama üzerinde karşılaştırmaktır. Çalışmada istatistikte yaygın olarak kullanılan L, M ve R sağlam tahmin edici sınıflarından 8 farklı tahmin edici ele alınmıştır. Uygulama bölümünde tabakalı örnekleme yönteminden yararlanılarak, ele alınan bu tahmin ediciler yardımıyla patates bitkisinin Türkiye'deki ortalama verim miktarı tahmin edilmeye çalışılmıştır.

2. EVREN ORTALAMASININ TAHMİN EDİCİLERİ

Bu bölümde evren ortalamasının tahmin edilmesinde yaygın olarak kullanılan EKK tahmin edicisi ile onun ağırlam alternatifleri kısaca tanıtılmıştır.

Sağlam tahmin ediciler, hesaplanış biçimlerine göre L-tahmin ediciler, R-tahmin ediciler ve M-tahmin ediciler olarak üç gruba ayrılır.

L-Tahmin edicileri ilk olarak Daniel [6] tarafından önerilmişlerdir. Sıra istatistiklerinin doğrusal kombinasyonları olarak tanımlanmaktadır [7].

M-Tahmin edicileri belirli bir amaç fonksiyonunun minimize edilmesini sağlayan, minimax prensibine dayanan sağlam tahmin edicilerdir [8-11].

R-Tahmin edicileri, rank (sıra) testlerine dayalı olarak belirlenirler. Tek örneklem durumunda en çok kullanılan R tahmin edicisi, Hodges-Lehmann tahmin edicisidir [12].

2.1. En Küçük Kareler Tahmin Edicisi

y_1, y_2, \dots, y_n , n birimlik bir örneklem olsun. Bu durumda sonlu evren ortalamasının EKK tahmin edicisi aşağıdaki gibi tanımlanır.

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (1)$$

2.2. Medyan

Medyan, bir veri setinde sadece ortadaki bir veya iki gözlemin farklı ağırlıklandırıldığı diğer tüm gözlem değerlerine sıfır ağırlığının verildiği bir L-tahmin edicisidir. $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$ sıra istatistikleri olmak üzere medyan aşağıdaki gibi tanımlanır [13].

$$Medyan = \begin{cases} y_{\left(\frac{n+1}{2}\right)} & , \quad n \text{ tek sayı ise} \\ \frac{y_{(n/2)} + y_{(n/2+1)}}{2} & , \quad n \text{ çift sayı ise} \end{cases} \quad (2)$$

2.3. Trimmed Ortalama (Trimmed Mean)

$y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$ sıra istatistikleri olmak üzere $(\alpha, 1 - \beta)$ trimmed ortalama aşağıdaki eşitlikle tanımlanır [7].

$$T_n = T_n(l_n, u_n) = \frac{1}{u_n - l_n} \sum_{i=l_n+1}^{u_n} y_{(i)} \quad (3)$$

Burada, $[\square]$ en büyük tam sayı fonksiyonu olmak üzere, verinin alt ucundan budanma oranı α , üst ucundan budanma oranı β , sıralı verilerin alt ucundan atılacak terim sayısı $l_n = [n\alpha]$, üst ucundan atılacak terim sayısı $u_n = [n\beta]$ şeklinde gösterilmektedir [13]. Genellikle kuyruklarının ağırlıklarına bağlı olarak $0 \leq \alpha, 1 - \beta \leq 0,25$ olarak alınır.

2.4. Winsorize Ortalama

$y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$ sıra istatistikleri olmak üzere $(\alpha, 1 - \beta)$ Winsorize ortalama,

$$W_n = W_n(l_n, u_n) = \frac{1}{n} \left\{ l_n y_{(l_n+1)} + \sum_{i=l_n+1}^{u_n} y_{(i)} + (n - u_n) y_{(u_n)} \right\} \quad (4)$$

şeklinde tanımlanır. Burada $[\square]$ en büyük tam sayı fonksiyonu olmak üzere, $l_n = [n\alpha]$ ve $u_n = [n\beta]$ 'dir. Trimmed ortalamada olduğu gibi Winsorize ortalamada da kuyrukların ağırlıklarına bağlı olarak genellikle $0 \leq \alpha$ ve $1 - \beta \leq 0,25$ olacak biçimde seçilir [7,13].

