

## Aykırı değer varlığında hızlı minimum kovaryans determinanı kestiricilerinin faktör analizinde kullanımı

Özlem Alpu\*

20.05.2016 Geliş/Received, 15.10.2016 Kabul/Accepted

doi: 10.16984/saufenbilder.56188

### ÖZ

Çok değişkenli analizlerden biri olan faktör analizi veri indirgeme, değişkenler arasındaki ilişkileri belirleme ve aynı zamanda sınıflandırma yöntemi olarak karşımıza çıkmaktadır. Çok sayıda değişken içeren veri setlerini analiz ederken araştırmacılar problemin sonuçları üzerinde olası zararlı etkileri olabilen ve aykırı değer olarak isimlendirilen olağandışı gözlemlerle karşılaşabilmektedir. Sıklıkla hatalı gözlemler olarak değerlendirilen aykırı değerler, veri seti hakkında önemli bilgileri içinde barındırabileceği gibi kurulan modelin yanlış belirlenmesi, yanlış parametre kestirimi ve yanlış analiz sonuçlarına da neden olabilmektedir. Bu çalışmanın amacı aykırı değerlerin parametre kestirimlerini yanlış yapmayacak özellikte faktör analizi metodunu kullanmaktır. Bu amaçla aykırı değerlerin etkisini azaltan sağlam konum ve ölçek kestiricileri içinden hızlı minimum kovaryans determinanı kestiricileri tercih edilip, bir faktör analizi uygulaması gerçekleştirilmiştir. Sonuç olarak aykırı değerlerin etkisini azaltarak verilerin çoğunluğuna uyan, açıklanan varyansı daha yüksek ve değişkenlerin daha anlamlı olarak faktörlerde toplandığı sağlam faktör analizi kestirim sonuçları elde edilmiştir.

**Anahtar Kelimeler:** aykırı değer, faktör analizi, hızlı minimum kovaryans determinanı

## Using fast minimum covariance determinant estimators for factor analysis in the presence of outliers

### ABSTRACT

Factor analysis, one of those multivariate methods, is used for data reduction, determining the relationships between variables, and also as a classification method. The researchers frequently face with the observations that might have bad affects on the results while analysing data sets with many variables. Those kind of observations are called outliers. Outliers, usually evaluated as erroneous observations, may reflect important information about the data set, but on the other hand may cause misspecification of the model, biased parameter estimates, and may lead incorrect analysis results. The aim of this study is to use factor analysis method which will not make biased parameter estimations of the outliers. For this aim, among the robust local and scale parameter estimators which reduce the affects of outliers, fast minimum covariance determinant estimators are choosen and an application is performed. As a result, robust factor analysis estimation results are obtained which are mostly fitted to the data by reducing the affect of the outliers, with higher explained variability and with explained variables which are gathered on more significant factors.

**Keywords:** outlier, factor analysis, fast minimum covariance determinant

---

\* Eskişehir Osmangazi Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü - oalpu@ogu.edu.tr

## 1. GİRİŞ (INTRODUCTION)

Gerçek hayatta karşılaşılan problemleri ortaya çıkaran birçok değişken söz konusu olabilir. Bu problemlerin çözümü için probleme neden olan tüm değişkenlerin incelenmesi gerekebilir. Problemin çözümü ancak bu değişkenlerin tamamının aynı anda incelenmesiyle mümkün olabilir. Bu değişkenlerin sayısı ne kadar fazla ise etkileri o kadar karmaşık ve problemin çözümü o denli zor olacaktır. Bilimsel çalışmalarda genellikle araştırmacının dikkate aldığı birden fazla sayıda değişkene ait veriler elde edilmekte, bu verilerin tek değişkenli analiz teknikleri kullanılarak yapılan analizlerinde, ele alınan değişkenlerin etkilerini tam olarak açıklamak mümkün olmayabilmektedir. Bu nedenle çok değişkenli istatistiksel analiz teknikleri geliştirilmiş ve bu teknikler bütün bilim dallarında yapılan çalışmalarda kullanıma sunulmuştur. Çok değişkenli istatistiksel analiz teknikleri olayların fazla sayıdaki değişkenler arasındaki ilişkilerine bağlı olarak, değişkenlerin daha anlamlı, kolay anlaşılır ve özet biçiminde yorumlanmasını sağlamaktadır. Faktör analizi de amaçlarından biri kendi aralarında önemli ilişkilere sahip değişkenleri gruplamak olan ve literatürde en sık kullanılan çok değişkenli istatistiksel tekniklerden biridir. Önceleri genellikle insan davranışını araştıran tüm sosyal bilimlerde kullanılan faktör analizi daha sonra diğer bilim alanlarında da kullanılmaya başlanmış, günümüzde uluslararası ilişkiler, sosyoloji, eğitim, ekonomi, insan-makine sistemleri, trafik kazaları araştırmaları, biyoloji, psikoloji ve tıp, jeoloji, meteoroloji, ulaştırma, finans ve risk yönetimi, sigortacılık ve bankacılık alanlarında kullanılmaktadır. Bu analiz için bazı önemli referanslar verilebilir [1-3].

Çok sayıda değişken içeren veri setlerini analiz ederken araştırmacılar problemin sonuçları üzerinde olası zararlı etkileri olabilen ve aykırı değer olarak isimlendirilen olağandışı gözlemlerle karşılaşabilmektedir. Aykırı değerler her hangi bir uygulama alanındaki neredeyse her veri setinde ortaya çıkabilmektedir. Bazen zararsızdırlar ve analize aykırı değerlerle devam edilse de, analiz öncesi silinse de sonuçları değiştirmeyebilirler. Klasik çok değişkenli konum ve ölçek kestiricilerinden olan örneklem ortalaması ve örneklem kovaryans matrisi eğer veriler normal dağılımdan geliyorsa optimaldir, ancak tek bir aykırı değer varlığına dahi aşırı derecede duyarlıdırlar. Veri setinde aykırı değer olması durumunda bu iki kestirici etkileneceğinden klasik faktör analizinin performansı da buna bağlı olarak kötüleşecektir [4]. Tek bir aykırı değer varlığı korelasyon kestirimlerinin yanı sıra Cronbach alfa ölçüsünü, kestirimi yapılan faktör analizi model parametrelerini ve modelin uyum iyiliği ölçülerini bile bozabilir [5-7]. Sonuç olarak aykırı değerlerin yanlış model

belirlenmesine yol açabildiği ve modele ekstra bir faktörün eklenmesine bile sebep olabileceği ileri sürülmüştür [8]. Ayrıca aykırı değerlerin Heywood durumları olarak bilinen negatif hata varyanslarına sebep olabileceği de belirtilmiştir [9]. Bu nedenlerle aykırı değerlerin etkisine dirençli olan yapıyı oluşturmak amacıyla sağlam faktör analizi geliştirilmiştir.

