EN KÜÇÜK KARELER VE TEMEL BİLEŞENLER REGRESYON ANALİZLERİNİN KARŞILAŞTIRILMASI

Zeynep TUNÇ

BİYOİSTATİSTİK ve TIP BİLİŞİMİ ANABİLİM DALI

Tez Danışmanı
Dr. Öğr. Üyesi Harika Gözde GÖZÜKARA BAĞ

Yüksek Lisans Tezi – 2018

T.C.

İNÖNÜ ÜNİVERSİTESİ

SAĞLIK BİLİMLERİ ENSTİTÜSÜ

EN KÜÇÜK KARELER VE TEMEL BİLEŞENLER REGRESYON ANALİZLERİNİN KARŞILAŞTIRILMASI

Zeynep TUNÇ

Biyoistatistik ve Tıp Bilişimi Anabilim Dalı

Yüksek Lisans Tezi

Tez Danışmanı

Dr. Öğr. Üyesi Harika Gözde GÖZÜKARA BAĞ

MALATYA

2018

KABUL VE ONAY SAYFASI

İnönü Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik ve Tıp Bilişimi Anabilim Dalı Yüksek Lisans Programı çerçevesinde yürütülmüş olan; **Zeynep TUNÇ 'un "En Küçük Kareler ve Temel Bileşenler Regresyon Analizlerinin Karşılaştırılması"** konulu bu çalışması, aşağıdaki jüri tarafından Yüksek Lisans tezi olarak kabul edilmiştir.

Tez Savunma Tarihi: 18/12/2018

Prof.Dr.Cemil ÇOLAK
İnönü Üniversitesi
Jüri Başkanı

Dr.Öğr.Üyesi Harika Gözde GÖZÜKARA BAĞ
İnönü Üniversitesi
Tez Danışmanı
Üye

Dr.Öğr.Üyesi Adem DOĞANER
Kahramanmaraş Sütçü İmam Üniversitesi
Üye

ONAY

Bu tez, İnönü Üniversitesi Lisansüstü Eğitim-Öğretim Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından kabul edilmiş ve Enstitü Yönetim Kurulu'nun ...../...../2018 tarih ve 2018/..... sayılı Kararıyla da uygun görülmüştür.

Prof. Dr. Yusuf TÜRKÖZ
Enstitü Müdürü

İÇİNDEKİLER

TEŞEKKÜR

Akademik eğitimim ve çalışmalarımın yanında günlük yaşantımda bilgi, birikim ve deneyimleri ile bana yol gösteren ve destek olan değerli danışman hocam Sayın Dr. Öğretim Üyesi Harika Gözde GÖZÜKARA BAĞ'a, eğitimim boyunca desteklerini esirgemeyen ve önerileriyle bana ışık tutan değerli hocalarım Prof. Dr. Saim YOLOĞLU ve Prof. Dr. Cemil ÇOLAK'a, aynı bölümde görev yaptığım çok değerli asistan arkadaşlarıma sonsuz saygı ve teşekkürlerimi sunarım. Bu süreçte yardımını hiç esirgemeyen, destekleriyle beni hiçbir zaman yalnız bırakmayan aileme ve ikizlerim Okyanus Balın TUNÇ ve Rüzgar Diren TUNÇ'a sonsuz teşekkürlerimi sunarım.

Arş. Gör. Zeynep TUNÇ

ÖZET

En Küçük Kareler ve Temel Bileşenler Regresyon Analizlerinin Karşılaştırılması

Amaç: Bu çalışmanın amacı, veride çoklu bağlantı olduğunda En Küçük Kareler (EKK) Regresyonu ile Temel Bileşenler Regresyonu (TBR) sonuçlarının karşılaştırılmasıdır.

Materyal ve Metot: Çoklu bağlantının derecesinin ve örneklem genişliğinin etkisinin incelenmesi amacıyla iki farklı veri grubu türetilmiştir. Birinci veri grubu; farklı çoklu bağlantı düzeyine sahip 10 veri setinden, ikinci veri grubu; aynı korelasyon yapısına sahip ancak örneklem genişliği farklı 10 veri setinden oluşmaktadır. Üç bağımsız ve bir bağımlı değişkenden oluşan tüm veri setleri için değişkenler standart normal dağılımdan türetilmiştir. Türetilen verilerde çoklu bağlantının varlığı yaygın olarak kullanılan ölçüler ile ispatlanmıştır. Tüm veri setlerine En Küçük Kareler ve Temel Bileşenler Regresyonu uygulanmıştır.

Bulgular: Çoklu bağlantı elde edebilmek için yapılan veri türetiminde tüm ilişkiler pozitif yönde tanımlanmıştır. Ancak, En Küçük Kareler çözümlemesinde çoklu bağlantının beklenen etkilerinden biri olarak ikinci (X_2) ve üçüncü (X_3) bağımsız değişkenler için regresyon katsayılarının işareti ters (negatif) olacak şekilde elde edilmiştir. Temel Bileşenler Regresyonu çözümlemesinde ise katsayıların işareti doğru yönde (pozitif) bulunmuştur. EKK çözümlemesinde elde edilen katsayılar ile TBR analizi sonucunda elde edilen katsayılar işaretçe farklı olmakla beraber büyüklük olarak da birbirinden farklıdır. Ayrıca, TBR sonuçlarında katsayıların standart hataları EKK sonuçlarına göre daha düşüktür.

Sonuç: Çoklu doğrusal regresyon çözümlemesi yapılırken çoklu bağlantının varlığı mutlaka incelenmeli ve bu duruma çözüm olabilecek yöntemlerden biri kullanılmalıdır. Aksi taktirde yapılacak kestirimler yanlış sonuçlara götürebilecektir. Yapılan bu çalışmanın sonuçları doğrultusunda veride çoklu bağlantı olduğu durumda karşılaştırılan iki yöntemden En Küçük Kareler regresyonu yerine Temel Bileşenler Regresyonunun kullanılması önerilmektedir.

Anahtar Kelimeler: Çoklu bağlantı, Doğrusal regresyon, En Küçük Kareler, Örneklem genişliği, Temel Bileşenler Regresyonu.

ABSTRACT

Comparison of Ordinary Least Squares and Principal Components Regression Analyses

Aim: The aim of this study is to compare the results of Ordinary Least Squares (OLS) and Principal Components Regression (PCR) analyses when there is multicollinearity in the data.

Material and Method: Two different data groups were simulated in order to examine the effect of the degree of multicollinearity and the sample size. The first data group consisted of 10 data sets with different multicollinearity degree and the second data group consisted of 10 data sets with the same correlation structure but with different sample sizes. All datasets had one dependent and three independent variables, and all the variables were derived from standard normal distribution. The presence of multicollinearity in the derived data was proven by commonly used measures. The least squares and principal components regression were applied to all datasets.

Results: When generating multicollinearity, all relationships were defined as positive in data simulation. However, the sign of the regression coefficients for the second (X_2) and third (X_3) independent variables were obtained as reverse (negative) as one of the expected effects of multicollinearity in Least Squares analysis. In the analysis of the Principal Components Regression, the sign of coefficients was found to be in the right direction (positive). The sign of the coefficients obtained from OLS and PCR were different and they also differed in magnitude. In addition, the standard errors of the coefficients in PCR results were lower than OLS results.

Conclusion: In the case of multiple linear regression analysis, the existence of multicollinearity must be examined and one of the methods that can handle this problem should be used. Otherwise, predictions may lead to incorrect results. Based on the results of this study that compares two methods when there is multicollinearity in data, it is recommended to use Principal Components Regression instead of Ordinary Least Squares.

Key words: Multicollinearity, Linear regression, Ordinary least squares, Sample size, Principal components regression.

SİMGELER VE KISALTMALAR DİZİNİ

EKK:	En Küçük Kareler
TBR:	Temel Bileşenler Regresyonu
VIF:	Variance Inflation Factors=Varyans Şişme Faktörü
OLS:	Ordinary Least Squares
PCR:	Principle Components Regression

ŞEKİLLER DİZİNİ

TABLOLAR DİZİNİ

1. GİRİŞ

Sağlık alanında yapılan bazı çalışmalar, bağımlı değişken ile bağımsız değişken ya da değişkenler arasındaki ilişkilerin belirlenmesi ve aralarındaki ilişkinin matematiksel bir model yardımıyla ifade edilmesi temeline dayanır. Bu modelin elde edilmesi için kullanılan yöntem ise regresyon çözümlemesi olarak adlandırılır. Bu tür çalışmalarda en temel amaç bağımlı değişkeni, bağımsız değişken(ler) yardımıyla kestirebilmektir. Birden fazla bağımsız değişken olduğunda bir diğer amaç, hangi bağımsız değişken(ler)in bağımlı değişkeni daha çok etkilediğini belirlemek olabilir (1).

Bağımlı değişken Y'yi açıklamak için tek bir bağımsız değişken X kullanılacaksa bu yönteme basit regresyon çözümlemesi adı verilirken, iki ya da daha çok bağımsız değişken X kullanılacaksa çoklu regresyon çözümlemesi olarak adlandırılır. Örneğin, vücut yağ yüzdesini kestirmek için bağımsız değişken olarak sadece vücut ağırlığı modele alınırsa bu iki değişken arasında basit regresyon modeli kurulabilir. Vücut yağ yüzdesini kestirmek için vücut ağırlığına ek olarak modele boy uzunluğu ve cinsiyet de eklenirse kurulan model çoklu regresyon modeli olur.

Bağımlı ve bağımsız değişken(ler) arasında model kurmadan önce aralarındaki ilişkinin nasıl olduğunun belirlenmesi gerekir. Bu amaçla en sık kullanılan yöntem bağımlı değişken ile bağımsız değişken arasındaki ilişkinin şeklini, yönünü ve kuvvetini gösteren saçılım grafiği çizilmesidir.

Saçılım grafiği yardımıyla değişkenler arasındaki ilişkinin doğrusal ya da doğrusal olmadığı gözlemlenebilir. Bağımlı ve bağımsız değişken(lerin) arasındaki ilişkinin şekline bağlı olarak doğrusal olan ya da doğrusal olmayan regresyon çözümleme yöntemleri kullanılır.

Regresyon yöntemlerinin sınıflandırılması aşağıdaki değişik şekillerde verilmektedir:

- Doğrusal/doğrusal olmayan regresyon yöntemleri

1) Doğrusal regresyon yöntemleri, regresyon modelinde yer alan bağımsız değişken/değişkenlerin Y_i bağımlı değişkene/değişkenlere etkilerini doğrusal ve eklenebilir formda ele alan regresyon yöntemlerini içerir.
2) Doğrusal olmayan (eğrisel) regresyon yöntemleri, regresyon modelinde yer alan bağımsız değişken/değişkenlerin Y_i bağımlı değişkene/değişkenlere etkilerinin

toplanabilir olmadığını (çarpımsal, eğrisel, üssel) varsayan regresyon yöntemlerini içerir.

- Parametrik/parametrik olmayan regresyon yöntemleri

1) Parametrik regresyon yöntemleri, bağımlı değişkenin/değişkenlerin normal dağılım/çok değişkenli normal dağılım göstermesini ön koşul kabul eden regresyon yöntemlerini içerir.
2) Parametrik olmayan regresyon yöntemleri, bağımlı değişkenin/değişkenlerin normal dağılım/çok değişkenli normal dağılım göstermesini ön koşul olarak ileri sürmeyen regresyon yöntemlerini içerir.

- Basit/çoklu/çok değişkenli regresyon yöntemleri

1) Basit (simple) regresyon yöntemleri, regresyon modelinde bir bağımlı bir bağımsız değişken olması durumunda oluşan doğrusal ve eğrisel regresyon modellerini içerir.
2) Çoklu (multiple) regresyon yöntemleri, regresyon modelinde bir bağımlı birden çok bağımsız değişken olması durumundaki doğrusal ve eğrisel regresyon modellerini içerir.
3) Çok değişkenli (multivariate) regresyon yöntemleri, regresyon modelinde birden çok bağımlı değişken ve bir ya da daha çok bağımsız değişken olması durumundaki doğrusal ve eğrisel regresyon modellerini içerir (2).

Bu çalışmada, çoklu doğrusal regresyon çözümlemesi dikkate alınmaktadır. Doğrusal regresyon çözümlemesinde en yaygın kullanılan yöntemlerden biri olan en küçük kareler yöntemi, gözlem değerleri, değişkenler ve hataların dağılımı hakkında birtakım varsayımların sağlandığı durumlarda geçerlilik kazanır. Bu varsayımlar geçerli olmadıkça elde edilen sonuçlar güvenilir olmaz. Çünkü varsayımların bozulmasının kestirilen parametreler üzerine çok önemli etkileri olabilmektedir. Buna bağlı olarak elde edilen regresyon denkleminden yapılacak kestirimlerin hatalı olma olasılığı yüksek olur (3). Bu yöntemin varsayımlarından biri bağımsız değişkenler arasında kuvvetli bir ilişki olmamasıdır. Bağımsız değişkenler arasında bir ya da daha fazla doğrusal bağıntının olması çoklu bağlantı (multicollinearity) sorununu gündeme getirir (1).

Çoklu regresyon denkleminin yorumu bağımsız değişkenlerin kuvvetli bir şekilde ilişkili olmaması varsayımına bağlıdır. Bu varsayımın bozulması, yani bağımsız değişkenler arasında bir ya da daha fazla doğrusal bağıntının olması çoklubağlantı sorununu gündeme getirir. Bağımsız değişkenler arasında ilişki olmaması durumunda bu değişkenlerin dik (ortogonal) olduğu söylenir. Ancak uygulamaların çoğunda bağımsız değişkenler arasında

ilişki söz konusudur. Hatta bazı durumlarda bağımsız değişkenler arasında çok kuvvetli doğrusal ilişki vardır ve böyle durumlarda regresyon modeli yardımıyla yapılacak çıkarsamalar yanlış yönlendirmelere ve hatalara neden olur (1).

Bu çalışmanın amacı, bağımsız değişkenler arasında çoklu bağlantı olması durumunda, doğrusal regresyon çözümlemesinde sıklıkla kullanılan en küçük kareler (EKK) yönteminin sonuçlarının nasıl etkilendiğini göstermektir. Bir diğer amaç ise çoklu bağlantı sorunu olması durumunda kullanılan temel bileşenler regresyonu (TBR) ile EKK yöntem sonuçlarının karşılaştırılmasıdır. Bu amaçla, benzetim tekniği yardımıyla örneklem genişliği 1000 ve çoklu bağlantı derecesi farklı olan 10 veri seti türetilmiştir. Aynı zamanda, çoklu bağlantının örneklem genişliğine göre etki derecesindeki değişimini gözlemlemek amacıyla aynı çoklu bağlantı düzeyinde örneklem genişliği 1000-10000 arasında değişen 10 veri seti türetilmiştir. Benzetim tekniği ile elde verilen setlerine hem en küçük kareler regresyon çözümlemesi hem de temel bileşenler regresyonu uygulanarak sonuçları karşılaştırılmıştır.