2.5. Trimmed L (Trimmed L-TL) Ortalama

$y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$ sıra istatistikleri olmak üzere konum parametresinin Trimmed L ortalaması,

$$\hat{\mu}_{TL} = \sum_{i=1}^n w_i y_{(i)} \quad (5)$$

olarak tanımlanır [14-15]. w_i ağırlıkları (6) numaralı eşitlikteki gibi verilir.

$$w_i = \begin{cases} \frac{\binom{i-1}{l_n} \binom{n-i}{l_n}}{\binom{n}{2l_n+1}} & ; \quad l_n + 1 \leq i \leq n - l_n \\ 0 & ; \quad \text{d.d.} \end{cases} \quad (6)$$

Burada $l_n = [n\alpha]$ 'dir.

2.6. Uyarlanmış En Çok Olabilirlik (Modified Maximum Likelihood, MML) Tahmin Edicisi

Bu çalışmada dağılımın şekline göre değişen tek tahmin edici MML tahmin edicisidir. Uygulama bölümündeki verilerin dağılımının olasılık yoğunluk fonksiyonu aşağıdaki gibi verilen uzun kuyruklu simetrik (Long Tailed Symmetric, LTS) dağılım olduğu belirlenmiştir.

$$f(y, p) = \frac{1}{\sigma} \left\{ 1 + \frac{(y - \mu)^2}{k\sigma^2} \right\}^{-p}, \quad -\infty < y < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0. \quad (7)$$

Burada $E(y) = \mu$, $V(y) = \sigma^2$, $k=2p-3$ ve $p \geq 2$ dir. Evrenin dağılımının LTS olması durumunda konum parametresinin MML tahmin edicisi aşağıdaki gibidir [1].

$$\hat{\mu} = \sum_{i=1}^n \frac{\beta_i y_{(i)}}{m} \quad (8)$$

Burada $k = 2p - 3$ olmak üzere,

$$\alpha_i = \frac{(2/k)t_{(i)}^3}{[1 + (1/k)t_{(i)}^2]^2}, \quad \beta_i = \frac{1 - (1/k)t_{(i)}^2}{[1 + (1/k)t_{(i)}^2]^2}, \quad m = \sum_{i=1}^n \beta_i, \quad t_{(i)} = E(z_{(i)}) \text{ 'dir.}$$

Sıra istatistiklerinin beklenen değerleri olarak tanımlanan $t_{(i)}$ değerleri $p=2(0,5)10$, $n \leq 20$ için tablolaştırılmıştır [16].

2.7. Huber Tahmin Edicileri

y_1, y_2, \dots, y_n , LTS dağılıma sahip olsun. Bu bölümde W24 ve BS82 olarak bilinen ve literatürde yaygın olarak kullanılan, iteratif yöntemlere gerek duymadan elde edilebilen tahmin ediciler tanıtılacaktır. Bu tahmin ediciler bir adımda çözülebildiklerinden bir-adım (one-step) tahmin ediciler olarak adlandırılırlar.

2.7.1. W24 Kestircisi

Konum parametresinin W24 tahmin edicisi,

$$\hat{\mu}_{W24} = T_0 + (hS_0) \tan^{-1} \left[\frac{\sum \sin(z_i)}{\sum \cos(z_i)} \right] \quad (9)$$

olarak tanımlanır [17]. Burada $T_0 = \text{medyan}(y_i)$; $S_0 = \text{medyan}|y_i - T_0|$; $h=2,4$; $z_i = \frac{y_i - T_0}{hS_0}$ ($1 \leq i \leq n$)'dir [17]. Konum parametresinin W24 tahmini hesaplanırken sadece $|z_i| \leq \pi$, ($1 \leq i \leq n$) koşulunu sağlayan z_i değerleri kullanılmaktadır [5, 9].