Bu çalışma veri setinde aykırı değer olması durumunda gizli veri yapısını belirlemek için klasik faktör analizinden ziyade sağlam faktör analizinin kullanımını incelemektedir. Bu amaçla gerçekleştirilen çalışmada klasik faktör analizi modeline değinildikten sonra sağlam faktör analizi ayrıntılı olarak ele alınmış, bu analizde kullanılan minimum kovaryans determinanı (MCD) ve hızlı MCD algoritmalarına yer verilmiş, ardından örnek bir uygulama üzerinde klasik ve sağlam faktör analizi sonuçları karşılaştırılmıştır.

## 2. MATERYAL VE YÖNTEM (MATERIAL AND METHOD)

### 2.1. Faktör Analizi Modeli (Factor Analysis Model)

Çok değişkenli analizde çok sık kullanılan tekniklerden biri olan faktör analizinin temel amacı gözlenen değişkenler kümesinden elde edilecek hipotetik faktörleri veya nicelikleri üretmektir. Faktörler mümkün olduğu kadar basit olmalı ve kolayca yorumlanabilmeleri için yeterli kesinlikte gözlemleri açıklayabilmeli/ tanımlayabilmelidir. Bu amaçla analiz  $X_1, X_2, \dots, X_p$  olmak üzere  $p$  tane rassal değişken varlığında, doğrudan gözlemlenemeyen ve varsayımsal nicelikler olan daha az sayıda ( $1 \leq k < p$ ) gizli değişkenin (faktörün)  $f_1, f_2, \dots, f_k$  varlığını varsaymaktadır. Faktörler aşağıdaki denklem aracılığıyla orijinal değişkenlerle bağlantılıdır:

$$X_j = \lambda_{j1}f_1 + \lambda_{j2}f_2 + \dots + \lambda_{jk}f_k + \varepsilon_j, \quad 1 \leq j \leq p. \quad (1)$$

$\varepsilon_1, \dots, \varepsilon_p$  hata terimleri veya spesifik faktörler olarak isimlendirilir, kendi aralarında ve faktörlerle bağımsız oldukları varsayılır.  $\lambda_{jl}$  katsayıları yükleri gösterir ve  $\Lambda \in \mathbb{R}^{p \times k}$  yük matrisinde yer alırlar. Vektör gösterimini kullanarak  $\mathbf{X} = (X_1, \dots, X_p)'$ ,  $\mathbf{f} = (f_1, \dots, f_k)'$  ve  $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_p)'$  olmak üzere, klasik  $k$  faktörlü model aşağıdaki gibi yazılabilir:

$$\mathbf{X} = \Lambda \mathbf{f} + \boldsymbol{\varepsilon}. \quad (2)$$

Bu eşitlikte  $E(\mathbf{f}) = E(\boldsymbol{\varepsilon}) = 0$ ,  $Cov(\mathbf{f}) = \mathbf{I}_k$ , ve  $Cov(\boldsymbol{\varepsilon}) = diag(\boldsymbol{\Psi})$  olmak üzere,  $\boldsymbol{\Psi} = (\psi_1, \dots, \psi_p) \in \mathbb{R}^p$  köşegen matrisi ifade etmektedir. Hata terimlerinin varyansları olan  $\psi_1, \dots, \psi_p$  özgül (spesifik) varyanslar olarak isimlendirilir. Ayrıca  $\boldsymbol{\varepsilon}$  ve  $\mathbf{f}$ 'nin birbirlerinden bağımsız oldukları varsayılmaktadır.

Faktör analizindeki temel adım  $\Lambda$  (sadece dik dönüşümde belirlenir) ve  $\Psi$  matrislerinin kestirimidir. Temel faktör analizi (PFA) ve en çok olabilirlik (ML) gibi klasik faktör analizi teknikleri  $X$ 'in kovaryans matrisi  $\Sigma$ 'nın ayrıştırılmasına dayalıdır [10]:

$$\Sigma = \Lambda\Lambda' + \text{diag}(\Psi). \quad (3)$$

Kovaryans matrisinin aykırı değerlere çok hassas olması, aykırı değer içeren veri setlerinde kullanımını kısıtlamaktadır. Aykırı değerlerin etkisine dirençli olan yapıyı oluşturmak amacıyla geliştirilen sağlam faktör analizine izleyen bölümde yer verilecektir.

## 2.2. Sağlam Faktör Analizi (Robust Factor Analysis)

Faktör analizi modeli örneklem kovaryans matrisinin hesabıyla başlar. Bazı yazarlar örneklem kovaryans matrisinin sınırlı olmayan etki fonksiyonuna ve sıfır bozulma noktasına sahip olduğunu, bu nedenle de tek bir aykırı değer varlığında bile güvenilir olmayan sonuçlar üreteceğine dikkat çekmişlerdir [11]. Ayrıca aykırı değerlerden dolayı normallik varsayımının sağlanmaması nedeniyle faktör modelinde sağlam kovaryans matrisinin kullanılması gerekliliği önerilmiştir [12]. Çok değişkenli sürekli değişkenler için sağlam konum ve yayılım (dispersion) ölçülerini elde etmek için alternatif yolları inceleyen pek çok çalışma söz konusudur [13-26]. Ayrıca sağlam faktör analizi üzerine de çalışmalar mevcuttur [4], [7], [11] [27-30]. Tüm bu yaklaşımlarda aykırı değerlere düşük ağırlık verilerek, orijinal değişkenlerin kovaryans matrisinin sağlam başlangıç kestirimi elde edilince, sonraki tüm parametre kestirimlerinin de sağlam olacağı düşüncesi söz konusudur.

Klasik yöntemlerle  $\Sigma$  matrisinin kestirimini yapmak yerine sağlam kestirici kullanıp, bilinmeyen  $\Lambda$  ve  $\Psi$  parametrelerinin kestirimleri sağlam kovaryans matrisini ayrıştırarak elde edilebilir. Bu yöndeki ilk çalışmalardan biri örneklem kovaryans matrisi çok değişkenli  $M$  kestiricisi ile yer değiştirilerek yapılmıştır [31]. Bu kestiricinin en büyük dezavantajı hesaplama karmaşıklığı ve düşük bozulma noktasına sahip olmasıdır.  $p$  boyutlu veriler için  $M$  kestiricisinin bozulma değeri en fazla  $1/p+1$ 'dir, bu değer çok yüksek boyutlu veriler için oldukça düşük değerdir [27].