2. GENEL BİLGİLER

2.1. Regresyon

Regresyon terimi 19. yüzyılda İngiliz istatistikçisi Francis Galton tarafından bir biyolojik inceleme için ortaya atılmıştır. Bu incelemenin ana konusu kalıtım olup, aile içinde baba ve annenin boyu ile çocukların boyu arasındaki bağlantıyı araştırmakta ve çocukların boylarının bir nesil içinde eski ata nesillerinin ortalamasına geri döndüklerini yani bir nesil içinde ortalamaya geri dönüş olduğu inceleme konusudur. Galton geri dönüş terimi için ilk yazısında İngilizce olarak "reversion" terimi kullanmışsa da sonradan aynı anlamda olan "regression" sözcüğünü kullanmıştır. Bu çalışmalarında Galton istatistiksel 'regresyon' kavramını ve yöntemini de geliştirmiştir (4, 5).

Regresyon çözümlemesinde iki değişken türü söz konusudur. Bunlar, bağımlı ve bağımsız değişken kavramlarıdır. Bir ya da daha fazla faktörün (etkenin) etkisiyle oluşabilen ve bu faktör(ler)le ilişkisi aranan değişkene bağımlı değişken adı verilirken, bu bağımlı değişkeni etkilediği düşünülen değişken(ler)e bağımsız değişken adı verilir (6). Farklı kaynaklarda, bağımlı değişken; etkilenen, açıklanan veya sonuç değişkeni olarak da anılırken, bağımsız değişken; etkileyen, açıklayan, neden olan değişken isimleri ile de tanımlanmaktadır.

Regresyon çözümlemesi ile bağımlı ve bağımsız değişkenler arasında bir ilişki var mıdır? Eğer bir ilişki varsa bu ilişkinin gücü nedir? Değişkenler arasında ne tür bir ilişki vardır? Bağımlı değişkene ait ileriye dönük değerleri kestirmek mümkün müdür ve nasıl kestirim yapılmalıdır? Belirli koşulların kontrol edilmesi durumunda özel bir değişken veya değişkenler grubunun diğer değişken veya değişkenler üzerindeki etkisi nedir ve nasıl değişir? ve benzeri sorulara cevap aranmaya çalışılır (7).

2.2. Basit Doğrusal Regresyon Analizi

Aralarında doğrusal bir ilişki olan bir bağımlı ve bir bağımsız değişken için kurulan matematiksel model basit doğrusal regresyon çözümlemesi adını alır.

n tane birimden oluşan örneklemden her birini kullanarak bağımlı değişken Y ve bağımsız değişken X değerleri saptanmış olsun. Bu durumda (Y_1,X_1), (Y_2,X_2), ... , (Y_n,X_n) olmak üzere n tane gözlem elde edilmiş olacaktır. Bu durumda Y ve X değişkenleri arasındaki ilişki nasıl bir ilişkidir? Bu ilişkiyi matematiksel şekilde ifade etmenin bir yolu var mıdır? Bu soruların yanıtlarını verebilmek için (Y_i,X_i) i=1,2, ... ,n gözlem çiftini koordinat eksenlerinde göstermek gerekir.

Bu işlemi yaparken kullanılan grafiğe saçılım grafiği denir. n tane gözlem çifti için saçılım grafiğinde kesişim noktaları bulunduğunda n tane nokta oluşacaktır. Bu noktaların konumuna bakılarak modelin nasıl olduğuna karar verilir. Eğer noktalar bir doğru etrafında toplanıyorsa modelin doğrusal olduğu söylenir (8).

X ve Y arasındaki doğrusal ilişki basit doğrusal regresyon modeli ile fonksiyonel olarak aşağıdaki gibi ifade edilir.

$$Y_i = \beta_0 + \beta_1 X_i \qquad i = 1, 2, \dots, n \qquad (1)$$

Burada; β_0 ve β_1 regresyon katsayılarıdır. β_0, regresyonun doğrusunun y eksenini kestiği noktayı göstermektedir ve sabit veya kesim noktası olarak da adlandırılmaktadır. β_1 ise regresyon doğrusunun eğimidir ve bağımsız değişken X'de bir birim değişiklik olduğunda bağımlı değişken Y'deki değişimi ifade etmektedir (1).

Eşitlik 1'deki regresyon denklemine göre X bağımsız değişkeni bağımlı değişken Y'yi kesin bir şekilde belirlemektedir. Ancak, iki değişken arasında gerçek dünyada bu tür ilişkilerle nadiren karşılaşılır (2).

Örneğin, yukarıdaki ilişkide bağımlı değişken Y diastolik kan basıncı, bağımsız değişken X yaş olsun. Yukarıdaki ilişki bu haliyle eksik kalacaktır. Çünkü diastolik kan basıncını belirleyen, ek başka hastalıkların olması, kullanılan ilaçlar, cinsiyet vb. gibi başka faktörler de olabilir. Öte yandan tansiyon aletinin ölçümünden kaynaklanan hatalardan dolayı diastolik kan basıncının ölçülmesinde hatalar yapılmış olabilir.

Yukarıda sayılan ve fonksiyona dahil edilmemiş faktörlere rassal faktörler denir ve önceden bilinmezler. Bu rassal etkiler, istatistiksel bir ilişki kurulurken modele bir rassal terim olarak ilave edilir. Bu rassal terimi ε_i ile gösterelim. Bu durumda X bağımsız değişkeni ve Y bağımlı değişkeni arasındaki, (9).

Gerçek ilişki $\qquad Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad i = 1, 2, \dots, n \qquad (2)$

Gerçek regresyon $\qquad E(Y_i) = \beta_0 + \beta_1 X_i \qquad i = 1, 2, \dots, n \qquad (3)$

Bu şekilde Y bağımlı değişkeninde meydana gelen değişmeler rassal değişmelerin toplam etkisiyle gerçek bir şekilde ortaya konmuş olur. Yukarıda bahsedilen gerçek ilişki ve gerçek regresyon doğrusu Y ve X değişkenlerine ait bütün anakitle verileri elde edilemediği sürece bilinemezler. Ancak, aynı model X ve Y değişkenine ait anakitlelerden alınmış bir grup gözlemle(örneklem) aşağıdaki gibi tanımlanabilir.

Kestirilen ilişki $Y_i = b_0 + b_1 X_i + e_i \qquad i = 1, 2, \dots, n$ (4)

Kestirilen regresyon $\hat{Y}_i = b_0 + b_1 X_i \qquad i = 1, 2, \dots, n$ (5)

Burada bağımlı değişken *Y*, bağımsız değişken *X* üzerine bağlanmış olup, regresyon denkleminde b_0 ve b_1 sırasıyla gerçek ilişkideki β_0 ve β_1 parametrelerinin, e_i ise ε_i'nin kestirimidir. e_i parametresine regresyon artıkları adı verilir. (Y_i,X_i) gözlemleri grafik üzerinde işaretlendiğinde $\hat{Y}_i = b_0 + b_1 X_i$ ile verilen regresyon doğrusundan sapmalarının nedeni regresyon artıkları adı verilen e_i parametreleridir. Diğer bir deyişle, $Y_i = \hat{Y}_i + e_i$'dir. Regresyon doğrusunun üzerinde yer alan gözlem değerleri için e_i ler pozitif, altında kalan gözlem değerleri için e_i ler negatif olmakla beraber, e_i lerin cebirsel toplamı sıfırdır.

Basit doğrusal regresyon 2 farklı amaç için kullanılabilir (1):

1) Kestirim yapmak,
2) *X* bağımsız değişkeninde bir birim artış olduğunda *Y* bağımlı değişkendeki değişiklik miktarını gösteren b_1 katsayısını kestirmek,

2.2.1 Basit Doğrusal Regresyon İçin Varsayımlar

Basit doğrusal regresyon çözümlemesinin bazı varsayımları aşağıda belirtilmiştir.

1) Bağımsız değişkenin değerleri hatasızdır yani hatasız ölçülür; ancak hiçbir ölçüm işleminde mükemmel ölçüm yapılamadığı için bu ifade şöyle açıklanabilir: bağımsız değişkendeki ölçüm hatalarının önemsenmeyeceği düşünülür.
2) Bağımsız değişkenin her bir değeri için birden çok bağımlı değişken değeri vardır. Yapılan kestirimlerin ve kurulan hipotez testlerinin geçerli olabilmesi için bu alt kümelerin normal olarak dağılması gerekir.
3) Bağımsız değişkenin her bir değerine karşılık gelen bağımlı değişken değerlerinin alt kümelerinin varyansları homojenlik gösterir.
4) Bağımlı değişkenin alt kümeleri bir ortalama üzerinde dağılır (10).

2.3. Çoklu Doğrusal Regresyon Modeli

Sağlık alanında yer alan bağımlı değişkenler genellikle iki ya da daha çok bağımsız değişken tarafından etkilenebilmektedir. Biyolojik sistem karmaşık bir etkileşim gösterir. Sağlık alanında bir değişkeninin değeri çok sayıda değişkenin etkileşimi sonucu ortaya çıkmaktadır. Bunlardan bazıları çok daha önemli etkilere sahip olan değişkenler iken diğerleri daha az öneme sahip ya da önemsiz etkiye sahip olan değişkenlerdir. Bir değişkeni etkileyen

iki veya daha fazla bağımsız değişken arasındaki neden- sonuç ilişkilerini doğrusal bir modelle açıklamak ve bu bağımsız değişkenlerin etki düzeylerini belirleyebilmek için yararlanılan metoda **çoklu doğrusal regresyon** analizi denir (11).

Bağımlı değişkenin birden fazla bağımsız değişken tarafından etkilendiği çoklu doğrusal regresyon analizinde, araştırmacıların üç genel amacı vardır (12)**:**

1) Bağımsız değişkenlerden hangisi ya da hangilerinin bağımlı değişkeni daha çok açıkladığını belirlemek,
2) Bağımlı değişkeni etkilediği belirlenen bağımsız değişkenler ile bağımlı değişkenin değerini kestirebilmek,
3) Veriyi özetlemek.

Bir örnek ile yukarıdaki durumları açıklamak istersek bağımlı değişken olarak anne karnındaki bir bebeğin doğum ağırlığını aldığımızı düşünelim. Ve bu ağırlığı tahminlemeyi amaçlayalım. Bu doğum ağırlığını önceden kestirebilmek için gebenin gebelik süresince beslenme durumu dikkate alınarak iki değişken arasında bir regresyon modeli oluşturulsun. Annenin beslenme durumunun, eğer bebeğin doğum kilosunu açıklamakta yetersiz kaldığı görülürse bağımlı değişkenimiz doğum ağırlığına etkisi olduğu düşünülen, anne yaşı, gebelik haftası, gebelik sayısı, canlı doğum sayısı gibi farklı bağımsız değişkenleri de modele ekleyerek çoklu regresyon denklemi oluşturulabilir. Örnek için birinci amaç, kurulan çoklu regresyon denklemi ile bebeğin doğum kilosunun en çok hangi faktörden etkilediğini bulmak, ikinci amaç da bebeğin doğum kilosunu önceden belirleyerek riskli gebelikleri belirleyerek gebeliklere zamanında müdahaleler yapabilmektir (13).

Y bağımlı değişken $X_1, X_2, \dots, X_p$ ler bağmsız değişkenler olmak üzere çoklu regresyon denklemi

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon \qquad (1)$$

ile verilir. Denklem de yer alan β_j, $j = 0,1, \dots, p$ değerlerine regresyon katsayıları denir. β_j değerleri, $i \neq j$ olmak üzere tüm Xi bağımsız değişkenleri sabit olduğunda, Xj deki her bir birimlik değişime karşılık Y bağımlı değişkenindeki beklenen değişimi gösterir. Bu nedenle β_j değerlerine kısmi regresyon katsayıları da denir (14).

Çoklu regresyon modelinde verilerin tablo ve matrisler yardımıyla gösterimi aşağıdaki şekilde olur

Gözlem	Y	X_1	X_2	…	X_p
1	y_1	x_{11}	x_{12}	…	x_{1p}
2	y_2	x_{21}	x_{22}	…	x_{2p}
3	y_3	x_{31}	x_{32}	…	x_{3p}
.					
.					
.					
n	y_n	x_{n1}	x_{n2}	…	x_{np}

Bu gösterim denklem ile ifade edilecek olursa, bir başka deyişle regresyon denklemi gözlemler cinsinden

$$y_i = \beta_0 + \beta_1 x_{ij} + \beta_2 x_{i2} + \cdots + \beta_j x_{ij} + \cdots + \beta_p x_{ip} + \varepsilon_i \qquad (2)$$

şeklinde gösterilir.

Ayrıca bağımlı değişkenimiz Y, $n\,x\,1$ vektör ve bağımsız değişken kümemiz X, $n\,x\,(p+1)$ boyutlu matris, β $(p+1)\,x\,1$ boyutlu katsayılar vektörü ve ε, $(n\,x\,1)$ boyutlu hata vektörü olmak üzere regresyon denklemi

$$Y = \beta X + \varepsilon \qquad (3)$$

şeklinde yazılır.

2.3.1. Çoklu Doğrusal Regresyon Modelinin Varsayımları

Basit doğrusal regresyonda da olduğu gibi çoklu doğrusal regresyon için de parametre kestirimleri yapılırken ilk önce bazı varsayımların sağlanıp sağlanmadığı kontrol edilmelidir. Varsayımların yerine getirilmemiş olması bazı problemleri ortaya çıkabilir ve bu problemler model üzerinde bazı olumsuz sonuçlar oluşturabilir. Bahsedilen varsayımlar aşağıdaki gibidir (15).

1) Hata terimlerinin aritmetik ortalaması sıfır olmalıdır.
2) Hata terimleri normal bir dağılım göstermelidir.
3) Hata terimlerinin varyansı sabit olmalıdır.
4) Hata terimleri birbirinden bağımsız olmalıdır.
5) Gözlem sayısı parametre sayısından büyük olmalıdır.
6) Bağımlı değişken ile bağımsız değişkenler arasında doğrusal bir ilişki olmalıdır.
7) Bağımsız değişkenler arasında ilişki olmamalıdır.

2.3.1.1. Hata Terimlerinin Aritmetik Ortalamasının Sıfır Olması

Hata terimi gözlem değerlerinin her bir değeri için farklı farklı değerler alabilir. Elde edilen regresyon doğrusunun altında kalan gözlem değerleri için elde edilen hatalar negatif değerler alırken, regresyon doğrusunun üstünde kalan gözlem değerleri için hesaplanan hata terimleri pozitif değerler alır. Yukarıda adı geçen varsayım tüm bu hata terimlerinin yani ε_i değerlerinin cebirsel toplamının sıfır olması varsayımıdır.