2.7.2. BS82 Kestircisi

Konum parametresinin BS82 tahmin edicisi,

$$\hat{\mu}_{BS82} = T_0 + hS_0 \frac{\sum \psi(z_i)}{\sum \psi'(z_i)} \quad (10)$$

olarak tanımlanmaktadır. Burada,

$$T_0 = \text{medyan}(y_i); S_0 = \text{medyan}|y_i - T_0|; h=8,2; z_i = \frac{y_i - T_0}{hS_0} \quad (1 \leq i \leq n)$$

$$\psi(z) = \begin{cases} z(1-z^2)^2 & ; |z| \leq 1 \\ 0 & ; |z| > 1 \end{cases}; \quad \psi' = 1 - 6z_i^2 + 5z_i^4$$

şeklindedir [9, 17].

2.8. Hodges-Lehman tahmin edicisi

Bu tahmin edici Wilcoxon sıra işaret test istatistiğinden elde edilmiştir [18]. Hodges-Lehman tahmin edicisi tüm ikili gözlem çiftlerinin ortalamalarının medyanı olarak tanımlanır.

y_1, y_2, \dots, y_n , n birimlik rassal bir örneklem olsun. Evren ortalamasının Hodges-Lehman tahmin edicisi,

$$\hat{\mu}_{HL} = \text{medyan} \left\{ \frac{y_{(i)} + y_{(j)}}{2}, \quad 1 \leq i < j \leq n \right\} \quad (11)$$

şeklinde tanımlanır.

3. UYGULAMA

Bu bölümde tabakalı örnekleme yönteminden yararlanılarak, patates bitkisinin Türkiye'deki ortalama verim miktarı ikinci bölümde ele alınan tahmin ediciler yardımıyla tahmin edilmeye çalışılmıştır. Bu amaçla Türkiye'deki 546 ilçeye ait patates üretimine ilişkin verim miktarı verileri TÜİK tarafından yapılan ve son olarak yayınlanan 2005 yılı genel tarım sayımı sonuçlarından elde edilmiştir. TÜİK tarafından tamsayım olarak yapılan çalışmadan elde edilen veriler, evrenin elemanları olarak benimsenmiştir. 2005 yılında Türkiye'de 546 ilçede patates üretimi yapılmıştır. Coğrafi bölgelere göre patates bitkisinin verim miktarının değişeceği düşünülmüş ve coğrafi bölgeler tabaka olarak benimsenmiştir. Öncelikle, TÜİK tarafından tamsayım olarak yapılan çalışmadan elde edilen verim miktarı değerleri coğrafi bölgelere göre ayrılmıştır. Güneydoğu Anadolu Bölgesinde patates üretimi bulunmadığı için bu bölge çalışmada yer almamıştır. Daha sonra, bu tabaklardan orantılı dağıtımla örneklem seçilmiş ve her coğrafi bölge için patates bitkisinin verim miktarı tahmin edilmiştir. Son olarak, her tabaka için elde edilen tahmin değerleri yardımıyla Türkiye'nin tamamında patates bitkisinin verim miktarı tahmin edilmiştir.

Ele alınan örneklemin tabakalar arasında paylaştırılmasının en basit ve en sık kullanılan yolu, örneklemin tabaka çaplarıyla orantılı olarak paylaştırılmasıdır. Bu nedenle tabaka örneklem büyüklükleri, orantılı dağıtım yöntemiyle bulunmuştur. Tabakalı örneklemede orantılı dağıtım uygulandığında incelenecek olan örneklem hacmi aşağıdaki gibi hesaplanır. d_0 kabul edilebilir hoşgörü miktarını, z_0 hata oranına göre standart normal dağılım tablosundaki değerini (güvenilirlik düzeyini) göstermek üzere, arzu edilen varyans $D^2 = d_0^2 / z_0^2$ şeklinde belirlenir. Tabakalı örneklemede orantılı dağıtım uygulandığında incelenecek olan örneklem hacmi aşağıdaki eşitlik yardımıyla hesaplanır [19].

$$n = \frac{N \sum_{h=1}^L N_h S_h^2}{N^2 D^2 + \sum_{h=1}^L N_h S_h^2}$$

Burada; N , evrendeki birim sayısını, N_h , h 'inci tabakadaki birim sayısını ve S_h^2 , h 'inci tabakanın varyansını göstermektedir.