Faktör analizini sağlam yapmak için farklı bir yaklaşım evren kovaryans matrisinin sağlam kestiricisi olarak minimum hacimli elipsoid (MVE) kestiricisinin kullanılmasıdır [27]. MVE kestiricisi mümkün olan en büyük bozulma değeri %50'ye sahip olduğundan, bu yaklaşım sağlam faktör analizi sonuçlarını verir. Buna rağmen, değişken sayısı çok fazla olduğunda hesapsal

karmaşıklık hızlıca artar ve yaklaşım cazip olmaktan uzaklaşır.

Bu dezavantajlardan dolayı Eşitlik (3)'teki  $\Sigma$  matrisinin sağlam bir kestiricisi olarak hızlı MCD kestiricisi kullanılmıştır [4]. MCD algoritmasının çalışması için gereken zaman değişken ve gözlem sayısına bağlı olarak oldukça fazla olabilir. Pek çok araştırmacı bu problemin üstesinden gelebilmek için algoritmayı iyileştirmeye/hızlandırmaya odaklanmışlardır [32-35]. Son olarak Rousseeuw ve Van Driessen [34] tarafından hazırlanan hızlı bir algoritma ile son derece cazip hale gelen sağlam faktör analizi aşağıdaki temel özelliklere sahip olmuştur:

- MCD kestiricisi için parametre seçimine bağlı olarak yöntemin bozulma değeri en fazla %50 olmaktadır.
- Faktör sayısının belirlenmesi, yüklerin kestirimi ve faktör skorları için yaygın olan yöntemleri kullanabilmektedir.
- Yöntem büyük veri setleriyle başa çıkabilmekte ancak yöntem değişken sayısının gözlem sayısından fazla olduğu durumlarda çalışmamaktadır.
- Pison ve arkadaşları [4] tarafından etki fonksiyonu türetilmiş ve ampirik etki fonksiyonu etkili gözlemlerin belirlenmesi için kullanılabilmiştir.

Hızlı MCD kestiricisinin kullanımıyla, aynı zamanda simülasyon çalışmaları aracılığıyla iki faktör çıkarma tekniği (PFA ve ML) karşılaştırılmıştır [4]. Yüklerin ve spesifik varyansların daha kesin bir şekilde kestirimi yapıldığından PFA'nın ML tekniğine tercih edildiği sonucunu ortaya çıkarmışlardır. Ayrıca Pison ve arkadaşları [4] PFA için etki fonksiyonunu türetmişlerdir. PFA için örneklem kovaryans matrisi kullanılarak Tanaka ve Odaka'nın [36] çalışmalarını doğrulayan sınırlı olmayan etki fonksiyonunu vermişlerdir. Bununla birlikte MCD kestiricisi sınırlı etki fonksiyonuna sahiptir. Pison ve arkadaşları [4] aynı zamanda korelasyon matrisi  $\rho$ 'ya dayalı PFA'nın etki fonksiyonunu hesaplamışlardır. Korelasyon matrisi

$$\rho = \Sigma_D^{-1/2} \Sigma \Sigma_D^{-1/2} \quad (4)$$

olmak üzere  $\Sigma_D$  matrisi  $\Sigma$  matrisinin köşegen elemanlarını içerir ve matrisin geri kalan elemanları sıfırdır. Eğer MCD kestiricisi kovaryans matrisinin kestirimi için kullanılırsa, sağlam korelasyon matrisi Eşitlik (4) kullanılarak kolaylıkla elde edilebilir. Bu sağlam korelasyon matrisine dayalı PFA'nın etki fonksiyonu yine sınırlıdır (bounded) ve bu nedenle yöntem aykırı değerlere karşı sağlam olur.

### 2.2.1. Minimum kovaryans determinanı metodu (Minimum covariance determinant method)

Çok değişkenli veri seti söz konusu olduğunda,  $X_1, X_2, \dots, X_p$  olmak üzere  $p$  tane rassal değişkeni  $n$  birim için  $n \times p$  boyutlu matris formunda  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$  ile  $i$ . gözlem ise  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$  olarak gösterilsin. Araştırmacılar bu çok değişkenli veri setinde aykırı değer olması halinde klasik konum ( $\bar{x}$ ) ve ölçek kestiricisi ( $s^2$ ) yerine yüksek bozulma değerine sahip kestiriciler ile ilgilenir. Yüksek bozulma değerine sahip konum ve ölçek kestirimindeki alternatiflerden biri MCD kestiricileridir.

MCD metodunun amacı,  $n$  birimlik örneklem hacmi için hesaplanan kovaryans matrisi determinanı minimum olan alt kümeleri ve  $h$  adet gözlemi inceleyerek, kovaryans matrisi determinantını minimum yapan  $h$  adet gözlemden oluşan alt kümeyi bulmaktır [34].

$n$  birimlik örneklemde seçilecek gözlem sayısı  $h$  değerine genellikle  $h = \left\lfloor \frac{n+p+1}{2} \right\rfloor \approx \frac{n}{2}$  eşitliğiyle, burada [.] tamsayı kısmı göstermek üzere, veya  $h \approx [0.75n]$  bazen de  $h \approx [0.80n]$  kullanılarak karar verilmektedir.  $h$  değerinin aykırı değer içermeyen minimum gözlem sayısından oluştuğu kabul edilir. Seçilecek  $h$  adet gözlem için hesaplanan ortalama MCD'nin konum parametre kestirimi, aynı gözlemler için hesaplanan varyans-kovaryans matrisi de ölçek parametre kestirimi olacaktır [23], [35], [37].

MCD yönteminde öncelikle  $C_h^n$  kombinasyonu kadar alt örnek oluşturulur. Daha sonra bu alt örneklerin her biri için örneklem aritmetik ortalaması  $\bar{x}$  ve örneklem varyansı  $s^2$  hesaplanır. Hesaplanan  $s^2$ 'ler içinden determinanı en küçük olan değer belirlenerek, ölçek parametresinin kestirimi ( $s_{MCD}^2$ ) kabul edilir ve bu değeri oluşturan alt örneğin örneklem aritmetik ortalaması konum parametresinin kestirimi ( $\bar{x}_{MCD}$ ) olur. Bu  $p \times 1$  boyutlu ortalama vektörü ( $\bar{x}_{MCD}$ ) ve  $p \times p$  boyutlu kovaryans matrisine ( $S_{MCD}^2$ ) MCD kestiricileri adı verilir.  $\mathbf{x}_{(MCD)} = \{i_1, \dots, i_h\}$ , MCD alt kümesine ait gözlemlerin indislerini göstermek üzere,

$$\bar{x}_{MCD} = \frac{1}{h} \sum_{i \in \mathbf{x}_{(MCD)}} \mathbf{x}_i \quad (5)$$

MCD alt kümesinin aritmetik ortalamasıdır. Ölçek parametresinin kestirimi de eşitlik 6'daki gibi verilir.

$$S_{MCD}^2 = \frac{k_{MCD}(h,n,p)}{h-1} \times \sum_{i \in \mathbf{x}_{(MCD)}} (\mathbf{x}_i - \bar{x}_{MCD})(\mathbf{x}_i - \bar{x}_{MCD})' \quad (6)$$

Burada  $k_{MCD}(h, n, p)$   $h, n$  ve  $p$  değerlerine bağlı olan bir sabittir ve her bir  $\mathbf{x}_i \sim N(\mu, \Sigma)$  olduğunda  $S_{MCD}^2$

kestiricisinin tutarlı ve yansız olmasını sağlamaktadır [38].