Bu varsayımın sağlanması koşuluyla örneklemden yola çıkılarak kestirimi sağlanan regresyon doğrusu anakütle doğrusu için iyi bir kestirim olabilmektedir. Bu varsayımın sağlanamaması durumunda elde edilen regresyon modeliyle bulunan parametre değerleri gerçek değerlerinden, hataların negatif olması durumunda daha küçük, pozitif olması durumunda daha büyük olacak şekilde elde edilir. Diğer bir deyişle, parametre kestirimleri sapmalı kestirimler olarak elde edilir (3).

2.3.1.2. Hata Terimlerinin Normal Dağılması

Aralık tahmini ve regresyon katsayılarıyla ilgili testlerin yapılabilmesi için hata terimlerinin dağılımının, ortalaması sıfır ve standart sapması sabit olan bir normal dağılım

gösterdiği kabul edilir (16). Yapılan testlerin güvenilebilir olması için bu varsayımın yerine gelmiş olması gerekmektedir. Bu nedenle hataların normal dağılıp dağılmaması durumu büyük önem arz eder.

Hataların normal dağılıma uygunluğunun değerlendirilmesi için en sık kullanılan yöntemler aşağıda verilmiştir:

1. Q-Q nokta grafik yöntemi
2. Ki-kare uygunluk testi
3. Kolmogorov-Smirnov testi
4. Shapiro-Wilk testi
5. Anderson-Darling testi(16).

Ayrıca kullandığımız paket programlarda örneğin SPSS gibi programlarda normalliği kolay şekillerde görmemizi sağlayacak grafiksel metotlarda vardır. Hataların normal dağılmamasına ilişkin bazı sebepler vardır. Örneğin, aşırı ve etkili değerlerin veri setinde olması ya da unutulan önemli bir değişkenin veri setinde yer almaması gibi nedenler hataların normal dağılmamasına neden olabilmektedir. Bu gibi durumları da ortadan kaldırarak hataların normal dağılması sağlanabilir.

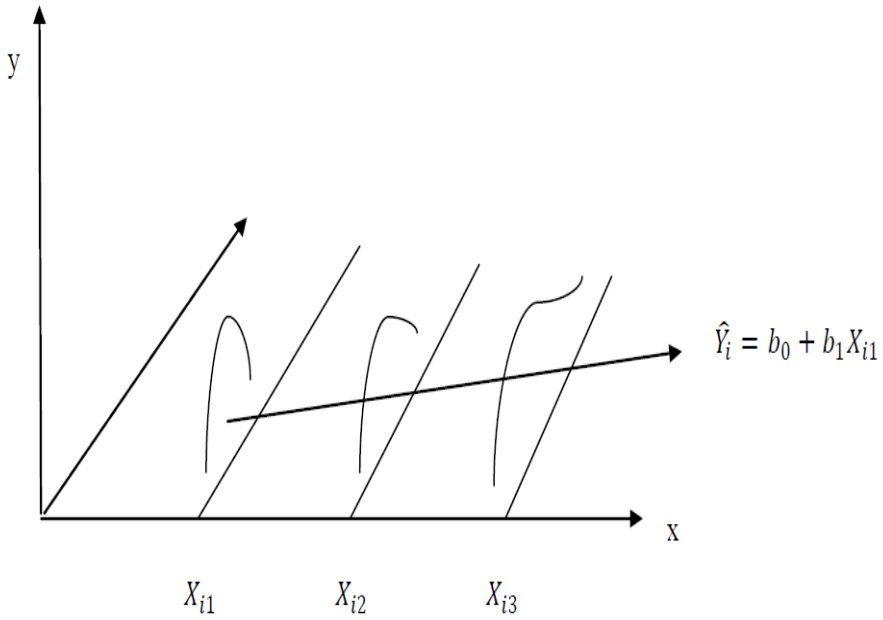
Ancak, unutulmaması gereken bir nokta ise eğer eşit varyanslılık ve hataların birbirinden bağımsız olması koşulları sağlanıyorsa hataların normal dağılamaması durumu büyük bir sorun oluşturmaz (16).

2.3.1.3. Hata Terimlerinin Varyansının Sabit Olması

Bu varsayım ile hata terimleri, X in tüm değerleri için kendi ortalamaları etrafında aynı dağılımı gösterir. Bu sonuç ise doğrusal regresyon modelinde elde edilen kestirimlerin standart hatalarının küçük olmasını ve kestirimlerin daha isabetli olmasını sağlar (12).

Bu varsayıma “eş varyanslılık (homoscadasticity)” denir. Bu varsayımın bozulması ise “değişen varyanslılık (heteroscadasticity) olarak adlandırılır.

Değişen varyanslılık grafiksel olarak Şekil 2.1 ile örneklenmiştir (17).



Şekil 2.1: Değişen varyanslılık

Değişen varyanslılık sorunun varlığı kontrol edilmediği ve sorunun giderilmesi için gerekli önlemler alınmadığı taktirde bulunan regresyon katsayıları yansız olmasına rağmen büyük standart hatalara sahip olacaktır. Bu durum etkisiyle, parametrelere ilişkin güven aralıkları genişleyecek ve katsayılara ilişkin testlerin düşük duyarlılıkta olması durumu ortaya çıkacaktır (12).

Bu sorunun varlığının araştırılabilmesi için kullanılacak yöntemlerden bazıları Grafik Yöntemi, Glejser testi, Spearman'ın Sıra Korelasyon Testi, Goldfield Quandt Testi ve Breusch Pagon Testi'dir (17).

Bu testler yardımıyla değişen varyanslılık durumu olduğu belirlenirse, bu sorunun giderilmesi için değişkenler üzerinde bazı dönüşümler yapılabilir. Bu dönüşümlere varyans dengeleme dönüşümleri adı verilir. Bu dönüşümler bağımlı ve bağımsız değişkenlerde yapılabilir.

Bazı varyans dengeleme dönüşümleri aşağıdaki tablo ile verilmiştir (12).

Tablo 2.1: Varyans dengeleme dönüşümleri tablosu

Açıklama ve Y değişkeninin olasılık dağılışı	Dağılımın ortalaması açısından Y'nin varyansı	Dönüşüm	Artık durumu
Y'ler Poisson dağılışına uyan sayımlar ise	μ	$\sqrt{y}$	Sağa ya da sola megafon
Y'lerin Poisson dağılışına uyan sayımlar ve Y'ler sıfıra yakın ya da çok küçükse	μ	$\sqrt{y}+\sqrt{y+1}$ $\sqrt{y+0.5}$ $\sqrt{y+1}$	Sağa ya da sola megafon
Y'lerin dağılım genişliği çok büyük ve tüm Y_i'ler pozitif ise	μ^2	log(y)	Sağa ya da sola megafon
Yukarıdakine ek olarak Y_i'lerin bazıları sıfıra eşit ise	μ^2	log(y+1)	Sağa ya da sola megafon
Y'lerin sıfıra yakın olacak şekilde toplandığı ve pozitif olduğu durumlarda	μ^4	$\frac{1}{y}$	Sağa ya da sola megafon
Yukarıdakine ek olarak bazı Y_i'ler sıfır ise	μ^4	$\frac{1}{y+1}$	Sağa ya da sola megafon
Binom oranları için $0 \leq y_i \leq 1$	$\frac{\mu(\mu+1)}{n}$	$sin^{-1}(\sqrt{y})$	Elips biçimi

2.3.1.4. Hata Terimlerinin Bağımsız Olması (Otokorelasyon Olmaması)

Bu varsayım altında farklı iki gözlem değerine ait hata terimleri birbirinden bağımsız $i \neq j$ iken $kov(\varepsilon_i, \varepsilon_j) = 0$ olmalıdır.

Bu varsayımın sağlanmaması durumuna otokorelasyon adı verilir. Otokorelasyon sorunu birçok sebepten dolayı ortaya çıkabilir. Bu nedenler aşağıdaki gibi sıralanabilir:

1. Açıklayıcılığı yüksek önemli bir bağımsız değişkenin modelde bulunmuyor olması
2. Verideki gözlem sayısının yetersiz olması
3. Uygun olmayan bir modelin seçilmesi
4. Bağımsız değişkenlerin ilişkili olması

Otokorelasyon varlığının araştırılması için grafik yöntemi, Durbin-Watson ve Von-Neumann testleri kullanılabilecek yöntemlerdendir. Görsel olarak yorum yapmada kolaylık sağlamasına rağmen grafiklerle otokorelasyonun varlığına kesin karar vermek her zaman mümkün olmaz. Bu nedenle analitik testleri uygulamak daha kesin sonuçlar verecektir (18).

Kurulan regresyon modelinde otokorelasyon varlığı belirlendiği durumda bu sorunu ortadan kaldırmak için; modele farklı bir bağımsız değişken ilavesi yapılabilir, gözlem sayısı arttırılabilir, model yeniden tanımlanabilir veya model üzerinde uygun olan çeşitli dönüşümler yapılabilir (19).

Otokorelasyon (özilişki) varlığının regresyon analizine etkileri ise şunlardı (16):

1) EKK yöntemiyle bulunan regresyon katsayıları yansızlığı sağlar ancak standart hataları minimum değeri almaz
2) Örnek regresyon denklemi ile regresyon katsayılarının standart hataları beklenenden düşük çıkabilir.
3) Aralık tahmini ve istatistik testler bağımsızlık ve rastgelelik varsayımına dayandıkları için geçerliliklerini kaybeder.

2.3.1.5. Gözlem Sayısının Fazla Olma

Çoklu doğrusal regresyonda verideki gözlem sayısının yetersizliği başta çoklu bağlantı olmak üzere birçok soruna sebep olabilir. Bu nedenle n gözlem sayısını, p ise regresyon denkleminde yer alan parametre sayısını göstermek üzere n>p koşulu sağlanmalıdır. Genellenebilirlik için en az gözlem sayısının bağımsız değişken başına 5 olması gerekmekle beraber bu sayının 10'un üzerinde özellikle de 15 ile 20 arasında olması arzu edilir. Bu yaklaşımlar dışında bazı yaklaşımlar da vardır. Bunlar katsayıları test edebilmek için uygun gözlem sayısının en az $104 + p$ $(n \geq 104 + p)$ kadar olması önerilmektedir. Ek olarak korelasyon katsayısına yönelik hesaplamalar için önerilen bir kesim noktası ise $n \geq 50 + 8p$ ile verilmektedir (1).

2.3.1.6. Bağımlı Değişken ile Bağımsız Değişkenler Arasında Doğrusal İlişki Olması

Korelasyon katsayılarına dayanan çok değişkenli yöntemler; çoklu doğrusal regresyon analizi, yapısal eşitlik modeli, faktör analizi ve diskriminant analizinin varsayımlarından biri

de doğrusallığın var olması koşuludur. Doğrusal olmayan etkileşimlerde hesaplanacak doğrusal korelasyonlar gerçek ilişkiyi hep olduğundan daha düşük gösterecektir. Bağımlı ve bağımsız değişkenler arasında doğrusal ilişki sağlanmadığında, bağımlı ve bağımsız değişkenlere dönüşüm uygulayarak doğrusallık koşulu elde edilebilir (19).

2.3.1.7. Bağımsız Değişkenlerin İlişkili Olmaması

Bağımsız değişkenler arasında ilişki varsa buna çoklubağlantı sorunu veya değişkenlerin ilişkili olması denir. Bu sorunun var olması bazı sonuçlara yol açar. Tezimiz çoklubağlantı durumunda En Küçük Kareler ve Temel Bileşenler Regresyonu sonuçlarını karşılaştırmayı amaçladığı için çoklubağlantı kavramı ayrı bir bölüm olarak ayrıntılı bir şekilde incelenecektir.

2.3.2. Çoklu Regresyonda Hipotez Testleri

Çoklu regresyon denklemi elde edilme işleminden sonra, çeşitli hipotezler test edilebilir. Öncelikle varyans analizi yapılarak, bağımsız değişkenlerin bağımlı değişkeni açıklayıp açıklamadığı, başka bir anlatımla bağımlı değişkenle bağımsız değişkenler kümesi arasında doğrusal bir ilişki var olup olmadığı test edilir (12).

Bulunan kestirimlerin anlamlılığına karar verebilmek için t ve F testi gibi testler kullanılır. Bu testler regresyon katsayılarının ve ayrıca çoklu korelasyon katsayısının anlamlılığı için kullanılabilir.

Bu testlerin yanında, regresyon modelindeki değişkenler arasında var olan ilişkinin derecesini yani kestirimlerin anlam derecesini belirlemede çoklu korelasyon katsayısı kullanılabilir (20).

2.3.2.1. Regresyon Modelinin Anlamlılığı için *F* Testi

Regresyon analizi yapılırken, bağımlı değişken üzerinde birden çok bağımsız değişkenin etkisinin var olup olmadığını kestirebilmek amacıyla F testi kullanılabilir. F testi ile Y bağımlı değişkeninin bağımsız değişkenlerin hepsiyle doğrusal bir bağa sahip olup olmadığı test edilebilir. Ancak, bu test ile bağımlı değişken ile bağımsız değişkenler arasında ilişki olduğuna karar verilmesine rağmen ilişkiyi hangi değişkenlerin sağladığı hakkında bir karara varılamaz.

Modeldeki değişkenler arasındaki ilişkiyi test edecek ve ilişkinin anlamlı olup olmadığını ortaya koyacak hipotezler şu şekilde oluşturulur.

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

$$H_1: \beta_1 \neq \beta_2 \neq \cdots \neq \beta_k \neq 0$$

H_0 hipotezi katsayıların hepsinin sıfır olduğunu dolayısı ile bağımsız değişkenler tarafından bağımlı değişkenin açıklanamadığını ve bu sebeple kurulan modelin istatistiksel açıdan anlamlı olamayacağını belirtirken, H_1 hipotezi ise bu katsayılardan en az bir tanesinin 0 olmadığını söyleyerek modelin anlamlı olacağını belirtir (3).

F testinin formülü çoklu korelasyon katsayısı olan R nin karesi olarak tanımlanan açıklayıcılık katsayısı R^2 yardımıyla belirtilir ve

$$F = \frac{R^2}{1 - R^2} . \frac{n - k}{k - 1}$$

olarak verilir. Formülde yer alan n gözlem sayısı, k kestirimi yapılacak parametre sayısı ve (k-1) bağımsız değişken sayısını belirtir.

Belirlenen bir α güven düzeyi için F tablosundan (k-1) ve (n-1) serbestlik derecesindeki tablo değeri $F_{(k-1,n-1)}$ bulunur ve hesaplanan F_{hesap} istatistiği ile karşılaştırılır. $F_{hesap} < F_{(k-1,n-1)}$ ise yokluk hipotezi reddedilerek bağımlı değişkenin modeldeki bağımsız değişkenler tarafından açıklandığı, kurulan modelin istatistiksel açıdan anlamlı olduğu sonucuna varılır. Tam tersi durumda ise yokluk hipotezi kabul edilerek modelin anlamsız olduğu sonucuna varılır ki bu durumda veri kontrol edilerek yeni gözlem eklenerek, başka bağımsız değişkenler kullanarak ya da veriye bağımsız değişken eklemesi yapılarak tekrar model anlamlılığı kontrol edilebilir.