Tabakalı örnekleme yöntemiyle patates üretimine ilişkin verim miktarının tahmini için örneklem büyüklüğünün belirlenmesinde 2004 yılına ait standart sapma değerleri kullanılmıştır. Tablo 1'de tabakalarda bulunan ilçe sayıları ve her bir tabakadaki patates üretimine ilişkin verim miktarının standart sapma değerleri verilmiştir. Hoşgörü miktarı $d_0 = 300$ kg/ha alındığında, örneklem büyüklüğü $n=63$ bulunmuştur. Orantılı dağıtım altında herbir tabaka için örneklem büyüklükleri belirlenmiş ve Tablo 2'de verilmiştir.

Tablo 1. Patates Üretimine İlişkin Verim Miktarına ait Evren Bilgileri

Bölge Adı (Tabaka)	Üretim yapılan ilçe sayısı	Standart Sapma
Akdeniz	48	1299,220
Doğu Anadolu	90	731,145
Ege	71	808,721
İç Anadolu	112	961,747
Karadeniz	165	679,331
Marmara	60	740,945
TOPLAM	546	

Tablo 2. Tabakalara Göre Örneklem Büyüklükleri

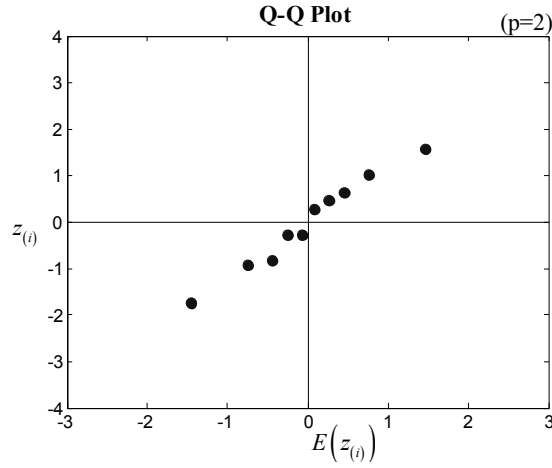
Bölge Adı (Tabaka)	Üretim yapılan ilçe sayısı	Tabaka Oranı	Örneklem Büyüklüğü
Akdeniz	48	0,09	6
Doğu Anadolu	90	0,16	10
Ege	71	0,13	8
İç Anadolu	112	0,21	13
Karadeniz	165	0,30	19
Marmara	60	0,11	7
TOPLAM	546	1,00	63

Tabakalı örnekleme yönteminde her bir tabaka ayrı birer evren olarak düşünülmektedir. Çalışmanın bu kısmında da her bir bölge birer evren olarak düşünülerek, bu evrenlerden basit rassal örnekleme yoluyla örneklem seçilmiştir. Bu bölgelere ait 2005 yılı verilerinin tamamı bilinmesine rağmen, elde edilen tahminlerin yanlılığının belirlenebilmesi amacıyla sadece örneklem verilerinin bilindiği varsayılmıştır. Bu kısımda sadece Doğu Anadolu bölgesi için yapılan analizler ayrıntılı olarak verilmiş, benzer şekilde diğer bölgeler için elde edilen sonuçlar tablolastırılmıştır.

Tablo 2’den de görülebileceği gibi, Doğu Anadolu Bölgesinde toplam 90 ilçede patates üretimi yapılmaktadır. Türkiye’deki 546 ilçe içerisinde, Doğu Anadolu bölgesinin oranı 0,16’dır. Türkiye’nin tamamında patates üretimine ilişkin verim miktarının tahmini için incelenmesi gereken ilçe sayısı 63 olduğundan, Doğu Anadolu Bölgesinde incelenmesi gereken ilçe sayısı 10 olarak belirlenmiştir. Doğu Anadolu Bölgesinden seçilecek olan bu 10 ilçe, basit rassal örnekleme yoluyla seçilmiştir. Örneğe çıkacak olan ilçeler, rassal sayılardan yararlanılarak bilgisayar yardımıyla belirlenmiştir.