$n$  ve  $p$ 'nin büyük değerleri için bu algoritma yardımıyla MCD kestirimleri oldukça fazla işlem yükü ve zaman gerektirmektedir. Bunun nedeni MCD kestirimlerini bulurken her  $h$  birimlik örneğin seçilmesi gerektiği ve seçilen bu örneklerin her birinin kovaryans matrisinin determinantının hesaplanması gerekmesidir. Bilgisayarla bile hesabı yoğun işlemlerin üstesinden gelebilmek için Hardin ve Rocke [35] tarafından MCD'nin kestiriminde çeşitli alternatif algoritmalar önerilmiştir.

Bu algoritmalarından biri olarak uygun çözüm algoritması (Feasible Solutions Algorithm) ile Hawkins [32]  $h$  birimlik örnekte yer alan gözlemlerin tek tek değiştirilmesine dayanan bir yöntem önermiştir. Bu yöntemde öncelikle  $h$  büyüklüğünde bir alt örnek olan  $H_1$  rasgele belirlenir. Ardından  $H_1$ 'in elemanı olan bir  $x'_i \in H_1$  gözlemi  $x'_i \notin H_1$  gözlemi ile yer değiştirilerek, yeni alt örnek  $H_2$  olarak isimlendirilir, eğer

$$m_i = \det(\text{Cov}(H_1)) - \det(\text{Cov}(H_2)) > 0 \quad (7)$$

olursa yukarıdaki  $m_i$ , sırasıyla  $H_1$ 'in elemanı olan ve  $H_1$ 'in elemanı olmayan tüm mümkün  $x'_i$  ve  $x'_j$  çiftlerinin yer değiştirmesi üzerinden maksimize edilir. (7) no'lu koşul sağlanmıyorsa  $x'_i$  ve  $x'_j$  çiftleri yer değiştirilmez.

Buna göre, maksimum  $m_i$ 'ye sebep olan  $H_2$ ,  $h$  büyüklüğünde yeni alt örnek olarak belirlenmiş olur. Yapılan değişiklik  $\det(\text{Cov}(H_1))$  değerini azaltmayınca kadar bu süreç tekrarlanır [34].

### 2.2.2. Hızlı minimum kovaryans determinanı metodu (Fast minimum covariance determinant method)

Daha hızlı bir yöntem olarak Rousseeuw ve Van Driessen [34] tarafından önerilen bu yöntemin temeli C-adıma (C-Step) dayalıdır. Bu yöntemde bir gözlem çiftini yer değiştirmek yerine, her bir adımda daha fazla gözlem yer değiştirilir. Yöntemde  $h$  büyüklüğünde  $C_1$  alt örnekle yonteme başlanır. Her bir  $i = 1, 2, \dots, n$  gözlemi için  $H_1$  alt örneğine dayalı,  $\bar{x}_{H_1}$ ,  $s_{H_1}^2$  ve  $d_{H_1}^2(x_i, \bar{x}_{H_1}) = d_{H_1}^2(i)$  hesaplanır. Daha sonra bu uzaklıklar,  $d_{H_1}^2(i)$ , sıralanır.

$$d_{H_1}^2(\pi(1)) \leq d_{H_1}^2(\pi(2)) \leq \dots \leq d_{H_1}^2(\pi(n)) \quad (8)$$

Daha sonra sıralanan bu uzaklıklardan minimum  $h$  gözlemin ait olduğu gözlemler  $\{\pi(1), \pi(2), \dots, \pi(h)\}$   $H_2$ 'ye atanır. Bu  $H_2$  alt kümesi üzerinden aritmetik ortalama ( $\bar{x}_{H_2}$ ), varyans ( $s_{H_2}^2$ ) ve uzaklıklar ( $d_{H_2}^2(x_i, \bar{x}_{H_2})$ ) hesaplanır. Bu süreç  $\pi$  dizilimi değişmeye kadar tekrarlanır.

İki algoritmada da görüldüğü gibi hangi  $H_1$  alt örneğinin seçileceği belli değildir. Hawkins [32] veriden  $h$  birimlik

rastgele seçilen bir alt grubu kullanmıştır. Eğer veri oldukça fazla kirlenme (contamination) gösteriyorsa, seçilen  $h$  birimlik rastgele örnek neredeyse hep kirlenmiş veri içereceğinden, iki algoritma da kirlenmemiş verinin parametre kestirim sonuçlarına yakınsamayacaktır [35]. Kirlenmiş veri için, Rousseeuw  $p+1$  (bir tekil olmayan kovaryans matrisinin sahip olması gereken en az gözlem sayısı) büyüklüğünde rastgele bir alt örnek ile başlamayı ve  $h$  tane gözlem oluşturulana kadar gözlem eklemeyi önermiştir. Algoritma  $p+1$  büyüklüğünde rasgele bir alt örnek ( $H_0$ ) ile başlar.  $\bar{x}_{H_0}$  ve  $s_{H_0}^2$  hesaplanır. Eğer  $\det(s_{H_0}^2)=0$  ise  $\det(s_{H_0}^2)>0$  olana kadar rasgele gözlemler alt örneğe eklenir.  $d_{s_{H_0}^2}^2(x_i, \bar{x}_{H_0}) = d_{H_0}^2(i)$  uzaklıkları hesaplanır ve sıraya  $\{d_{H_0}^2(\pi(1)) \leq d_{H_0}^2(\pi(2)) \leq \dots \leq d_{H_0}^2(\pi(n))\}$  dizilir, burada  $H_1 := \{\pi(1), \pi(2), \dots, \pi(h)\}$  olarak tanımlanır [34].