2.3.2.2. Regresyon Katsayılarının Anlamlılığı için t Testi

t testi regresyon modelinde yer alan bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin gösterimi olan β parametrelerinin her birinin tek tek test edilmesi amacıyla kullanılır. Katsayıları test ederken anakütle varyansı bilinmiyor ve gözlem sayısı $n < 30$ ise t testi kullanılırken, anakütlenin varyansı bilindiği ve gözlem sayısı $n > 30$ olduğu durumda Z testi kullanılır (17).

β_j gibi bir katsayının test edilmesi amacıyla kurulacak hipotez testi

$$H_0: \beta_j = 0$$

$$H_0: \beta_j \neq 0$$

şeklinde olur.

F testinde olduğu gibi elde edilen test istatistiği tablo değeri ile karşılaştırılarak katsayıların anlamlı veya anlamsız olduğu, bu katsayıya ait bağımsız değişkenin bağımlı değişkeni açıklayıp açıklamadığına karar verilir. Diğer bir deyişle, H_0 hipotezi kabul edilirse β_j katsayısına denk gelen X_j bağımsız değişkenin Y bağımlı değişkenini açıklamadığı ve modelden çıkarılması gerektiği söylenebilir.

2.3.2.3. Çoklu Korelasyon Katsayısının Anlamlılığının Test Edilmesi

Bağımsız değişken sayısının birden çok olduğu regresyon modelinde, bağımlı değişkene ait gözlenen değerler ile kestirilen değerler arasındaki Pearson korelasyon katsayısına çoklu korelasyon katsayısı adı verilir (21). Genellikle yorumunun daha kolay olmasından dolayı uygulamalarda çoklu korelasyon katsayısının karesi olarak bilinen açıklayıcılık katsayısı R^2 değeri daha çok tercih edilir ve hesaplanır. R^2 değeri bağımlı değişkenin yüzde kaçının modeldeki bağımsız değişkenler tarafından açıklandığını belirten bir değerdir. R^2 nin 0,80 ve üstü bir değer olması kabul edilebilirdir (20) ve bağımsız değişkenlerin bağımlı değişkeni açıklayıcılığının iyi olduğu söylenir. Bu değer, 0-1 arasında değişir ve 1 değerine ne kadar çok yakınlaşırsa bağımsız değişkenlerin bağımlı değişkeni açıklayıcılığı o kadar artar. Kurulan model ne kadar iyi olursa ρ ve R^2 değeri de o kadar büyük olur (22).

Regresyon katsayılarının istatistiksel açıdan anlamlı olup olmadığını belirlemek için korelasyon katsayılarının anlamlılığı da kontrol edilmelidir. Y bağımlı değişken ve X_i ler bağımsız değişkenler olmak üzere regresyon modelindeki değişkenlikler aşağıdaki şekilde tanımlanır.

Toplam değişkenlik: $\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2$

Regresyonla açıklanan değişkenlik: $\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y}_i)^2$

Regresyonla açıklanmayan değişkenlik (Hata) : $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$ (20).

Regresyondaki değişkenler için varyans analizi tablosu Tablo 2.2 de verilmiştir.

Tablo 2.2: Değişkenler için varyans analizi tablosu

Değişimin Kaynağı	Kareler toplamı	Serbestlik derecesi	Kare ortalama	F oranı
Regresyonla açıklanan	$\sum(\hat{Y}_i - \bar{Y})^2$	k-1	$\frac{\sum(\hat{Y}_i - \bar{Y})^2}{k-1}$	$\frac{\frac{\sum(\hat{Y}_i - \bar{Y})^2}{k-1}}{\frac{\sum(Y_i - \hat{Y}_i)^2}{n-k}} = \frac{\sigma^2\hat{Y}_i - \bar{Y}}{\sigma^2 Y_i - \hat{Y}_i}$
Hata	$\sum(Y_i - \hat{Y}_i)^2$	n-k	$\frac{\sum(Y_i - \hat{Y}_i)^2}{n-k}$	
Toplam	$\sum(Y_i - \bar{Y}_i)^2$	n-1		

Bu tablodaki F oranı yardımıyla korelasyonlar için kurulan hipotezler test edilir. Hipotezler ise şöyle kurulur.

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

Hesaplanan F değeri, belirlenen bir α anlamlılık düzeyinde $k-1$ ve $n-k$ serbestlik dereceli F tablo değeri ile karşılaştırılır. $F_{hesap} > F_{tablo}$ değeri ise H_0 hipotezi reddedilir, H_1 hipotezi kabul edilir ve bağımlı değişkenin bağımsız değişkenler tarafından açıklandığı ve modelin anlamlı olduğu sonucuna varılır.

2.4. Çoklu Doğrusal Bağlantı Problemi

Çoklu doğrusal regresyon analizinin varsayımlarından biri olan bağımsız değişkenlerin birbirleriyle ilişkisinin olmaması varsayımı yerine getirilmezse çoklu doğrusal bağlantı problemi ile karşılaşılır. Regresyon modelinde yer alan bağımsız değişkenlerin hiçbiri arasında herhangi bir ilişki yoksa, diğer bir anlatımla bu değişkenlerle elde edilebilecek tüm ikili basit korelasyon değerleri sıfır oluyor ise değişkenlerin dik yani ortogonal olduğu söylenir. Ancak, çoğu uygulamada bağımsız değişkenler arasında ilişkiye rastlanmaması çok az karşılaşılan bir durumdur. Regresyondaki değişkenler arasında küçük de olsa bir ilişkiden bahsedilebilir. Belirlenen ilişki doğrusal bir ilişki ise sonuçta çoklu doğrusal bağlantının varlığından söz edilir. Bağlantının doğrusal olarak elde edilmemesi durumunda çoklu doğrusal bağlantının varlığından bahsedilemez. Çünkü çoklu doğrusal bağlantı bağımsız

değişkenler arasında doğrusal bağlantılarla ilişkili olup doğrusal olmayan ilişkilerle ilgisi yoktur (3).

Çoklu bağlantı, $nx(p+1)$ boyutlu girdi matrisini göstermek üzere $x_1, x_2, \dots, x_p$ kolonlarının doğrusal bağımsızlığı açısından tanımlanabilir. $k \leq p$ olmak üzere $x_1, x_2, \dots, x_t$ bağımsız değişkenleri hepsi sıfıra eşit olmayan $t_1, t, \dots, t_p$ katsayılarıyla sırasıyla çarpıldığında

$$\sum_{i=1}^{k} t_i x_i = t_1 x_1 + t_2 x_2 + \cdots + t_k x_k = 0$$

oluyorsa $x_1, x_2, \dots, x_t$ bağımsız değişkenleri doğrusal bağımlı olur ve bu durumda tam çoklu bağlantıdan söz edilir. Yukarıda belirtilen denklemden de anlaşılacağı üzere herhangi bir *Xi* değişkeni diğer değişkenler türünden yazılabilir. Böylece $X'X$ matrisinin rankı k+1 değerinden den küçük olur ve $X'X$ matrisinin tersi $(X'X)^{-1}$ hesaplanamaz. Eğer,

$$\sum_{i=1}^{k} t_i x_i = t_1 x_1 + t_2 x_2 + \cdots + t_k x_k \cong 0$$

durumu varsa güçlü çoklu bağlantı vardır. Bu durumda $(X'X)^{-1}$ ifadesi hesaplanabilir ancak bu durumun regresyonla elde edilecek sonuçlar üzerinde bazı olumsuz etkileri ortaya çıkacaktır (12).

Bu olumsuzluklar şu şekilde sıralanabilir:

1. EKK yöntemiyle kestirilmek istenen parametrelerin kestirimleri gerçek sonuçlarından çok farklı olacaktır.
2. Yapılan kestirimlerde yansızlık korunacaktır ancak bulunan kestirim değerlerinin mutlak değerleri çok büyük olacaktır. Bu durum ise veride çok küçük değişiklikler yapıldığında kestirilen parametrelerin işaret değiştirmesine neden olacaktır.
3. Parametre kestirimlerinin karasız olduğu görülecektir. Kestirimlerin geçerliliğini test etmek için farklı örneklemler kullanılarak kestirimler yapıldığında çok farklı sonuçlar elde edilecektir.
4. Çoklu bağlantı durumunda EKK için kullanılan bilgisayar algoritmaları, model kestirimi yapılan parametreler için çok farklı kestirimler ve işaretler verebilir.

5. Modelin tümel anlamlılığı için kullanılan varyans çözümlemesi ile yapılan F testi anlamlı bulunurken modeldeki katsayıların anlamlılığının değerlendirildiği t testi sonuçları anlamsız bulunabilir (12).

2.4.1. Çoklu Bağlantının Kaynakları

Çoklu bağlantının ortaya çıkma nedeninin bilinmesi çözüm bulunması konusunda bazı ipuçları verebilir. Çoklu bağlantı aşağıda sayılacak olan nedenlerden bir veya bir kaçının birleşmesi sonucu olarak ortaya çıkabilir.

1) **Aşırı tanımlanmış model**: Veriyi oluşturan gözlem sayısının kullanılan parametre sayısından küçük olması (n<p) durumudur. Bu sebepten ortaya çıkan bir çoklu doğrusal bağlantı sorununu aşabilmek için önem derecesine göre bazı değişkenleri regresyon modelinden çıkartmak veya verideki gözlem sayısını artırmak çözüm olabilir (15).
2) **Örnekleme yöntemleri**: Veriyi toplama sürecinde; araştırıcının isteyerek veya istemeyerek bağımsız değişkenler uzayından bir alt uzayı örnekleme alması durumunda çoklu doğrusallık oluşabilir (14). Bunun nedeni, gerçekte modelin kendisinde çoklu doğrusal bağlantı olmamasına rağmen bağımsız değişkenlerin eksik veya sayıca yetersiz bir alt kümesinin alınmasından kaynaklı bir çoklu bağlantının ortaya çıkmasıdır.
3) **Model ve anakütle üzerindeki fiziksel kısıtlar**: Bu durum, anakütlede gerçekte var olan ilişkilerin örneklemde de ortaya çıkmasıdır. Kitledeki zorunluluklar daha çok bağımsız değişkenlerin kimyasal veya üretim proseslerinden ortaya çıkar. Örneğin bir kimyasal reaksiyonun gerçekleşmesi için belli içeriklerin sabit oranlarda olması vb. (23).

Bu üç nedene ek olarak, araştırmacının çalışmaya başlarken seçtiği bağımsız değişkenler de bazı durumlarda çoklu doğrusal bağlantıya neden olabilir. Örneğin, hamile bir kadının yaşı, gebelik sayısı ve doğum sayısı gibi değişkenler farklı değişkenlermiş gibi düşünülse de gerçekte her üçü de sonuçları bakımından birbirleriyle ilişki oluşturan değişkenlerdir. Çünkü kadının yaşı arttıkça gebelik sayısı da artacaktır ve buna bağlı olarak gebelik sayısı arttıkça da doğum sayısında artış gözlenecektir. Bu sebeple yapılacak bir çalışmada araştırıcı tarafından bu üç değişkeninde de farklı değişkenlermiş gibi regresyon modeline koyulması çoklu bağlantıya sebep olabilir (22).

2.4.2. Çoklu Bağlantının Etkileri

Kurulan çoklu doğrusal regresyon modelinde, çoklu bağlantı olmasının olumsuz etkileri alt başlıklar ile açıklanacaktır.

2.4.2.1. Çoklu Bağlantının EKK Yöntemiyle Elde Edilen Kestirimlere Etkileri

Veride çoklu bağlantının olması durumunda, regresyon katsayıları için elde edilecek EKK kestirimleri etkilenir. Bu etkilerin daha kolay ifade edilmesi için iki bağımsız değişkenden oluşan doğrusal bir regresyon modeli dikkate alınacaktır.

$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + e$

olmak üzere vektörel olarak

$$e'e = \sum_{i=1}^{n} {e_i}^2 = (Y - \hat{Y})'(Y - \hat{Y})$$

$$= (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

$$= Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

yazılabilir. $\hat{\beta}'$ ye göre türev alınıp sıfıra eşitlenirse

$$\frac{\partial \sum_{i=1}^{n} {e_i}^2}{\partial \widehat{\beta}'} = -2X'Y + 2X'X\hat{\beta} = 0$$

olup, bu eşitlikten En Küçük Kareler denklemi

$$X'X\hat{\beta} = X'Y$$

olarak elde edilir. Bu ifade, r_{12}; X_1 ve X_2 değişkenleri arasındaki korelasyonu, r_{1y} ve r_{2y} ise bağımsız değişkenler X_1 ve X_2 ile Y bağımlı değişkeni arasındaki korelasyonu göstermek üzere aşağıdaki şekilde

$$\begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} \begin{bmatrix} \widehat{\beta_1} \\ \widehat{\beta_2} \end{bmatrix} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}$$

gösterilebilir.

C; $X'X$ matrisinin tersini göstermek üzere

$$C = (X'X)^{-1} = \begin{bmatrix} \frac{1}{1-{r_{12}}^2} & \frac{-r_{12}}{1-{r_{12}}^2} \\ \frac{-r_{12}}{1-{r_{12}}^2} & \frac{1}{1-{r_{12}}^2} \end{bmatrix}$$

olarak elde edilir. Bu eşitlik yardımıyla, $\hat{\beta}_1$ ve $\hat{\beta}_2$ katsayılarının kestirimleri

$$\hat{\beta}_1 = \frac{r_{1y} - r_{12} r_{2y}}{1-{r_{12}}^2}$$

$$\hat{\beta}_2 = \frac{r_{2y} - r_{12} r_{2y}}{1-{r_{12}}^2}$$

şeklinde elde edilir. Bağımsız değişkenler X_1 ve X_2 arasında kuvvetli bir ilişki olduğunda bu iki değişken arasındaki korelasyon katsayısı $|r_{12}| \to 1$ olacaktır.

Bu durum, $var(\hat{\beta}_1) = C_{11} = \frac{1}{1-{r_{12}}^2} = \infty$ ve $cov(\hat{\beta}_1, \hat{\beta}_2) = C_{12} \to \pm\infty$

olmasına sebep olacaktır (14).

Diğer bir anlatımla, bağımsız değişkenler X_1 ve X_2 arasındaki kuvvetli bir ilişki, EKK yöntemi ile kestirilen katsayıların olması gerekenden büyük varyans ve aynı şekilde büyük kovaryanslara sahip olmasına sebep olacaktır. Büyük varyanslılık durumunun, her bir örnek verisinde regresyon katsayıları için yapılan kestirimlerde hassasiyet açısından önemli bir gösterge olmasından yola çıkarak, farklı örnekler kullanıldığında oldukça farklı katsayı kestirimleri ortaya çıkacaktır (19).

Benzer şekilde ikiden fazla bağımsız değişkenden oluşan modellerde de çoklu bağlantının varlığı aynı sonuçlara yol açacaktır; katsayılara ait kestirimlerin varyansları büyüyecek ve katsayılar için elde edilecek güven aralıkları da genişleyecektir.