Doğu Anadolu Bölgesindeki patates üretimine ilişkin verim miktarının olasılık dağılımının belirlenebilmesi amacıyla Q-Q grafiği yönteminden yararlanılmıştır. Q-Q grafiği eldeki verilerin varsayılan dağılıma uyup uymadığını gösteren grafiksel bir yöntemdir. Bilindiği gibi Q-Q grafiği yönteminde standartlaştırılmış gözlem değerleri, ilgili olasılık dağılımında karşı gelen sıra istatistiklerinin beklenen değerleriyle karşılaştırılmaktadır. Q-Q grafiğinde, gözlem değerleri bir ekseninde, teorik olarak hesaplanan değerler diğer ekseninde yer almak üzere koordinat düzlemi üzerinde noktalar işaretlenir. Bu noktalar işaretlendiğinde, noktaların büyük bir çoğunluğu bir doğru etrafında yer alıyorsa, o zaman gözlem verilerinin araştırmaya konu olan teorik dağılımdan olduğu söylenir.

Bu amaçla ilk olarak, seçilen 10 ilçeden elde edilen patates üretimine ilişkin verim miktarları küçükten büyüğe doğru sıralanmış ve standartlaştırılmıştır. Farklı şekil parametreleri için Q-Q grafikleri çizilmiş, Doğu Anadolu Bölgesindeki patates üretimine ilişkin verim miktarının $p=2$ şekil parametresine sahip LTS dağılıma sahip olduğu belirlenmiştir. $p=2$ şekil parametresi için çizilen Q-Q grafiği Şekil 1’de verilmiştir.



Şekil 1. Doğu Anadolu Bölgesindeki Patates Üretimine İlişkin Verim Miktarının $LTS(p=2)$ için Q-Q Grafiği

Doğu Anadolu bölgesindeki patates üretimine ilişkin verim miktarının olasılık dağılımı belirlendikten sonra, ele alınan tahmin ediciler yardımıyla patates üretimine ilişkin ortalama verim miktarı tahmin edilmiştir. Doğu Anadolu bölgesi için $n=10$ birimlik örneklem verileri kullanılarak elde edilen tahminler Tablo 3’te verilmiştir. Ayrıca 2005 yılında Doğu Anadolu bölgesinin 90 ilçesindeki patates üretimine ilişkin verim miktarı değerlerinden bu bölgeye ait evren ortalamasının kesin değeri 17402,4 kg/ha olarak belirlenmiştir. Bu değer yardımıyla herbir tahmin edici için yanlılık miktarları hesaplanmış ve Tablo 3’te verilmiştir.

Tablo 3. Doğu Anadolu Bölgesindeki Patates Üretimine İlişkin Verim Miktarı Tahminleri