Sonuç olarak kullanılacak algoritmanın birinci adımında elde edilen alt örnek için örneklem aritmetik ortalaması ve örneklem varyans-kovaryans matrisi hesaplanır. İkinci adım olarak birinci adımdaki kestirimlere dayalı olarak, örneklemdeki her bir gözlem için Mahalanobis uzaklıkları ( $d_i^2$ ) hesaplanır. Üçüncü adım olarak ikinci adımdan en küçük  $d_i^2$ 'ye sahip noktalardan bir alt örneklem seçilir. 1-3 adımları alt örneklem değişmeyinceye kadar tekrarlanır.  $\bar{x}_{MCD}$  ve  $s_{MCD}^2$ , MCD'nin konum ve ölçek parametrelerinin kestiricileri olmak üzere, aykırı gözlemlerin bu kestiricileri etkilemesi ihtimali düşük olduğu için, sağlam bir kestirici olan  $d_{s_{MCD}^2}^2(x_i, \bar{x}_{MCD})$ 'nin aykırı gözlemleri belirlemesi ihtimali yüksek olacaktır. Aykırı gözlemler için ise  $d_{s_{MCD}^2}^2(x_i, \bar{x}_{MCD})$  değerleri büyük olacaktır [35].

### 3. UYGULAMA VE BULGULAR (APPLICATION AND RESULTS)

Bu çalışmada İzmir kent içi raylı sistem toplu taşımacılık hizmeti vermekte olan İzmir Banliyösü (İZBAN)'nden memnuniyeti etkileyen faktörlerin belirlenmesi amaçlanmaktadır. Bu amaç doğrultusunda çalışmada kullanılan veri seti İzmir'de İZBAN'dan faydalanan yolculara İZBAN'dan memnuniyetlerini etkileyebileceği düşünülen sorulardan oluşan bir anket uygulanarak elde edilmiştir. Anket 2014 yılında toplamda 500 yolcuya amaçlı örnekleme (purposive sampling) tekniği benimsenerek uygulanmıştır. Ancak anketlerden 34 tanesi uygun şekilde cevaplandırılmadığı için çalışma dışı bırakılmıştır. Anket iki kısımdan oluşmaktadır. İlk kısım yolcuların cinsiyet, yaş, meslek ve İZBAN'ı kullanım sıklığına yönelik sorulardan oluşmakta (Tablo 1), ikinci kısımda ise yolcuların memnuniyetinde etkili olabileceği düşünülen soruların yer aldığı 5'li likert tipi 35 sorudan oluşmaktadır.

Yolcuların İZBAN'dan memnuniyetlerinde etkili olabilecek faktörleri belirlemek amacıyla kullanılacak faktör analizinde tekniğin sağlaması gereken koşulları belirlemek için bazı hesaplamalar yapılmıştır.

Tablo 1. İzban kullanıcılarının demografik ve kullanım sıklığına ilişkin bilgileri (Information of demographic and usage frequency of Izban users)

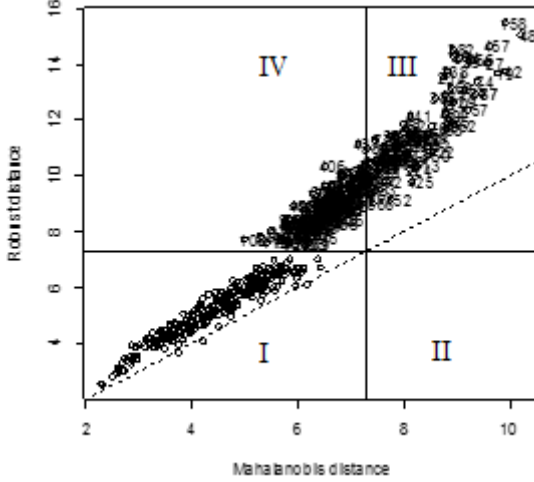
Değişken	Değişken düzeyi	Frekans (n=466)	Yüzde (%)
Cinsiyet	Kadın	221	47.4
	Erkek	245	52.6
Yaş	18-25 yaş	135	29.0
	26-35 yaş	104	22.3
	36-45 yaş	84	18.0
	46-55 yaş	78	16.7
	56+ yaş	65	14.0
Meslek	Kamu çalışanı	126	27.0
	Özelde çalışan	93	20.0
	Öğrenci	187	40.1
	Emekli	33	7.1
Kullanım sıklığı	Diğer	27	5.8
	Her gün	175	37.6
	Haftada dört-beş kez	187	40.1
	On beş günde bir	81	17.4
	Ayda bir kez ve daha az	23	4.9

Çalışmadaki tüm hesaplamalar R (ver. 3.2.4) yazılımı kullanılarak gerçekleştirilmiştir. İlk olarak değişkenler arasında yeterli ilişki olup olmadığı değerlendirilmiştir. Sağlayan Bartlett küresellik testi sonucundan ( $p=0.0001 < \alpha=0.05$ ) değişkenler arasında faktör analizi yapmak için yeterli düzeyde bir ilişki olduğu görülmüştür. Aynı zamanda Kaiser-Meyer-Olkin (KMO) örneklem yeterliliği testi (KMO değeri=0.921) sonuçları da incelenmiş olup, her iki test sonucuna göre veri setinin faktör analizine uygunluğuna karar verilmiştir.

Ayrıca veri setinde aykırı değer olup olmadığı incelenmiştir. Bu inceleme sağlam analiz gerekliliği için önemlidir. Aykırı değerleri belirlemek için hızlı MCD kestiricisine dayalı sağlam uzaklıklar hesaplanmıştır. Aykırı değerleri uzaklık grafiği yardımıyla Mahalanobis uzaklığına karşı sağlam uzaklıklar çizilerek görselleştirmek mümkündür (Şekil 1).

Eşik değeri (7.294) Şekil 1'i dört parçaya yatay ve dikey ekseninde ayırarak araştırmacılara aykırı değer tespitinde yardımcı olmaktadır. Şekil 1'e göre her iki uzaklık ölçüsü tarafından aykırı değer olarak belirlenen gözlemleri ve maskelenmiş aykırı değerleri yatay eşik değerinin (7.294) üstünde görmek mümkündür. Eğer veri seti kirlenmediyse her iki uzaklık ölçüsü de aynı sonuçları verir ve şekilde belirtilen I. kısım ( $MD(xi) \leq 7.294$  ve  $RD(xi) \leq 7.294$ ) veri setinde her iki uzaklık ölçüsü tarafından işaretlenmeyen iyi gözlemleri içerir. 2. kısım ( $MD(xi) > 7.294$  ve  $RD(xi) \leq 7.294$ ) klasik metod tarafından yanlışlıkla aykırı değer olarak belirlenen

gözlemleri gösterir ancak şekilde hiç gözlem yoktur. 3. kısım ( $MD(x_i) > 7.294$  ve  $RD(x_i) > 7.294$ ) her iki metot tarafından aykırı değer olarak belirlenen gözlemleri içerir. 4. kısım ( $MD(x_i) \leq 7.294$  ve  $RD(x_i) > 7.294$ ) maskelenmiş aykırı değerleri içerir. Diğer bir deyişle, Mahalanobis uzaklığı tarafından aykırı değer olarak belirlenmeyen ancak sağlam uzaklık tarafından aykırı değer olarak işaretlenen gözlemleri içerir. Şekilden de görüldüğü üzere veri seti çok sayıda aykırı değer içermektedir.