Çoklu doğrusal bağlantı, regresyon katsayılarının kestirimlerinin işaretlerini de etkilenir. Böyle bir durumda katsayı kestirimlerinin işaretlerine bakılarak bağımsız değişkenlerle bağımlı değişken arasındaki ilişki yanlış gösterilmiş olacaktır. Örneğin, araştırıcının öngörüsüne göre pozitif bir değer almasını beklediği parametre kestiriminin işareti tam tersine negatif, negatif çıkmasını beklediği bir parametrenin kestiriminin işareti pozitif olarak elde edilebilir. Bu gibi durumlarda çoklu bağlantının varlığı, bağımlı değişken kestirimlerinin de yanlış olmasına neden olur.

2.4.2.2. Çoklu Bağlantının Bağımlı Değişkenin Kestirimlerine Olan Etkileri

Çoklu bağlantının en küçük kareler kestirimlerine etkilerinde ifade edildiği üzere, regresyon katsayıları değer olarak ve işaret bakımından etkilediği için bu denklem yardımıyla yapılacak kestirimler de etkilenir. Regresyon katsayılarının gerçek katsayılardan değerce ve işaretçe çok farklı olması bağımlı değişken Y kestirim değerlerini etkiler, Y kestirimlerinin de standart hataları büyür (13).

2.4.2.3. Çoklu Bağlantının Hipotez Testlerine Olan Etkileri

Regresyon katsayılarının geçerliliğini test etmek amacıyla daha önce ifade edildiği üzere anakütle varyansı bilinmiyor ve gözlem sayısı $n < 30$ ise t testi kullanılırken, anakütlenin varyansı bilindiği ve gözlem sayısı $n > 30$ olduğu durumda Z testi kullanılır. Çoklu bağlantı olması durumunda bu iki test istatistiğinin değeri 0 a yaklaşır (15). H_0 hipotezinin reddedilmesi gittikçe zorlaşarak test edilen parametrenin sıfırdan farklı olmadığı yani ilgili bağımsız değişkenin bağımlı değişken Y yi etkilemediği sonucuna varılır. Böylece çoklu bağlantı durumu test istatistiklerinin değerlerinin küçük olarak elde edilmesine ve sonuçların yanlış olmasına sebep olur.

2.4.3. Çoklu Bağlantının Belirlenmesi

Bir regresyon analizinde ilk adımlardan birisi, veride çoklu bağlantı olup olmadığının belirlenmesidir (24). Çoklu bağlantının varlığını işaret eden bazı göstergeler vardır. Bunlar içinde en basit olanı, iki bağımsız değişken arasındaki basit korelasyon katsayısının 1 değerine (teorik olarak 0.80 ve üstü olması) yaklaşmasıdır. Ancak, bu durum kesin olarak çoklu bağlantının varlığını kanıtlamaz.

Çoklu bağlantının etkilerinde ifade edildiği üzere bulunan regresyon katsayılarının değerce büyüklüğü ve beklenenin aksine işarete sahip olması da bazen bir çoklu bağlantı durumu göstergesidir. Ayrıca, regresyon modeli anlamlı iken katsayıların anlamlılığı için yapılan testlerde regresyon katsayılarının istatistiki olarak anlamlı olmaması ve kestirimleri elde edilen regresyon katsayılarının güven aralıklarının genişlemesi de çoklu bağlantı sonucunda ortaya çıkabilmektedir.

Çoklu bağlantı durumunun belirlenmesine ek olarak bağlantının derecesinin de belirlenmesi anlamlı olacaktır. Bu amaçla kullanılan bazı yöntemler aşağıda verilmiştir (12).

2.4.3.1. Çoklu Bağlantının X'X Korelasyon Matrisiyle Belirlenmesi

Çoklu doğrusal bağlantı durumunun belirlenmesinde kullanılan ve uygulanması en kolay yöntemlerden biri olan bu yöntemde

$$X_{ij} = \frac{X_{ji} - \bar{X}_j}{\sum_{i=1}^{p}(X_{ji} - \bar{X}_j)^2}$$

şeklinde standartlaştırılarak elde $X'X$ standartlaştırılmış korelasyon matrisinde köşegenin dışında yer alan r_{ij} değerleri kontrol edilir. Farrar ve Glauber 1967 yılında r_{ij} değerlerini geometriksel olarak X_i ve X_j bağımsız değişkenleri arasındaki açının kosinüs değeri olarak tanımlamıştır (25). X_i ve X_j bağımsız değişkenleri arasında doğrusal bir bağıntı olduğunda $|r_{ij}|$ değerinin 1 e yaklaşması ilgili değişkenler arasında doğrusal bir ilişkiye çok yakın bir ilişki olduğunu, çoklu bağlantı durumunun olabileceğini belirtir.

Ancak, iki bağımsız değişken arasında var olan kısmi korelasyon değerinin büyük değerler almıyor olması, her zaman çoklu doğrusal bağlantı sorununun var olmadığı anlamına gelmez. Benzer şekilde, istatistiksel olarak anlamlı her korelasyon değeri de her zaman çoklu doğrusal bağlantı problemini gündeme getirmez. Lawrence Klein' e göre basit korelasyon katsayısı olarak verilen r, çoklu korelasyon katsayısından değerce küçük olursa çoklu bağlantı problemi ortaya çıkmayabilir (26).

2.4.3.2. Çoklu Bağlantının Açıklayıcılık Katsayısı ile İncelenmesi

Bu yöntemde amaç mevcut modele yeni bağımsız değişkenler ilave ederek R^2deki değişimlerin gözlemlenmesidir. R^2 de önemli bir gelişme olmazsa model için çoklu bağlantı durumundan söz edilebilir (19).

2.4.3.3. Çoklu Bağlantının Kısmi Korelasyon Katsayıları ile İncelenmesi

Bağımsız iki değişken arasındaki basit korelasyon katsayısı anlamlı iken kısmi korelasyon katsayılarının anlamsız olması çoklu bağlantının bir göstergesi olabilir. Ancak, bu yöntem de her zaman sağlıklı sonuçlar vermeyebilir. Diğer bir deyişle, kısmi korelasyon katsayılarının yüksek çıkması durumu dahi çoklu bağlantı problemini ortaya çıkarabilir (26).

2.4.3.4. Çoklu Bağlantının Tolerans Değerleri İle Belirlenmesi

Çoklu bağlantının varlığının gösterilmesinde kullanılabilen başka bir ölçü ise tolerans değerleridir. Bağımsız değişkenler arasındaki çoklu açıklayıcılık katsayısı ${R_j}^2$ olmak üzere tolerans değeri

$$T = 1 - {R_j}^2 \quad j = 1,2,\dots,k$$

ile hesaplanır. Bir değişkenin tolerans değerinin 0 a yaklaşması, bu değişkenin diğer bağımsız değişkenlerle arasında çoklu bağlantı olduğunun göstergesidir (12).

2.4.3.5. Çoklu Bağlantının VIF ile Belirlenmesi

Bağımsız değişkenlere ait korelasyon matrisinin tersinin köşegen üzerinde bulunan elemanlarına varyans sişme faktörleri (Variance Inflation Factors=VIF) denir. Terimsel olarak ifade edilecek olursa $(X'X)^{-1}$ matrisinde yer alan j-inci köşegen elemanına, j-inci varyans şişme faktörü adı verilir ve $(VIF)_j$ ile gösterilir (27). Tarihte ilk kez 1967 yılında Farrar ve Glauber tarafından çoklu bağlantının varlığının araştırılması için kullanılmıştır, ancak 1970 yılında Marquardt tarafından VIF adı ile adlandırılmıştır (25).

$(VIF)_j$ değerleri tolerans değerlerine bağlı olarak aşağıdaki şekilde elde edilir.

$$(VIF)_j = {}^{1}/_{tolerans} = {}^{1}/_{1 - R_j{}^2}$$

1992 yılında Webster tarafından VIF değerinin yorumlanması için verilen kurala göre; $VIF \geq 10$ olması durumunda çoklu bağlantıdan söz edilebilir (19).

2.4.3.6. Çoklu Bağlantının X'X Matrisinin Özdeğerleri İle Belirlenmesi

Vinod ve Ullah 1981 yılında, çoklu bağlantıyı ayrıntılı olarak çalışan ilk araştırmacı olan Ragnar Frisch (1934) in çoklu bağlantı varlığını özdeğerlerle ilişkilendirerek açıkladığını söylemiştir. Ancak o zamanlarda bilgisayar programlarının yetersizliğinden dolayı $X'X$ matrisinin özdeğerlerinin sayısal analizi pek desteklenememiştir (25). İlk kez Vinod ve Ullah (1981) bulunan en büyük özdeğerin en küçük özdeğere bölümünün karekökünü koşul sayısı olarak adlandırmıştır, daha sonra 1982 yılında Montgomery ve Peck koşul sayısının tanımını en büyük özdeğerin en küçük özdeğere bölümü olarak yapmışlardır (21).

X'X matrisine ait minimum ve maksimum özdeğerler, sırasıyla λ_{min} ve λ_{max} olmak üzere;

$$K = \frac{\lambda_{max}}{\lambda_{min}}$$

olarak tanımlanan koşul sayısı çoklu bağlantı durumunu araştırırken kullanılan en yaygın yöntemlerden biridir. Gujarati'nin 1995 yılında koşul sayısı için verdiği genellemeye göre $K>30$ olması durumu genelde bir çoklu doğrusal bağlantı durumunun olduğunu gösterir. Ancak $K<100$ olduğu durum için bu pek önemli değildir. $100<K<1000$ olması çoklu

bağlantının güçlü olduğunu, K>1000 olması durumu ise çoklu bağlantı durumunun ciddi boyutlarda olduğunu gösterir (28, 29).

$$K_i = \frac{\lambda_{max}}{\lambda_i}$$

ifadesi koşul indeksi olarak bilinir. Fark edileceği üzere en büyük koşul indeki koşul sayısına eşit olur.

Koşul indekslerine bakılarak veride kaç tane doğrusal bağımlılığın olduğu söylenebilir. 1000 den büyük koşul indeksi sayısı, çoklu bağlantı sayısını göstermektedir (12).

Ayrıca korelasyon matrisine ilişkin özdeğerlerden bir yada daha fazlası 0 veya 0 a yakın bir değer alıyorsa doğrusal bağımlılık olduğunun göstergesidir. Örneğin, veride herhangi bir özdeğer 0 ise veride 1 tane çoklu bağlantı olduğu söylenebilir (12).

2.4.3.7. Çoklu Bağlantının Korelasyon Matrisinin Determinant Değeri ile Belirlenmesi

Korelasyon matrisinin determinantı 0 ile 1 arasında değişim gösterir. Eğer determinant değeri 1 e yakınsa bağımsız değişkenler birbirine diktir, çoklu bağlantı yoktur denir. Aynı şekilde determinantın değeri 0 a yakınsa güçlü bir çoklu bağlantının varlığından söz edilir. Fakat bu yöntemin uygulanması her ne kadar kolay olsa da bu yöntemle hangi değişkenlerin çoklu bağlantılı olduğu belirlenemez (12).

2.4.4. Çoklu Doğrusal Bağlantının Giderilmesi için Yapılabilecekler

Çoklu doğrusal regresyonun varsayımlarından biri olan bağımsız değişkenlerin ilişkili olmaması, diğer bir deyişle veride çoklu bağlantı olmaması koşulunun sağlanmaması daha önce ifade edildiği üzere bazı sonuçlar doğurmaktadır. Yukarıdaki bahsi geçen yöntemler ile çoklu bağlantı durumu belirlenmiş ise, regresyon modeli üzerinde oluşturduğu zararlı etkilerinden dolayı, bu durumu ortadan kaldırma ya da etkisini indirgeme yoluna gidilmelidir.

Çoklu bağlantı durumunun ortadan kaldırılması için önerilen birçok yöntem vardır. Temelde ilk önerilen başlangıçta veriyi oluşturacak değişkenlerin seçiminde dikkatli olunmasıdır (24). Ayrıca veriye yeni gözlemler eklenmesi, modelin yeniden oluşturulması ya da bazı yanlı kestirim yöntemlerinin kullanılması en yaygın olan yöntemlerdir. Her bir yöntemin kendine göre uygulama alanı ve sakıncalı yönleri de var olabilir. Örnek verecek olursak; oluşturulan örneklemin seçildiği evreni çok iyi temsil etmemesi sebebiyle ortaya çıkacak bir çoklu bağlantı varlığında veriye uygun yeni gözlemlerin eklenmesi tavsiye edilir. Ancak örnekleme birim ilave etmek her zaman mümkün olmayabilir. Bir veya birden fazla

bağımsız değişkenin modelden atılması gerekir. Bu işlem modelin yeniden tanımlanması durumudur. Fakat hangi değişkenlerin çıkarılacağı bir sorun teşkil edebilir. Böyle bir yöntem modeli yanlış tanımlamamıza neden olabilir. Ek olarak bu yöntemin kullanımı örneklemin evreni temsil edemediği durumlar için uygun olmaz. Çünkü modele gerçekten katkısı olan bir değişken çoklu bağlantı durumundan dolayı modelden atılabilir.

Bağımsız değişkenler arasındaki gerçek ilişki nedeniyle ortaya çıkan çoklu bağlantı durumunda, sorunun bir örnekleme metodu sorunu olmadığı durumlarda, yapılabilecek bir işlem ise aralarında çoklu bağlantı olan değişkenlerin birleştirilerek yeni bir değişken oluşturulması ve modeli yeniden tanımlarken var olan çoklu bağlantılı değişkenler yerine bu değişkenlerle elde edilen yeni değişkenin kullanılmasıdır. Bu işleme modeli yeniden tanımlama denir (12, 19, 26).

Ancak tüm bunları yapmak yerine yani veriye ekleme, çıkarma işlemi yapmak yerine yanlı kestirim yöntemlerini kullanmak daha iyidir (13). Çoklu bağlantının varlığı durumunda oluşan sorunları ortadan kaldırmada kullanılacak en etkili yol modelde yer alan değişkenleri ekleyip çıkarmadan regresyon katsayılarını yanlı tahmin etmektir. Yanlı tahmin sonuçlarını kullanan yöntemlerden en çok tercih edilenleri Temel Bileşenler Regresyonu, Ridge Regresyon ve Kısmi En Küçük Kareler Regresyonudur. Temel Bileşenler Regresyonu veride var olan gerçek değişkenleri kullanmak yerine bunlara ait dik dönüşümlerini kullanır. Ridge regresyonda ise amaç korelasyon matrisinin köşegen elemanlarına küçük bir sayı eklenerek kestirim varyanslarının küçültülmesini sağlamaktır. Kısmi En Küçük Kareler Regresyonu ise bağımsız değişkenlerin yüksek oranda korelasyon içermesi ve gözlem sayısının değişken sayısından çok olduğu durumlarda kullanılır. Temel Bileşenler Regresyonu ve Çoklu Regresyon özelliklerini bir araya getirir. Amaç bağımsız ve bağımlı değişkenler arasındaki kovaryansı mümkün olduğunca küçültmektir (30). Bu sayılan yöntemlerin tümünde yanlılık kadar artış olurken, tahminlerin varyansları azalır.