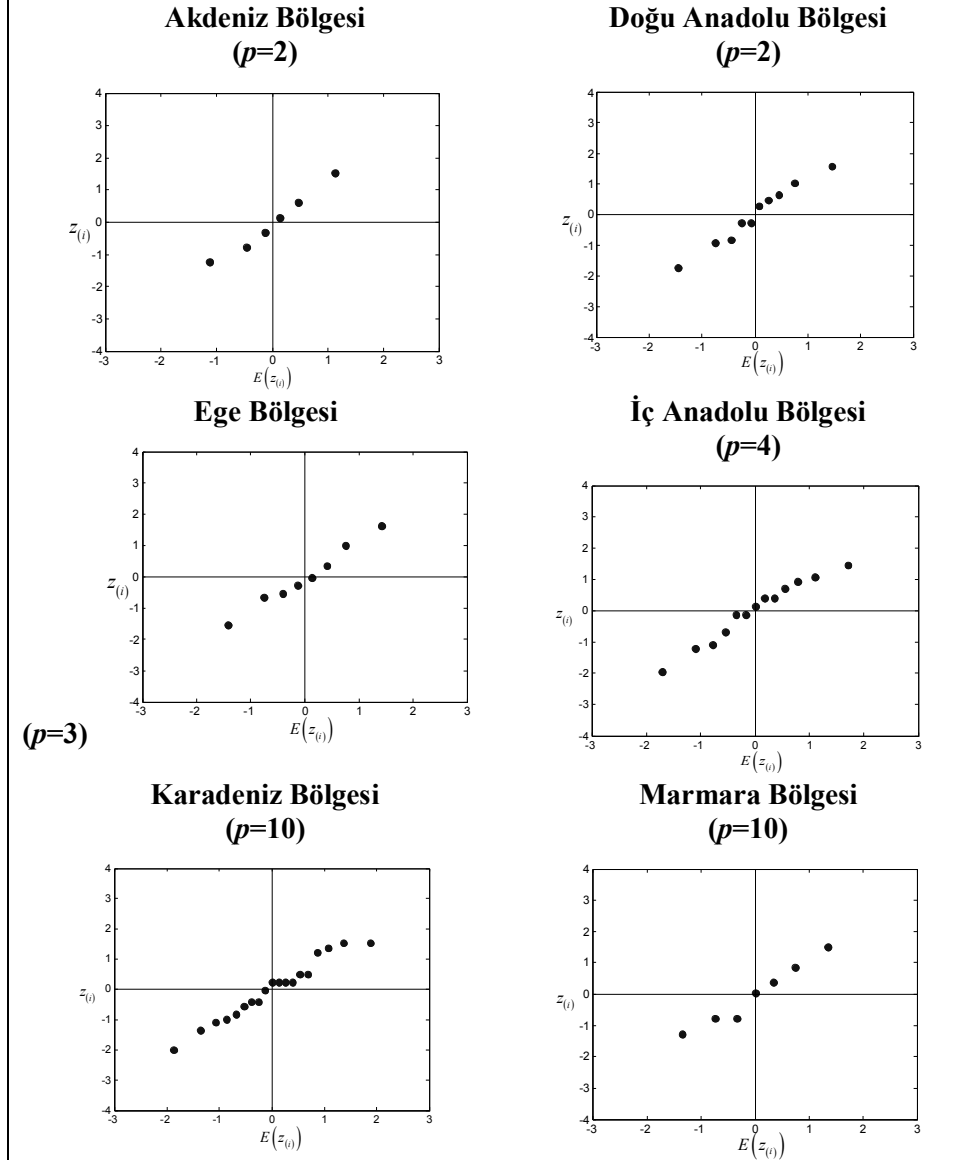
Tahmin ediciler	Tahmin Değeri	Yanlılık Miktarı
EKK	16460,7	941,7
Medyan	16500,0	902,4
Trim(0.1;0.1)	16562,5	839,9
Win(0.1;0.1)	16600,0	802,4
TL(0.1)	16566,7	835,8
MML	16634,1	768,3
W24	16495,2	907,2
BS82	16496,7	905,7
HL	16500,0	902,4

Tablo3'te yer alan evren ortalaması tahmin değerlerinden en küçük yanlılığa sahip olan tahmin edicinin MML tahmin edicisi olduğu görülmektedir. MML tahmin edicisinden sonra, Win(0.1;0.1), TL(0.1), Trim(0.1;0.1), medyan, Hodges Lehmann, BS82 ve W24 tahmin edicileri de EKK tahmin edicisinden daha az yanlılığa sahip olan diğer tahmin edicilerdir.

Bir tek örneklemeden elde edilen sonuca bakılarak, elde edilen bu değerlerden hangisinin daha güvenilir olduğuna karar verilemez. Altın [20], tabakalı örnekleme yönteminde $p=2$ şekil parametresine sahip LTS dağılım için ele alınan tüm tahmin edicileri 10,000 iterasyonluk Monte-Carlo simülasyonu ile karşılaştırmıştır. Bu karşılaştırma sonucunda evren ortalamasının en etkin tahmin edicisinin uygulamadaki sonucu destekler biçimde, MML tahmin edicisi olduğunu belirlemiştir. LTS dağılımın diğer şekil parametreleri içinde aynı sonuca ulaşılmıştır.

Bu sonuçlardan yararlanılarak, Doğu Anadolu bölgesindeki patates üretimine ilişkin verim miktarı ortalamasının en iyi tahminini MML tahmin edicisinin verdiği belirlenmiştir.