Şekil 1. Mahalanobis uzaklıklarına karşı sağlam uzaklıkların grafiği (Mahalanobis distance vs robust distance plot)

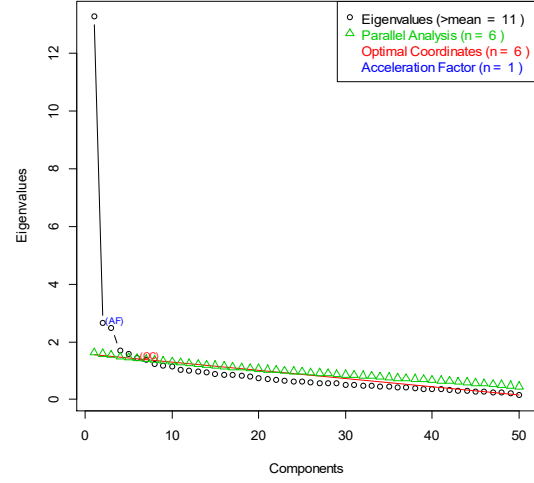
Faktör sayısını belirlemede en yaygın kullanılan iki metot özdeğerlerin 1'den büyük olanlarının seçildiği varyansa katılma kriteri (Kaiser's little jiffy) ve yamaç eğim testi (Scree test) aşırı büyük veya aşırı küçük faktör sayısı vermesinden dolayı güvenilir bulunmamaktadır [39]. Cattell [40] ise yamaç eğim testini subjektif bir metot olarak değerlendirmiş, uzmanlar tarafından bile geniş bir yorum yelpazesine sahip olduğu için güvenilirliğinin az olduğunu belirtmiştir [41]. Faktör sayısını belirlemede bu iki metoda alternatif istatistiksel metot olan paralel analiz ve optimal koordinatlar ise ele alınan uygulama için 6 faktörü önermektedir (Şekil 2).

Bu iki metodu dikkate alarak ve sağlam analizin gereklerinden biri olan aykırı değerlerin varlığının tespit edilmesiyle sağlam faktör analizi sonuçları ve klasik faktör analizi sonuçları hızlı MCD kestiricileri kullanılarak, temel bileşenler tekniği ile elde edilmiştir. Tablo 2'de varimax rotasyonlu faktör yükleri, değişkenlere özgü varyans (communality) ve her bir faktörün katkısı ile açıklanan toplam varyans bilgileri verilmiştir.

Tablo 2'ye göre klasik kestirici kullanılarak hesaplanan ilk 6 faktörün birikimli katkı oranı %51.6 iken, sağlam hızlı MCD kestirici ile bu değer %71'e ulaşmıştır.

Klasik faktör analizinin sonuçları aykırı değerlerden fazlasıyla etkilendiğinden, yorumlar sadece sağlam faktör analizi üzerinden yapılacaktır.

Non Graphical Solutions to Scree Test



Şekil 2. Grafiksel olmayan çözümlere karşı yamaç eğim grafiği (Non graphical solutions to scree test)

Tablo 2'den başlıca vagon içi aydınlatma, tutamak sayısı, tutamak yüksekliği, oturma yeri sayısı, oturma yerinin rahatlığı, vagon içinin temizliği, vagon iklimlendirme, vagon kapılarının geçiş rahatlığı, vagon içi boşlukların yeterliliğine ait değişkenlerin birinci faktör üzerinde etkili olduğu görülmektedir. Bu nedenle bu faktör "**araç (vagon) içi fiziksel koşullar (F3)**" olarak isimlendirilmiştir. İkinci faktörü İzban içindeki emniyet tedbirlerinin yeterliliği, duraklardaki güvenlik elemanı sayısının yeterliliği, İzban ile seyahat etmenin güvenliği, yolcu güvenliğini dikkate alan hızda ilerlemesi değişkenlerinin açıkladığı görülmektedir, dolayısıyla bu faktör "**güvenlik (F5)**" olarak isimlendirilmiştir. İzban içinde duraklara yaklaşırken yapılan anonsların zamanlaması, anonsların anlaşılabilirliği, vagon içi yol haritasının/varış duraklarının rahatça görünür yerde olması, geliş/varış süresi bilgilerinin doğruluğu, durak bilgilerinin okunurluğu ve anons sıklığının yeterliliği değişkenleri tarafından açıklanan üçüncü faktör ise "**yolcu bilgilendirme (F1)**" olarak isimlendirilmiştir. Dördüncü faktör durakların aydınlatılması, durak sayısı, durak yeri, duraklardaki oturma yeri sayısı, durakların temizliği, engelliler için düzenlemenin yapılması ve diğer ulaşım araçlarına ulaşılabilirlik değişkenlerine bağlı olarak ismi "**durak (istasyon) fiziksel koşulları (F2)**" olarak belirlenmiştir. Bilet fiyatlarının uygunluğu, indirimli biniş kartı fiyatlarının uygunluğu ve aktarma ücreti indirimi uygunluğu değişkenlerinden dolayı beşinci faktörün isminin "**ücret (F6)**" olmasına karar verilmiştir. Son olarak altıncı faktör hafta içi-hafta sonu hizmet saatlerinin yeterliliği, hafta içi-hafta sonu sefer

sıklığının yeterliliği, belirtilen sefer planına (zaman tablosuna) uygun çalışması ve aktarmalarda beklenen sürenin uygunluğu değişkenlerinden dolayı “**hizmet (sefer) süresi (sıklığı) (F4)**” olarak isimlendirilmiştir.