3. MATERYAL VE METOT

Bu çalışmada, veride çoklu bağlantı olduğu durumda regresyon analizi için en küçük kareler regresyonu ve temel bileşenler regresyonu kullanılmıştır.

3.1. En Küçük Kareler Yöntemi

Doğrusal regresyon çözümlemesinde en yaygın kullanılan En Küçük Kareler yöntemi 1805 yılında Legendre tarafından geliştirmiştir. 1809 yılında Gauss geliştirdiği metot ile hatalar normal dağılımlı olduğunda en küçük kareler yönteminin en uygun çözüm olduğunu göstermiştir (31).

Bu yöntemin amacı, hata terimlerinin varyanslarının homojen olması durumunda ve normal dağılım göstermesi halinde optimum sonuçları, bir başka ifade ile hata terimlerinin kareleri toplamını en küçük hale getirerek modeli en iyi duruma getirmektir (32). Bunun için En Küçük Kareler (EKK) yöntemi aşağıda verilen fonksiyonu minimize edecek katsayı kestirimini yapmaya çalışır (15).

$$Q_{EKK}(b) = \sum_{i=1}^{n} {e_i}^2$$

En Küçük Kareler metodunda geçen regresyon artıkları e_i ler için şu varsayımlar geçerli olur.

1) e parametresi rastgele bir değişkendir.

2) Rastgele e değişkeninin beklenen değeri E(e_i) sıfırdır.

Bu varsayımın gerçekleşmemesi durumunda regresyon modeliyle yapılan parametre kestirimleri gerçek değerinden, e_i 'lerin her birinin pozitif olması durumunda daha büyük, negatif olması durumunda daha küçük olur. Yani parametre kestirimleri yanlı kestirimler olarak elde edilir (13).

3) Rastgele e değişkeninin varyansı Var(e_i) sabit σ^2 dir.

Bu varsayım ışığında *X*'in tüm değerleri için hata terimleri, her bir terim için kendi ortalamaları etrafında aynı dağılımı gösterirler. Bu sonuç doğrusal regresyon modelindeki

kestirimlerin standart hatalarının küçük olmasını, sonuç olarak kestirimlerin daha isabetli olmasını sağlar. Eğer bu varsayım göz ardı edilirse değişken varyanslılık sorunu oluşur. Bu durumda yine model için elde edilen regresyon katsayıları yansız olmasına rağmen katsayıların standart hataları büyük olacaktır. Bu ise parametrelere ilişkin güven aralıklarının büyümesine ve katsayılara ilişkin testlerin düşük duyarlılık göstermelerine neden olacaktır (26).

4) Rastgele e değişkeni $e_i \sim N(0, \sigma^2)$ ile normal dağılım gösterir.
5) Rastgele e değişkeninin farklı terimleri için aralarındaki korelasyon değerleri 0'dır. Yani $Kov(u_i, u_j) = 0$ dır.

Bu varsayım bozulduğu durumlarda karşımıza otokorelasyon sorunu çıkmaktadır. Bu sorunun regresyon analizi sonuçlarına çeşitli etkileri vardır. Örneğin, istatistik testler ile aralık kestirimi bağımsızlık ve rastgelelik varsayımına dayanmaları sebebiyle geçerliliğini kaybederler, regresyon denkleminin standart hatası ve regresyon katsayılarının standart hataları bulunması gereken değerlerden oldukça düşük çıkabilir (16).

6) Rastgele e değişkeni bağımsız değişkenlerden bağımsızdır. Yani $Kov(u_i, X_i) = 0$ dır.

Belirtilen varsayımların sağlanamadığı durumlarda yapılan kestirimler yanlı olmakta ve böylece ilgili önemlilik testleri geçerliliğini yitirmektedir (1). Yukarıdaki koşullara göre artıkların kareleri en küçük olacak şekilde b_0 ve b_1 katsayılarını kestirerek modeli en iyi duruma getirmeye çalışan yönteme En Küçük Kareler yöntemi denir (33).

Regresyon modelinin belirlenmesinde saçılım grafiği incelendiğinde doğrusal bir eğilim görülüyorsa X'in Y'ye göre matematik modelinin doğrusal olduğuna kesin olmamakla beraber karar verilebilir. Ancak, gözlem noktaları için aralarından birden çok doğru geçtiği gözlemlenebilir. Bu doğrulardan model için en uygunu, tüm doğrusal fonksiyonlar içinde Y gözlem değerine en yakın Y kestirim değerini minimum hata ile veren doğrusal fonksiyon olacaktır:

$$\varepsilon = Y - \check{Y} = Y - \beta_0 - \beta_1 X$$

Bu denklemi sonucunu minimum yapacak bir doğrunun seçilmesi gerekir. Tüm gözlem değerleri için bu durum geçerliği olduğundan

$$\sum_{i=}^{n} \varepsilon_i{}^2 = \sum_{i}^{n} \left(Y_i - \breve{Y}_\iota\right)^2 = \sum_{i}^{n} \left(Y_i - \beta_0 - \beta_1 X_i\right)^2$$

ifadesi minimum olmalıdır.

İkinci dereceden bir fonksiyonun değerinin minimum olması için fonksiyonun birinci türevlerinin sıfıra eşit olması gerekmektedir. Bu nedenle, yukarıdaki ifadeyi minimum yapmak için ifadenin ayrı ayrı b_0 ve b_1 e göre türevleri alınarak ifadeleri sıfıra eşitlememiz gerekmektedir. Türevler alınıp gerekli sadeleştirme işlemleri yapıldığında aşağıdaki denklemler elde edilir.

$$\sum Y = n\beta_0 + \beta_1 X$$

$$\sum YX = \beta_0 \sum X + \beta_1 \sum X^2$$

Bu denklem sistemlerindeki β_0 ve β_1 katsayıları bilinmeyen değerler olmak üzere diğer değerler bilindiği için denklem sisteminde yerlerine konularak sistemin çözümü elde edilir. Buradan bilinmeyenler β_0 ve β_1 parametreleri denklemden çekilirse

$$\beta_0 = \frac{\sum X^2 \sum Y - \sum X \sum XY}{n\sum X^2 - (\sum X)^2}$$

$$\beta_1 = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$$

olarak elde edilir. Bulunan β_0 ve β_1 parametreleri regresyon denkleminde yerine yazılarak regresyon kestirimi elde edilmiş olur (16).

3.2. Temel Bileşenler Regresyonu

Çoklu bağlantı durumunu gidermek için kullanılabilen Temel Bileşenler Regresyonunda, veri setinde yer alan orijinal değişkenler yerine bunların dik dönüşümleri yardımıyla elde edilen yeni yani yapay değişkenlerin bir kümesi üzerinde, En Küçük Kareler regresyon yöntemi uygulanarak regresyon katsayılarının yani $\hat{\beta}$ ların tahmin işlemi yapılır. Temel Bileşenler Regresyonu kullanılarak yapılan tahminde hata kareleri ortalaması değerinin

En Küçük Kareler regresyonu yöntemiyle yapılan tahmine değerine göre daha küçük bir değer alması beklenmektedir. Regresyon modelini matris gösterimi ile

$$Y = \beta X + \varepsilon$$

olarak verildiğinde

En Küçük Kareler regresyonu yöntemi ile tahmin edilen $\hat{\beta}$ katsayılarının değeri

$$\hat{\beta} = (X'X)^{-1}X'Y$$

şeklindedir. Bağımsız değişkenlerin, temel bileşenlere dönüşümü aşağıdaki bağıntı yardımıyla gerçekleştirilir.

$$X'X = PDP' = W'W$$

Bağıntıdaki;

D; $X'X$ nün özdeğerlerinin köşegen matrisini

P; $X'X$ özvektör matrisini belirtmektedir.

Böylece temel bileşenlerden oluşan veri seti üzerinde uygulanan En Küçük Kareler tahmini;

$$\gamma = (W'W)^{-1}W'Y$$

şeklinde olur. Burada γ ile gösterilen terim Temel Bileşenler Regresyonu yapılarak tahmin edilen regresyon katsayılarıdır. Yapılan işlemler sonucunda modelde çoklu doğrusal bağıntı problemi kalmaz.

Temel Bileşenler Regresyonu ile elde edilen γ katsayıları ile ilk veriye uygulanan En Küçük Kareler metodu ile elde edilen $\hat{\beta}$ katsayıları arasında

$$\gamma = P'\hat{\beta} \text{ ve } \hat{\beta} = P\gamma$$

İlişkileri vardır (34-36).

3.2.1. Temel Bileşenlerin Elde Edilmesi

Veride çoklu bağlantı olduğu durumda kullanılan yanlı kestirim yöntemlerinden birisi Temel Bileşenler Regresyonudur. İlk kez 1933 yılında Hotelling tarafından ele alınmıştır. Bu yöntemde amaç, korelasyon matrisinden elde edilen ve temel bileşenler olarak adlandırılan işlemle elde edilen yapay değişkenlerin bir kümesi üzerinde En Küçük Kareler yönteminin uygulanmasıdır.

Temel bileşenleri elde etme aşamasında ham X_{pxn} veri matrisi olduğu gibi kullanılabilirken, bunun yerine verinin standartlaştırılmasıyla elde edilen Z_{pxn} veri matrisi de kullanılabilmektedir. Temel bileşenlerin bulunmasında işlem yapılmamış ham verinin matrisi kullanılırsa varyans-kovaryans matrisinden, standartlaştırılmış veri matrisinin kullanılması durumunda ise korelasyon matrisinden yardım alınır. Farklı sonuçları ortaya koyabilen bu iki yöntemden hangisinin uygulanacağı konusunda en önemli kriter, değişkenleri oluşturan ölçü birimleridir. Şayet tüm değişkenler için ölçü birimleri ile varyanslar birbirine benzerse kovaryans matrisinden, değilse korelasyon matrisinden faydalanılır. Ancak, değişkenlerin ölçü birimlerinin birbirine benzer olması durumuna pratikte pek rastlanmaz.

Bağımlı değişken Y, bağımsız değişkenler kümesi X, regresyon katsayıları kümesi β ve hata terimi e olsun.

X, pxn boyutlu veri matrisi olmak üzere p toplam bağımsız değişken sayısını n ise gözlem sayısını göstermektedir.

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ X_{p1} & X_{p2} & \cdots & X_{pn} \end{bmatrix}$$

Temel Bileşenler Regresyonunun ilk adımında hem bağımlı değişken hem de bağımsız değişkenler ortalamadan farkları alınarak standart sapmalarına bölünerek standartlaştırma işlemi yapılır.

i’ inci değişken için aritmetik ortalama ve standart sapma sırası ile

$$\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$$

ve

$$S_i = \sqrt{\sum_{j=1}^{n} \frac{(X_{ij} - \bar{X}_i)^2}{n}}$$

olarak elde edilir.

Böylece i’ inci değişkenin j’ inci gözlemine ait standartlaştırılmış değeri

$$Z_{ij} = \frac{(X_{ij} - \bar{X}_i)}{S_i}$$

olarak bulunur.

Bu işlemlerden sonra $i = 1, 2, \dots, p$ ve $j = 1, 2, \dots, n$ olmak üzere

$$Z = \begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1n} \\ Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ Z_{p1} & Z_{p2} & \cdots & Z_{pn} \end{bmatrix}$$

pxn boyutlu $Z = (Z_1, Z_2, \dots, Z_p)$ vektörlerinden oluşan standartlaştırılmış veri matrisi elde edilir. Bu yolla elde edilen standartlaştırılmış değişkenlerin aritmetik ortalaması 0 ve standart sapması 1 olur.

Z_i ile Z_k vektörleri arasında oluşan kovaryans değeri

$$Cov(Z_i, Z_k) = S_{ik} = \frac{1}{n-1}\sum_{j=1}^{n} \left(\left(Z_{ij} - \bar{Z}_i\right)\left(Z_{kj} - \bar{Z}_k\right)\right)$$

olarak gösterildiğinde i’ inci değişken ile k’ inci değişkenler arasındaki korelasyon katsayısı

$$r_{ik} = \frac{Cov(Z_i, Z_k)}{\sqrt{Var(Z_i)}\sqrt{Var(Z_k)}} = \frac{\sum_{j=1}^{n}\left(\left(Z_{ij} - \bar{Z}_i\right)\left(Z_{kj} - \bar{Z}_k\right)\right)}{\sqrt{\sum_{j=1}^{n}(Z_{ij} - \bar{Z}_i)^2}\sqrt{\sum_{j=1}^{n}(Z_{kj} - \bar{Z}_k)^2}}$$

ile elde edilir. Burada;

i=1,2,…,p

k=1,2,…,p

$r_{ik} = r_{ki}$

i≠k

R, pxp boyutlu bağımsız değişkenler için korelasyon matrisidir.

$$R = \frac{ZZ'}{n-1}$$

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix}$$

$\lambda_1, \lambda_2, \ldots, \lambda_p$ değerleri $\det(R - \Lambda I) = 0$ eşitliğini yerine getiren ve $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ koşuluna uygun özdeğerlerdir.

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}$$

dir.

$a = \left(a_1, a_2, \ldots, a_p\right)$ vektörleri ise sıfıra eşit olmayan, $Ra_i = \lambda_i a_i$ eşitliğinden $a_i' a_i = 1$ ile $a_i' a_j = 0$ $(i \neq j = 1,2, \ldots, p)$ koşullarının sağlanmasıyla elde edilen korelasyon matrisine ait standartlaştırılmış özvektör değerleridir.

$$a_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \cdots \\ \cdots \\ \cdots \\ a_{p1} \end{bmatrix} \quad a_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \cdots \\ \cdots \\ \cdots \\ a_{p2} \end{bmatrix} \quad \cdots \quad a_p = \begin{bmatrix} a_{1p} \\ a_{2p} \\ \cdots \\ \cdots \\ \cdots \\ a_{pp} \end{bmatrix}$$

a' ise a vektörüne ait devrik (transpoze) vektördür.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix}$$

Temel Bileşenler Regresyonu, birbirleriyle ilişkili olan Z_{ij} değerlerine bir dönüşüm uygulayarak birbirleriyle ilişkisi olmayan W_{ij} değerlerinin elde edilmesini sağlar. Bu dönüşüm aşağıdaki gibi gösterilebilir:

$$W_{pxn} = A_{pxp}{}' Z_{pxn}$$

$$W_1 = (a_1)'Z = a_{11}Z_1 + a_{21}Z_2 + \cdots + a_{p1}Z_p$$

$$W_2 = (a_2)'Z = a_{12}Z_1 + a_{22}Z_2 + \cdots + a_{p2}Z_p$$

...