Türkiye'deki patates üretimine ilişkin verim miktarını tahmin etmek için, Doğu Anadolu bölgesi dışında kalan diğer bölgeler içinde benzer hesaplamaların yapılması gerekir. Tüm coğrafi bölgelerden incelenmesi gereken örneklem hacimleri Tablo 2'te yer almaktadır. Bu bölgelerden incelenmesi gereken birimler, basit rassal örnekleme yöntemiyle seçilmiştir. Daha sonra her bölge için patates üretimine ilişkin verim miktarının olasılık dağılımının belirlenebilmesi amacıyla Q-Q grafikleri çizilmiştir ve en uygun olan Q-Q grafikleri Şekil 2'de verilmiştir. Akdeniz bölgesindeki patates üretimine ilişkin verim miktarı $p=2$, Ege bölgesindeki $p=3$, İç Anadolu bölgesindeki $p=4$, Marmara ve Karadeniz bölgelerindeki ise $p=10$ şekil parametrelili LTS dağılım olarak bulunmuştur.



Şekil 2. Bölgelere Göre Patates Üretimine İlişkin Verim Miktarlarının LTS(p) için Q-Q Grafikleri

Tüm bölgeler için patates üretimine ilişkin verim miktarının sahip olduğu olasılık dağılımı belirlendikten sonra, ele alınan tahmin ediciler yardımıyla her bölge için evren ortalamaları tahmin edilmiştir. Her bir bölge için ayrı ayrı elde edilen tahmin değerleri, 2005 yılı verileri kullanılarak belirlenen

evren ortalamalarının kesin değerleri ve bu değerler kullanılarak belirlenen yanlışlık miktarları Tablo 5’de verilmiştir. Bölge adlarının altında yer alan değerler evren ortalamalarının kesin değerleridir. Bu değerlerle çeşitli tahmin ediciler yardımıyla her bölge için belirlenen ortalama verim miktarı tahminleri karşılaştırıldığında, mutlak değerce en az yanlışlığa sahip olan tahmin edicinin MML tahmin edicisi olduğu görülmektedir. Her bölgeden farklı büyüklükte örneklem incelendiğinden diğer tahmin ediciler arasındaki sıralama değişmektedir. Özellikle örneklem hacmi $n < 10$ olduğunda EKK tahmin edicisi ile Trim(0.1;0.1), TL(0.1) ve Win(0.1;0.1) ortalamalar benzer sonuçlar vermektedir. Şekil parametresi p ’nin değeri 10 olduğunda EKK tahmin edicisi iyi sonuçlar vermektedir. Bunun nedeni $p > 10$ için LTS dağılımının Normal dağılıma yaklaşmasıdır.

Şu ana kadar her bir bölge için elde edilen sonuçlar üzerinde durulmuştur. Tabakalı örnekleme yöntemi uygulandığında, Türkiye’deki patates üretimine ilişkin verim miktarı tahmini Tablo 2’de verilen tabaka oranları ve Tablo 4’te verilen sonuçlar yardımıyla belirlenebilir. Türkiye’deki patates üretimine ilişkin verim miktarı tahmini için tabakalı örnekleme yöntemi uygulandığında elde edilen sonuçlar Tablo 5’te verilmiştir. Tablo 5’ten görüleceği üzere, tabakalı örnekleme yöntemi uygulandığında patates üretimine ilişkin verim miktarını en az hata ile yine MML tahmin edicisi tahmin etmektedir.

buraya yatık soyfa montajlanacak

Tablo 5. Türkiyede Patates Üretimine İlişkin Ortalama Verim Miktarı Tahminleri

Tahmin ediciler	Tahmin Değeri	Yanlılık Miktarı
EKK	21602,0	893,7
Medyan	21111,5	403,1
Trim(0.1;0.1)	21527,0	818,7
Win(0.1;0.1)	21605,6	897,2
TL(0.1)	21366,1	657,8
MML	20977,1	268,8
W24	21474,2	765,9
BS82	21470,0	761,7
HL	21372,5	664,2

4. SONUÇ

Bu çalışmada örnekleme alanında yaygın olarak kullanılan EKK tahmin edicisine alternatif bazı sağlam tahmin ediciler tanıtılmıştır. L, M ve R sağlam tahmin edici sınıflarından 8 farklı tahmin edici çalışmaya alınmıştır. Evrenin normal dağılıma sahip olmaması durumuna değinilmiş, evrenin dağılımının uzun kuyruklu simetrik olması durumunda ele alınan tahmin ediciler arasından evren ortalamasının en etkin tahmin edicisinin MML tahmin edicisi olduğu belirlenmiştir.