Tablo 2. Klasik ve sağlam faktör analizi sonuçları (Results of classical and robust factor analysis)

Değişkenler	Klasik faktör analizi						Sağlam faktör analizi							
	Faktör yükleri						Faktör yükleri							
	F1	F2	F3	F4	F5	F6	h <sup>2</sup>	F1	F2	F3	F4	F5	F6	h <sup>2</sup>
Vagon içi aydınlatma yeterliliği						0.582	0.524			0.397				0.515
Vagon içi tutamak sayısı yeterliliği					0.458		0.483			0.586				0.732
Vagon içi tutamak yüksekliği uygunluğu				0.466			0.406			0.632				0.634
Vagon içi oturma yeri sayısı yeterliliği				0.586			0.486			0.570				0.835
Vagon içi temizliği		0.560					0.585			0.509				0.806
Vagon içi oturma yeri rahatlığı		0.686					0.607			0.494				0.672
Vagon kapılarının geçiş rahatlığı		0.521					0.472			0.585				0.725
Vagon iklimlendirme yeterliliği		0.451					0.445			0.449				0.574
Vagon içi boşlukların yeterliliği		0.415					0.456			0.473				0.506
İzban içi emniyet tedbirleri yeterliliği	0.730						0.596				0.778			0.793
Durak güvenlik elemanı sayısı yeterliliği	0.785						0.686				0.824			0.941
Seyahat güvenliği	0.804						0.694				0.864			0.964
Yolcu güvenliği için İzban hızı						0.674	0.576				0.412			0.618
İzban içi anons zamanlaması			0.492				0.424	0.540						0.660
İzban içi anonsların anlaşılabilirliği			0.435				0.442	0.760						0.824
Vagon için yol haritası konumu		0.464					0.436	0.690						0.753
İzban geliş süresi bilgisi doğruluğu		0.740					0.599	0.771						0.750
Durak bilgilerinin okunurluğu		0.452					0.545	0.485						0.755
Anons sıklığının yeterliliği		0.642					0.487	0.662						0.670
Durakların aydınlatılması					0.670		0.603	0.560						0.546
Durak sayısı yeterliliği					0.710		0.604	0.622						0.535
Durak yeri uygunluğu			0.642				0.515	0.724						0.769
Durak oturma yeri sayısı yeterliliği					0.542		0.567	0.706						0.655
Durakların temizliği					0.580		0.488	0.574						0.538
Engelliler için düzenleme					0.445		0.586	0.509						0.606
Diğer ulaşım araçlarına ulaşılabilirlik		0.434					0.569	0.490						0.686
Bilet fiyatlarının uygunluğu	0.480						0.405					0.604	0.847	
İndirimli biniş kartı fiyat uygunluğu	0.427						0.376					0.633	0.752	
Aktarma ücret indirimi uygunluğu	0.601						0.442					0.865	0.959	
Hafta içi hizmet saatleri yeterliliği				0.463			0.565			0.654				0.847
Hafta sonu hizmet saatleri yeterliliği				0.613			0.577			0.806				0.897
Hafta içi sefer sıklığı yeterliliği							0.204			0.465				0.570
Hafta sonu sefer sıklığı yeterliliği				0.631			0.567			0.601				0.648
Sefer planına uygunluk				0.565			0.576			0.657				0.741
Aktarmada beklenen süre uygunluğu			0.524				0.453			0.508				0.540
Faktör yüklerinin kareler toplamı	3.515	3.385	3.347	2.889	2.672	2.240		4.892	4.780	4.419	4.337	3.916	2.522	
Açıklanan varyans oranı	0.100	0.097	0.096	0.083	0.076	0.064		0.140	0.137	0.126	0.124	0.112	0.072	
Birikimli varyans	0.100	0.197	0.293	0.376	0.452	0.516		0.140	0.277	0.403	0.527	0.639	0.711	

#### 4. SONUÇLAR (CONCLUSION)

Kovaryans ve korelasyon matrisi kestirimlerinin aykırı değerlerden oldukça fazla etkilendiği kolaylıkla görülebilir. Bunun sonucu olarak özdeğer ve özvektör kestirimleri de veri setindeki aykırı değerlerden etkilenmektedirler. Bu nedenle aykırı değerlerden etkilenmeyen sağlam bir yönteme ihtiyaç duyulmaktadır.

Aykırı değerlerin etkisinin faktör analizi sonuçları üzerindeki etkilerini görmek için klasik ve sağlam faktör analizi sonuçları aşağıdaki gibi karşılaştırılabilir.

Tablo 1'e göre klasik kestirici kullanılarak hesaplanan ilk 6 faktörün birikimli katkı oranı %51.6 iken, sağlam hızlı MCD kestirici ile bu değer %71'e ulaşmıştır.

Sağlam faktör analizinde değişkenlere özgü varyansın klasik faktördekilere göre daha büyük olduğu görülmektedir. Bu durum sağlam analizle her bir değişkenin faktörler içindeki açıklama miktarının arttığını göstermektedir. Klasik faktör analizinde en büyük açıklama oranı 0.100 iken, bu durum sağlam faktör analizinde 0.140 olarak görülmektedir.

Faktör analizinde faktörler katkı (açıklanan varyans) oranına göre sıraya dizildiğinde, aykırı değerlerin faktör sıralamasında değişiklik yapabildiği görülmektedir. Bir kaç faktörün konumu(sırası) klasik faktör analizinde ve

sağlam faktör analizinde farklı yerdedir, diğer bir deyişle  $F5 \rightarrow F2$  ve  $F1 \rightarrow F5+F6$  olduğu görülmektedir. Klasik analizdeki  $F1$  faktörünün, sağlam analizdeki  $F5$  ve  $F6$  faktörlerindeki soruların birleşiminden oluştuğu görülmektedir. Sağlam faktör analizindeki  $F3$  faktörünün değişkenlerinin klasik faktör analizinde birden fazla faktöre dağıldığı dikkat çekmektedir.  $F4$  faktörü her iki analizde de birikimli varyans sıralamasında aynı sırada yer almıştır. Her iki faktör analizinde de en az değişkene sahip faktör  $F6$ 'dır.

Veri setinde aykırı değerlerin olması durumunda, bu değerlerin etkisini azaltarak, değişkenlerin daha anlamlı bir şekilde bir arada toplanması ve daha fazla açıklama oranına sahip olması nedeniyle bu veri seti için sağlam faktör analizi kullanımı önerilmektedir.

Bu çalışma aykırı değerlerden dolayı normallik varsayımının bozulduğu durumlarda, aykırı değerlerin etkisini azaltarak verilerin çoğunluğuna uyan, parametre kestirimlerini yanlış yapmayacak özellikte sağlam faktör analizinin kullanılmasını önermektedir. Ayrıca bu çalışma ışığı altında, faktör analizinden sonra regresyon analizi ve yapısal eşitlik modellemesi gibi ileri istatistiksel teknikleri kullanacak araştırmacılar için de çalışmanın fayda sağlayacağı düşünülmektedir.