...

$$W_p = \left(a_p\right)'Z = a_{1p}Z_1 + a_{2p}Z_2 + \cdots + a_{pp}Z_p$$

a_{ij} değerleri her bir temel bileşen değerinin hangi değişken ile ne gibi bir oranla ilişki içinde olduğunu gösteren özvektör değerleridir. Bu değerlere temel bileşen yükleri de denilir. Temel bileşen değerlerine ait varyans ve kovaryans değerleri;

$$Var(W_i) = Var((a_i)'Z) = (a_i)'Ra_i$$

$Cov(W_i, W_k) = (a_i)'Ra_k$

dir.

Buradaki $W_1 = (a_1)'Z$ dönüştürülmüş vektörüne birinci temel bileşen değeri, eşitlikteki a_1 vektörüne ise birinci özvektör değeri denir.

Temel bileşenler $W_1, W_2, \dots, W_p$ belirlenirken $Var(W_1) > Var(W_2) > \cdots > Var(W_p)$ koşulunu sağlayacak ve birbirinden bağımsız olacak bir şekilde seçilmelidir.

Böylece toplam varyansa en çok katkısı olan birinci temel bileşen;

$W_1 = (a_1)'Z = a_{11}Z_1 + a_{21}Z_2 + \cdots + a_{p1}Z_p$ doğrusal birleşim denklemidir.

$MaxVar(W_1) = (a_1)'Ra_1$ eşitliğinde a_1 vektörü birinci temel bileşenin varyansı maksimum bir değer olacak şekilde belirlenmektedir. Bu sebeple a_i vektörlerini $a_i'a_i = 1$ olacak şekilde seçmek uygun olur. Bu denkleme göre seçilen

1. Birinci temel bileşen $MaxVar(W_1) = (a_1)'Z$ ve $a_1'a_1 = 1$ şartını gerçekleyen $(a_1)'Z$ doğrusal bileşeni olur.

2. İkinci temel bileşen $MaxVar((a_2)'Z)$, $a_2'a_2 = 1$ ve $Cov(W_1, W_2) = 0$ koşullarını sağlayan $(a_2)'Z$ doğrusal bileşeni olur.

3. Genellenecek olursa i inci temel bileşen değeri $MaxVar((a_i)'Z)$, $a_i'a_i = 1$ ve $i > k$ olmak üzere $Cov(W_i, W_k) = 0$ koşullarını sağlayan $(a_i)'Z$ doğrusal bileşimi olur.

$W_i = (a_i)'Z$ i=1, 2, … , p olmak üzere

$\sum_{i=1}^{p} Var(W_i) = \sum_{i=1}^{p} Var(Z_i) = p$

Standartlaştırılmış değişkenler için toplam varyans p dir. Aynı zamanda standartlaştırılmış veri için korelasyon matrisinin (R) determinant değeri, elde edilen p tane

özdeğerin toplamına, diğer bir söylemle p değerine eşit olur. Gösterim olarak $|R| = \sum_{i=1}^{p} \lambda_i$ dir.

Bileşenler için toplam varyansı açıklama oranları verilecek olursa; birinci temel bileşen için toplam varyansı açıklama oranı;

$$\frac{\lambda_1}{p} = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}$$

Birinci ve ikinci temel bileşenin birlikte toplam varyansı açıklama oranı;

$$\frac{\lambda_1 + \lambda_2}{p} = \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}$$

ve benzer olarak toplam p tane temel bileşen için toplam varyansı açıklama oranı;

$$\frac{\lambda_1 + \lambda_2 + \cdots + \lambda_p}{p} = \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_p}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} = 1$$

olarak elde edilir.

Sonuç olarak temel bileşenler analizinde, veri matrisini standartlaşma yoluyla korelasyon matrisi elde edilir. Korelasyon matrisi yardımı ile özdeğerler, standartlaştırılmış özvektörler ve her bir özvektör için devrik vektör elde edilir. Son aşamada devrik özvektör standartlaştırılmış veri matrisi ile çarpılarak temel bileşen değerleri elde edilir (34-37).

3.2.2. Temel Bileşenlerin Özellikleri

1. $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$ dır.

2. Temel bileşenlerden herhangi birinin toplam varyansı açıklama yüzdesi $\left({}^{\lambda_i}\!/_{\sum_{i=!}^{p} \lambda_i} \right) x100$ ile belirlenir.

3. $W_i = a_i' Z$ (i=1, 2, … , p) eşitliği kullanılarak

a. $Var(W_i) = \frac{W_i W_{i'}}{n-1} = \frac{(a_i'Z)(a_i'Z)'}{n-1} = a_i' \left[\frac{ZZ'}{n-1}\right] a_i = a_i' R a_i = a_i{}' \lambda_i a_i =$

b. $\lambda_i a_i{}' a_i = \lambda_i$

elde edilir. Bu işlemler sonucunda özdeğerlerin temel bileşenlerin varyansları olduğu söylenir.

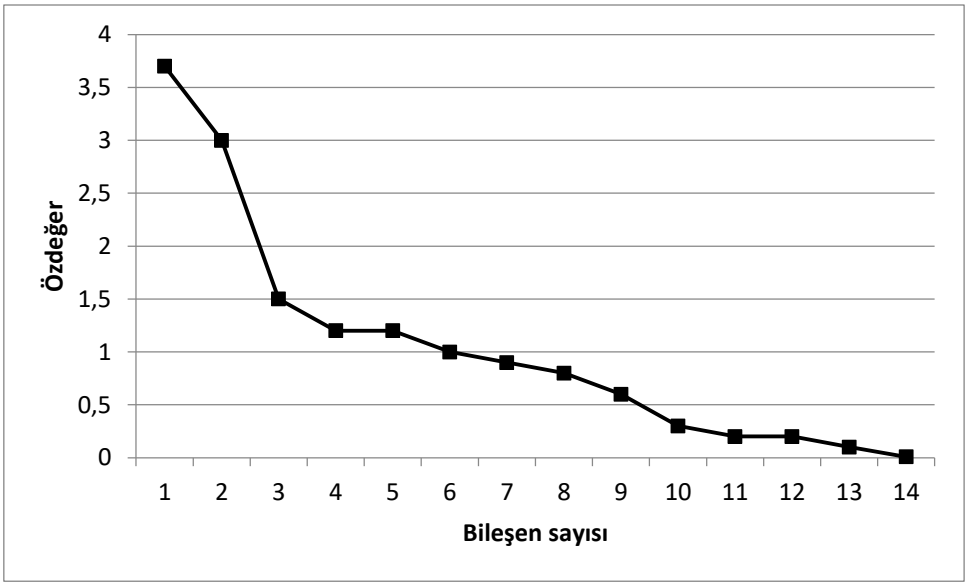
4. Orijinal değişkenlerin toplam varyans değeri, elde edilen temel bileşenlerin toplam varyansına eşit olur (37).

$$\sum_{i=1}^{p} Var(z_i) = \lambda_1 + \lambda_2 + \cdots + \lambda_p = \sum_{i=1}^{p} Var(y_i)$$

3.2.3. Temel Bileşen Sayısının Belirlenmesi

Veri matrisi R nin özdeğerlerinin elde edilmesinden sonra en önemli kısım q önemli özdeğer sayısının belirlenmesi aşamasıdır. Önemli özdeğer sayısı q'nun belirlenmesinde birkaç farklı yöntem vardır. Bunlar içinde en basiti ve en bilineni standartlaştırılmış veri matrisinin ele alındığı durumlarda 1 den büyük olan özdeğerlerin sayısının q olarak seçilmesidir. Bir diğer yaklaşımla, $\sum_{j=1}^{q} \lambda_j / p \geq {}^{2}/_{3}$ koşulunu sağlayan en küçük q değeri önemli bileşen sayısı olarak alınabilmektedir (36).

Temel bileşenlerin sayısının belirlenmesinde kullanılabilecek bir diğer yöntem ise grafik ile belirleme yöntemidir. Bileşen sayısına karşı özdeğerlerin grafiklenmesi ile, varyans açıklama oranlarındaki hızlı düşüş belirlenerek, temel bileşen sayısı belirlenebilmektedir.



Şekil 3.1: Özdeğerlerin Varyans Açıklama Oranları

Grafik yöntemi ile en hızlı düşüşün gözlendiği yerdeki bileşen sayısı temel bileşen sayısı olarak alınır (36).

3.3. Benzetim Çalışması

Bu çalışmanın amacı olan veri setinde çoklu bağlantı olduğu durumda En Küçük Kareler ve Temel Bileşenler Regresyon analizlerinin sonuçlarının karşılaştırılması için IBM SPSS Statistics sürüm 25.0 ile simülasyon tekniği kullanılarak veri türetilmiştir (30). İki farklı grubunda 10 ar adet veri seti bulunmaktadır. Türetilen tüm veri setleri 3 bağımsız (x1, x2 ve x3), 1 bağımlı (y) değişken içermektedir ve tüm değişkenler standart normal dağılımdan türetilmiştir.

İlk amaç, çoklu bağlantının derecesi değiştikçe EKK ve TBR analiz sonuçlarını karşılaştırmak olduğu için ilk veri grubunda farklı derecelerde çoklu bağlantı içerecek 10 adet veri seti türetilmiştir. Çoklu bağlantının derecesini değiştirmek amacıyla bağımlı değişken ile bağımsız değişkenler arasındaki korelasyonlar sabit tutulmuş, bağımsız değişkenler arasındaki korelasyon değerleri arttırılmıştır. Her veri seti 1000 gözlemden oluşmaktadır. Bu 10 veri seti türetilirken kullanılan korelasyon düzeyleri Tablo 3.1'de verilmiştir.

Tablo 3.1. Birinci veri grubundaki 10 veri setinin korelasyon yapısı

Değişkenler arasındaki korelasyon gösterimi	Değişkenler arasındaki korelasyonun değeri Veri setleri									
	1	2	3	4	5	6	7	8	9	10
x_1-x_2	0.75	0.80	0.80	0.85	0.90	0.90	0.95	0.95	0.95	0.95
x_1-x_3	0.90	0.90	0.9	0.90	0.90	0.90	0.90	0.90	0.90	0.90
x_2-x_3	0.50	0.50	0.55	0.50	0.50	0.55	0.50	0.55	0.6	0.65
y-x_1	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
y-x_2	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65
y-x_3	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55

Bu çalışmanın bir diğer amacı ise gözlem sayısının çoklu bağlantı üzerine etkisini incelemekti. Bu amaçla türetilen ikinci veri grubunda yine 10 veri seti bulunmaktadır. Bu veri grubunda tüm değişkenler arasındaki korelasyonlar sabit tutularak, veri setlerindeki gözlem sayısı arttırılmıştır. Gözlem sayısı 1000'den 10000'e kadar 1000'er gözlem arttırılarak 10 veri seti türetilmiştir. Böylece, çoklu bağlantı düzeyi aynı iken gözlem sayısının etkisi değerlendirilebilecektir. İkinci veri grubunda değişkenler arasındaki korelasyon yapısı Tablo 3.2 ile gösterilmiştir.

Tablo 3.2. İkinci veri grubundaki 10 veri seti için ortak korelasyon yapısı

Değişkenler	Değişkenler arasındaki korelasyonun değeri
x1-x2	0.80
x1-x3	0.90
x2-x3	0.55
y-x1	0.75
y-x2	0.65
y-x3	0.55

3.4. Veri Analizi

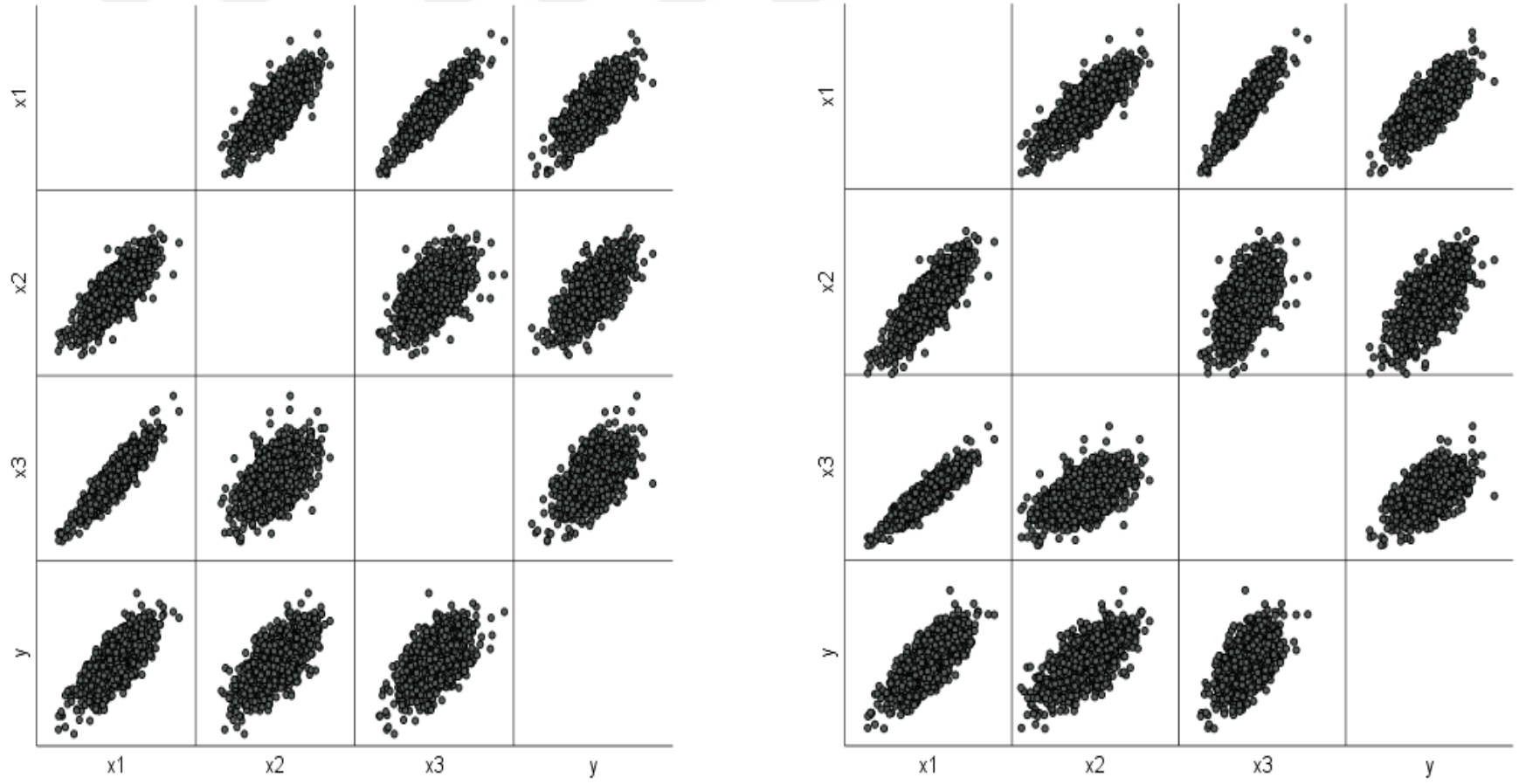
Türetilen verinin çoklu normal dağılıma uygunluğu, İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Anabilim Dalı tarafından geliştirilen "Normal Dağılım İnceleme Yazılımı" ile incelendi (38). Benzetim tekniği ile elde verilen setlerine IBM SPSS Statistics sürüm 25.0 paket programı ile öncelikle çoklu doğrusal regresyon analizi uygulanmıştır (30). Çözümleme sonucunda veri setinde çoklu bağlantının varlığı incelenmiştir. Çoklu bağlantının varlığı VIF, tolerans değerleri, özdeğerler ve koşul indeksine bakılarak desteklenmiştir. En küçük kareler regresyonu sonuçları ile karşılaştırılacak olan Temel bileşenler regresyonu için NCSS-2007 paket programı kullanılmıştır (39).