Uygulama kısmında, tabakalı örnekleme yönteminden yararlanılarak patates bitkisinin Türkiye genelinde ortalama üretim verimliliği, en az yanlılıkla MML tahmin edicisi tarafından elde edilmiştir. Ayrıca her tabakadan basit rassal örnekleme yöntemi ile örneklem seçildiğinden, her coğrafi bölge için elde edilen ortalama üretim verimliliği tahmin değerleri incelendiğinde, en küçük yanlılığa sahip olan tahmin edicinin yine MML tahmin edicisi olduğu görülmektedir.

Elde edilen tüm bu sonuçların ışığında, evrenin LTS dağılıma sahip olması durumunda konum parametresinin tahmini için en etkin tahmin edicinin MML tahmin edicisi olduğu söylenebilir.

KAYNAKLAR

1. Tiku, M.L., Akkaya, A.D., Robust estimation and hypothesis testing, New Age Int.Pub., New Delhi (2004).
2. Pearson, E.S., Adyanthaya, N.K., The distribution of frequency constants in small samples from non-normal symmetrical and skew populations, Biometrika, 21, 259-286, (1929).

3. Pearson, E.S., The analysis of variance in cases of nonnormal variation, *Biometrika*, 23, 114-133, (1932).
4. Scheffe, H., The Analysis of variances, John Wiley&Sons, New York (1959).
5. Tiku, M.L., Tan, W.Y. and Balakrishnan, N., Robust inference, Marcel Dekker Inc., New York (1986).
6. Daniel, C., Observations weighted according to order, *Amer. J. Math.*, 42, 222-236, (1920).
7. Hoaglin, D. C., Mosteller, F., Understanding robust and exploratory data analysis, John Wiley & Sons, New York (1983).
8. Huber, P.J., Robust Estimation of a location parameter, *Ann. Math. Statist.*, 35; 73-101, (1964).
9. Huber, P. J., Robust Statistics, Wiley, New York (1981).
10. Hlavka, Z., Robust sequential methods, PhD Thesis, Charles University, Faculty Mathematics and Physics, Department of Probability and Mathematical Statistics, Berlin, 131p., www.quantlet.de/mdstat/scripts/rsm/rsmps.ps, (2000).
11. Shevlyakov, G.L., Vilchevski, N.O., Robustness in data analysis: Criteria and methods, Brill Academic Publishers, Leiden, Netherlands (2002).
12. Staudte, R.G., Sheather, S.J., Robust estimation and testing, John Wiley & Sons, New York (1990).
13. Olive, D. J., Applied Robust Statistics, e-book, www.math.siu.edu/olive, (2005).
14. Elamir, E.A.H., Seheult, A.H., Trimmed L-moments, *Comput. Stat. Data. An.*, 43, 299-314, (2003).
15. Leonowicz, Z., Karvanen, J., Shishkin, S.L., Trimmed estimators for robust averaging of event-related potentials, *Journal of Neuroscience Methods*, 142, 17-26, (2005).
16. Tiku, M.L., Kumra, S., Expected values and variances and covariances of order statistics for a family of symmetric distributions (Student's t). In selected tables in mathematical statistics, Vol. 8. American Mathematical Society: Providence, RI; 1985, 141-270, (1981).
17. Gross, A.M., Confidence interval robustness with long-tailed symmetric distributions, *J. Amer. Stat. Assoc.*, 71, 409-416, (1976).
18. Hodges, J.L.Jr., Lehmann, E.L., Estimates of location based on ranks tests, *Annals of Mathematical Statistics*, 34 (2), 598-611, (1963).
19. Yamane, T., Temel örnekleme yöntemleri, Çev. Esin, A., Bakır, M.A., Aydın, C., Gürbüzel, E., Literatür, Ankara, (2001)
20. Altın, A., Sağlam Kestiricilerin Etkinliklerinin Farklı Örnekleme Yöntemleri için Karşılaştırılması ve Uygulaması, Doktora Tezi, İstatistik Bölümü, Eskişehir Osmangazi Üniversitesi, 186, Eskişehir, (2007).