#### KAYNAKÇA (REFERENCES)

- [1] K. G. Jöreskog, "Some contributions to maximum likelihood factor analysis," *Psychometrika*, cilt 32, p. 443-482, 1967.
- [2] D. N. Lawley and A. E. Maxwell, Factor analysis as a statistical method, London: Butterworth, 1971.
- [3] D. J. Bartholomew and M. Knott, Latent variable models and factor analysis, London: Arnold, 1999.
- [4] G. Pison, P. J. Rousseeuw, P. Filzmoser and C. Croux, "Robust factor analysis," *Journal of multivariate analysis*, cilt 84, no. 7, pp. 145-172, 2003.
- [5] J. P. Stevens, "Outliers and influential data points in regression analysis," *Psychological Bulletin*, cilt 95, no. 2, pp. 334-344, 1984.
- [6] A. Christmann and S. Van Aelst, "Robust Estimation of Cronbach's Alpha," *Journal of Multivariate Analysis*, cilt 97, no. 7, p. 1660-1674, 2006.
- [7] D. Mavridis and I. Moustaki, "Detecting outliers in factor analysis using the forward search algorithm," *Multivariate behavioral research*, cilt 43, no. 3, pp. 453-475, 2008.
- [8] K. A. Bollen and G. Arminger, "Observational residuals in factor analysis and structural equation models," *Sociological methodology*, cilt 21, pp. 235-262, 1991.
- [9] K. A. Bollen, Structural equations with latent variables, New York: Wiley, 1989.
- [10] A. Basilevsky, Statistical factor analysis and related methods : theory and application, New York: Wiley, 1994.
- [11] K. M. Yuan and P. M. Bentler, "Structural equation modeling with robust covariances," *Sociological methodology*, cilt 28, pp. 363-396, 1998b.
- [12] K. G. Jöreskog, "Structural equation models in the social sciences: Specification, estimation and testing," in *Advances in factor analysis and structural equation models*, K. Jöreskog and D. Sörbom, Eds., Cambridge, Abt Books, 1979, pp. 105-127.
- [13] N. A. Champbell, "Robust procedures in multivariate analysis I: Robust covariance estimation," *Applied Statistics*, cilt 29, no. 3, pp. 231-237, 1980.
- [14] T. C. Cheng and M. P. Victoria-Feser, "High breakdown estimation of multivariate mean and covariance with missing observations," *British Journal of Mathematical and Statistical Psychology*, cilt 55, pp. 317-335, 2002.
- [15] P. L. Davis, "Asyptotic behaviour of S-estimators of multivariate location parameters and dispersion matrices," *Annals of Statistics*, cilt 15, pp. 1269-1292, 1987.
- [16] S. J. Devlin, R. Gnanadesikan and J. Kettenring, "Robust estimation of dispersion matrices and principal components," *Journal of the american statistical association*, cilt 76, no. 374, pp. 354-362, 1981.
- [17] D. L. Donoho, "Breakdown properties of multivariate location estimators," *Ph.D. Qualifying paper*, p. 69, 1982.
- [18] A. S. Hadi, "Identifying multiple outliers in multivariate data," *Journal of the Royal Statistical Society, Series B*, cilt 54, no. 3, p. 761-771, 1992.
- [19] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, Robust statistics, the approach based on influence functions, New York: Wiley, 1986.
- [20] P. J. Huber, Robust statistics, New York: Wiley, 1981.



- [21] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *Annals of statistics*, cilt 4, pp. 51-67, 1976.
- [22] P. J. Rousseeuw, "Multivariate estimation with high breakdown point," in *Mathematical statistics and applications*, W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, Eds., Dordrecht:, Reidel, 1985, p. 283–297.
- [23] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*, New York: Wiley, 1987.
- [24] P. J. Rousseeuw and B. C. Van Zomeren, "Unmasking multivariate outliers and leverage points," *Journal of the American Statistical Association*, cilt 85, pp. 633-651, 1990.
- [25] W. A. Stahel, *Robust estimation: Infinitesimal optimality and covariance matrix estimators*, Zurich: Ph.D Dissertation, ETH, 1981.
- [26] D. Woodruff and D. M. Roche, "Computable robust estimation of multivariate location and shape in high dimension using compound estimators," *Journal of the American Statistical Association*, cilt 89, p. 888–896, 1994.
- [27] P. Filzmoser, "Robust principal component and factor analysis in the geostatistical treatment of environmental data," *Environmetrics*, cilt 10, p. 363–375, 1999.
- [28] I. Moustaki and M. P. Victoria-Feser, "Bounded-influence robust estimation in generalized linear latent variable models," *Journal of the American Statistical Association*, cilt 101, no. 474, p. 644–653, 2006.
- [29] K. M. Yuan and P. M. Bentler, "Robust mean and covariance structure analysis," *British Journal of Mathematical and Statistical Psychology*, cilt 51, p. 63–88, 1998a.
- [30] K. M. Yuan and P. M. Bentler, "Effects of outliers on estimators and tests in covariance structure analysis," *British Journal of Mathematical and Statistical Psychology*, cilt 54, p. 161–175, 2001.
- [31] R. Kosfeld, "Robust exploratory factor analysis," *Statistical Papers*, cilt 37, pp. 105-122, 1996.
- [32] D. Hawkins, "The feasible solution algorithm for the minimum covariance determinant estimator in multivariate data," *Computational Statistics and Data Analysis*, cilt 17, pp. 197-210 , 1994.
- [33] A. Atkinson, "Fast very robust methods for the detection of multiple outliers," *Journal of the American Statistical Association*, cilt 89, p. 1329–1339, 1994.
- [34] P. J. Rousseeuw and K. Van Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, cilt 41, no. 3, pp. 212-223, 1999.
- [35] J. Hardin and D. M. Roche, "The Distributions of Robust Distances," *Journal of Computational and Graphical Statistics*, cilt 14, no. 4, pp. 1-19 , 2005.
- [36] J. S. Tanaka and Y. Odaka, "Influential observations in principal factor analysis," *Psychometrika*, cilt 54, p. 475–485, 1989.
- [37] H. P. Lopuhaä and P. J. Rousseeuw, "Breakdown points of affine equivariant estimators of multivariate location and covariance matrices," *The Annals of Statistics*, cilt 19, no. 1, pp. 229-248 , 1991.
- [38] A. Cerioli, "Multivariate outlier detection with high-breakdown estimators," *Journal of the American Statistical Association-Theory and Methods*, cilt 105, no. 489, pp. 147-156, 2010.
- [39] B. Thomson, *Exploratory and confirmatory factor analysis: Understanding concepts and applications.*, Washington, DC: American Psychological Association, 2004.
- [40] R. B. Cattell, "The scree test for the number of factors," *Multivariate Behavioural Research*, cilt 1, pp. 245-276, 1966.
- [41] D. L. Streiner, "Factors affecting reliability of interpretations of scree plots," *Psychological Reports*, cilt 83, pp. 687-694, 1998.