4.BULGULAR

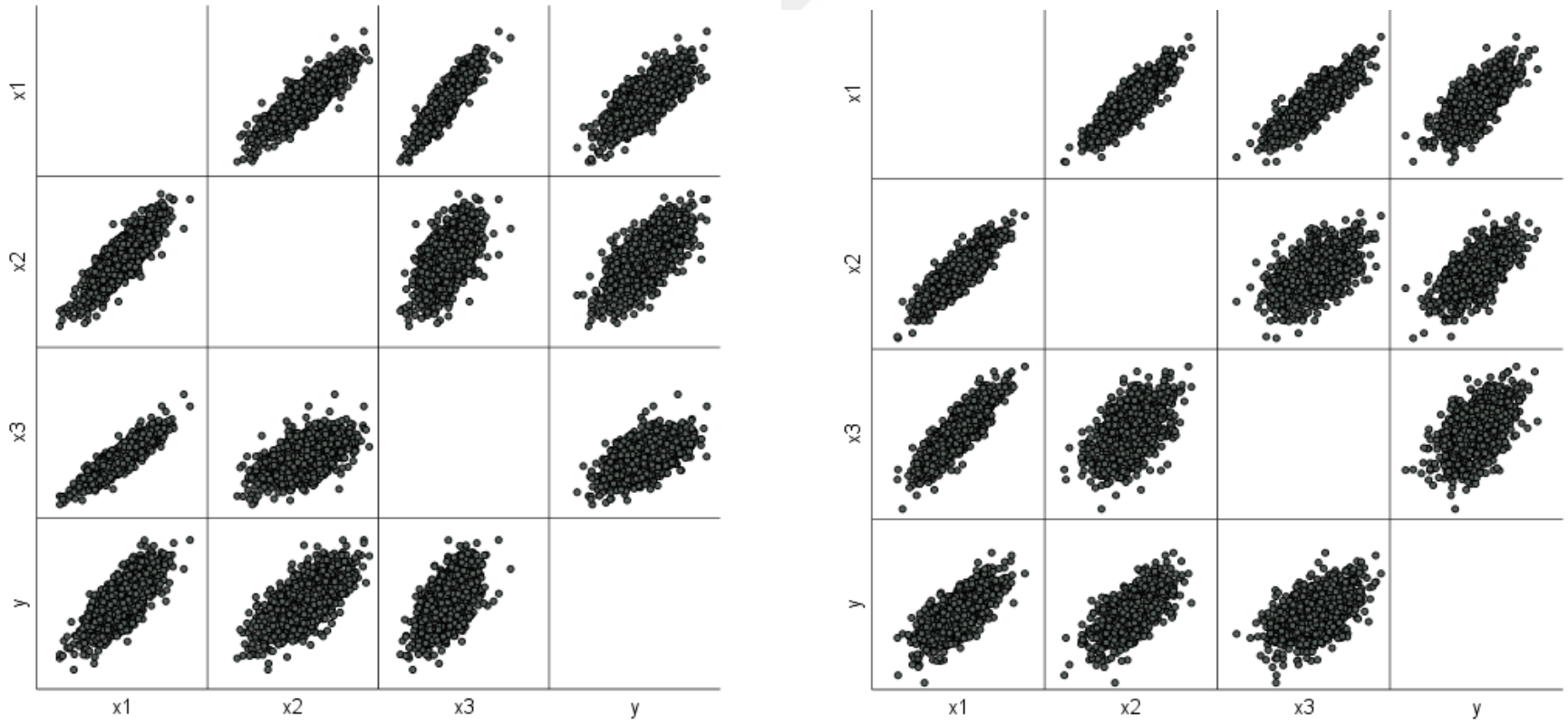
Çalışmada kullanılan ilk veri grubu olan çoklu bağlantı derecesi farklı 10 veri seti için matris şeklindeki saçılım grafikleri çoklu bağlantının varlığının belirlenmesi için kullanılan ölçülere ait bilgiler Tablo 4.1 ve matris şeklindeki saçılım grafikleri Şekil 4.1 – Şekil 4.5 ile sunulmuştur.

Tablo 4.1. Farklı derecelerde çoklu bağlantı içeren ilk veri grubu için çoklu bağlantı belirleme kriterleri tablosu

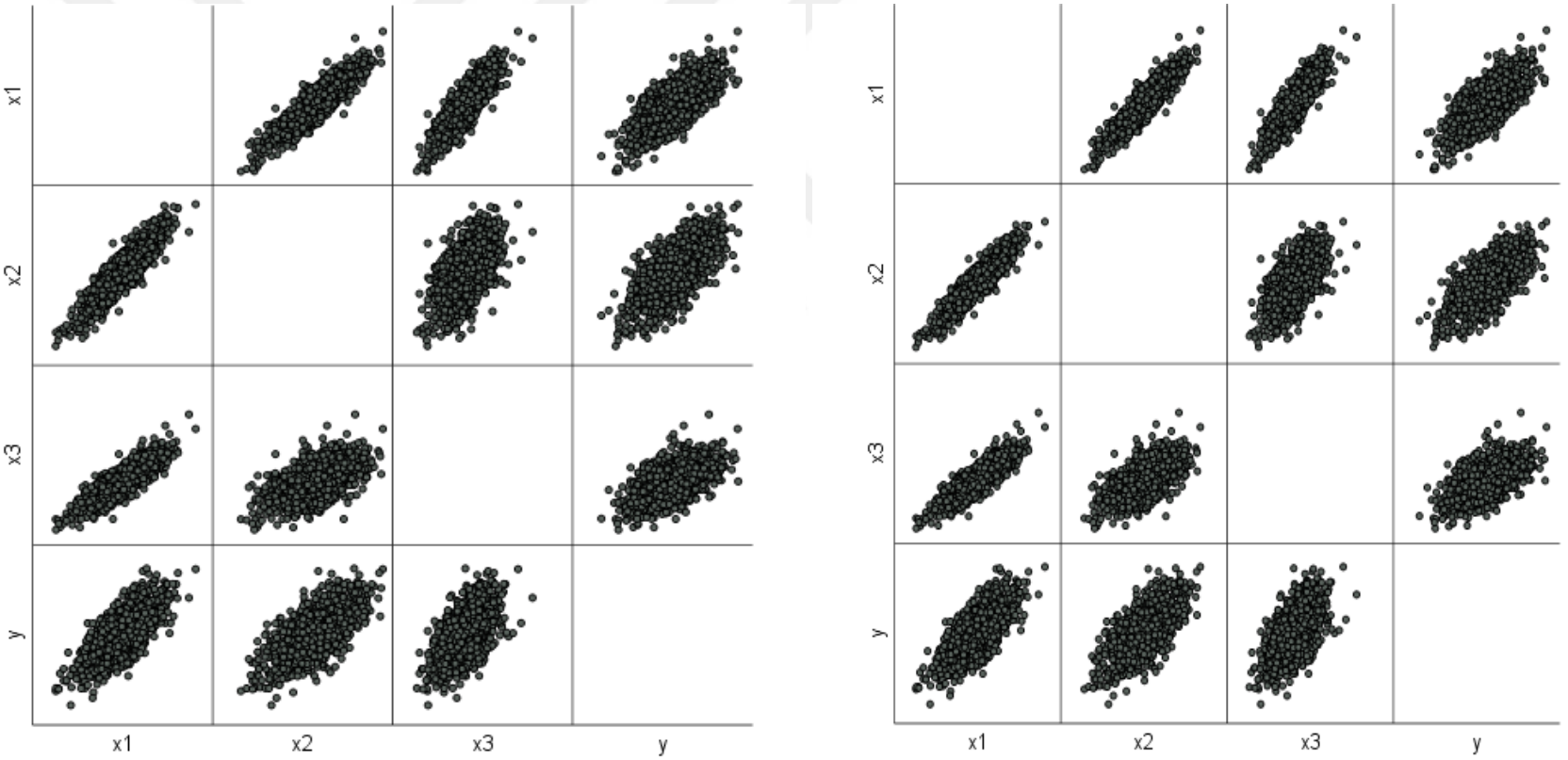
Veri Sayısı	Bağımsız Değişkenler	VIF	Tolerans Değeri	Özdeğer	Koşul İndeksi
n=1000	x_1	16.87850	0.05925	2.47096	1.00
	x_2	4.04280	0.24735	0.49355	5.01
	x_3	9.67960	0.10331	0.03550	69.61
n=1000	x_1	44.18220	0.02263	2.50335	1.00
	x_2	10.89020	0.09183	0.48303	5.18
	x_3	20.83430	0.04800	0.01362	183.84
n=1000	x_1	20.8475	0.04796	2.533271	1.00
	x_2	5.4184	0.18455	0.437455	5.79
	x_3	10.5746	0.09456	0.029274	86.54
n=1000	x_1	165.55410	0.00604	2.52092	1.00
	x_2	46.75720	0.02139	0.47543	5.30
	x_3	64.36160	0.01554	0.00365	691.17
n=1000	x_1	235.73770	0.00424	2.50118	1.00
	x_2	79.79830	0.01253	0.49627	5.04
	x_3	78.98630	0.01266	0.00255	980.73
n=1000	x_1	160.0686	0,006247	2.570841	1.00
	x_2	52.3712	0,019094	0.425333	6.04
	x_3	51.6331	0,019367	0.003827	671.85
n=1000	x_1	441.51720	0.00226	2.53002	1.00
	x_2	168.38760	0.00594	0.46861	5.40
	x_3	123.11860	0.00812	0.00137	1848.17
n=1000	x_1	249.43130	0.00401	2.57379	1.00
	x_2	94.99260	0.01053	0.42376	6.07
	x_3	66.44860	0.01505	0.00245	1050.42
n=1000	x_1	158.99470	0.00629	2.61782	1.00
	x_2	60.96010	0.01640	0.37830	6.92
	x_3	40.47790	0.02470	0.00389	673.85
n=1000	x_1	109.18950	0.00916	2.66186	1.00
	x_2	42.65640	0.02344	0.33243	8.01
	x_3	26.57870	0.03762	0.00571	465.93



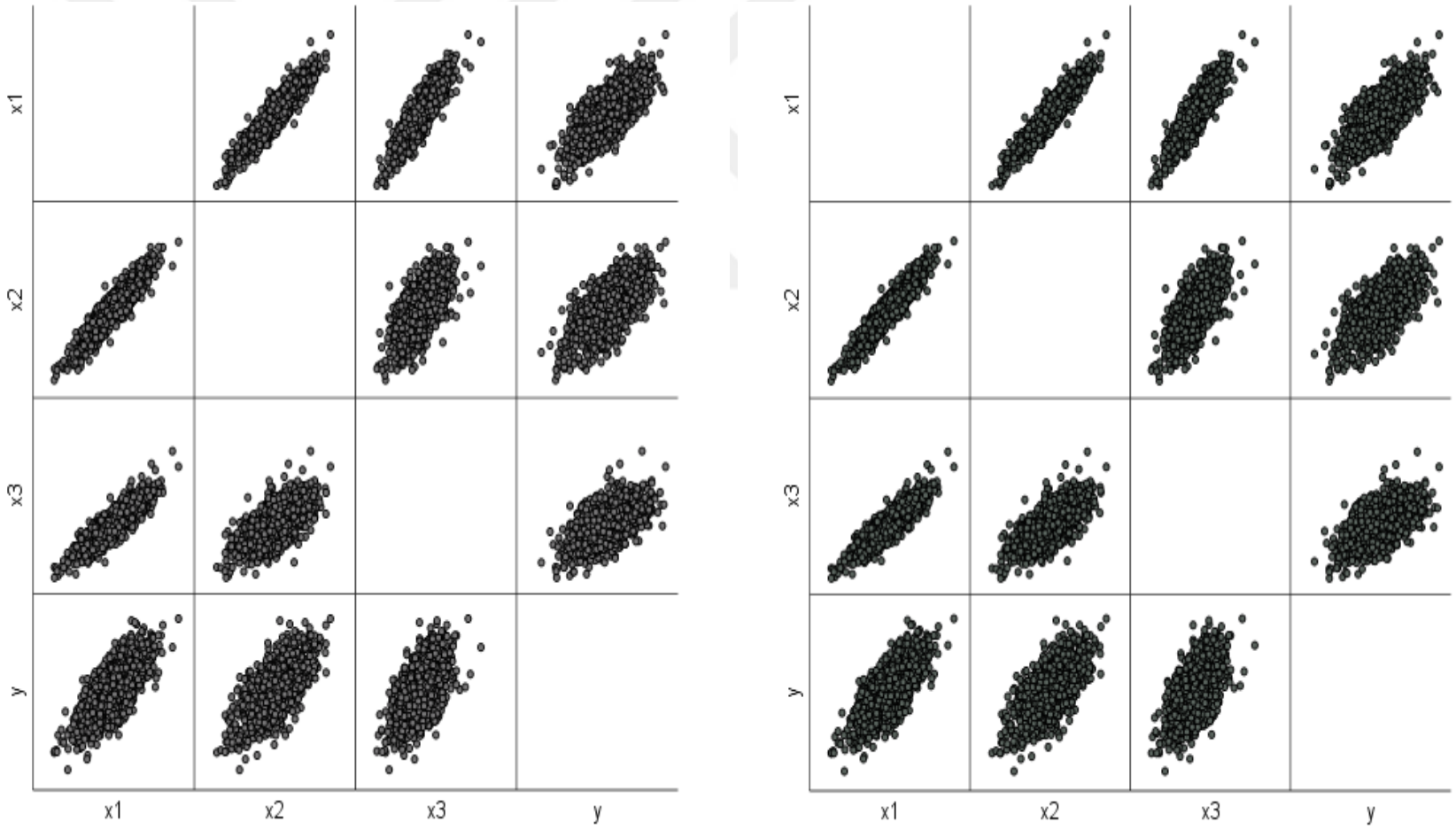
Şekil 4.1: Farklı derecelerde çoklubağlantı içeren veri seti grubu içindeki birinci ve ikinci veri seti için saçılım grafikleri



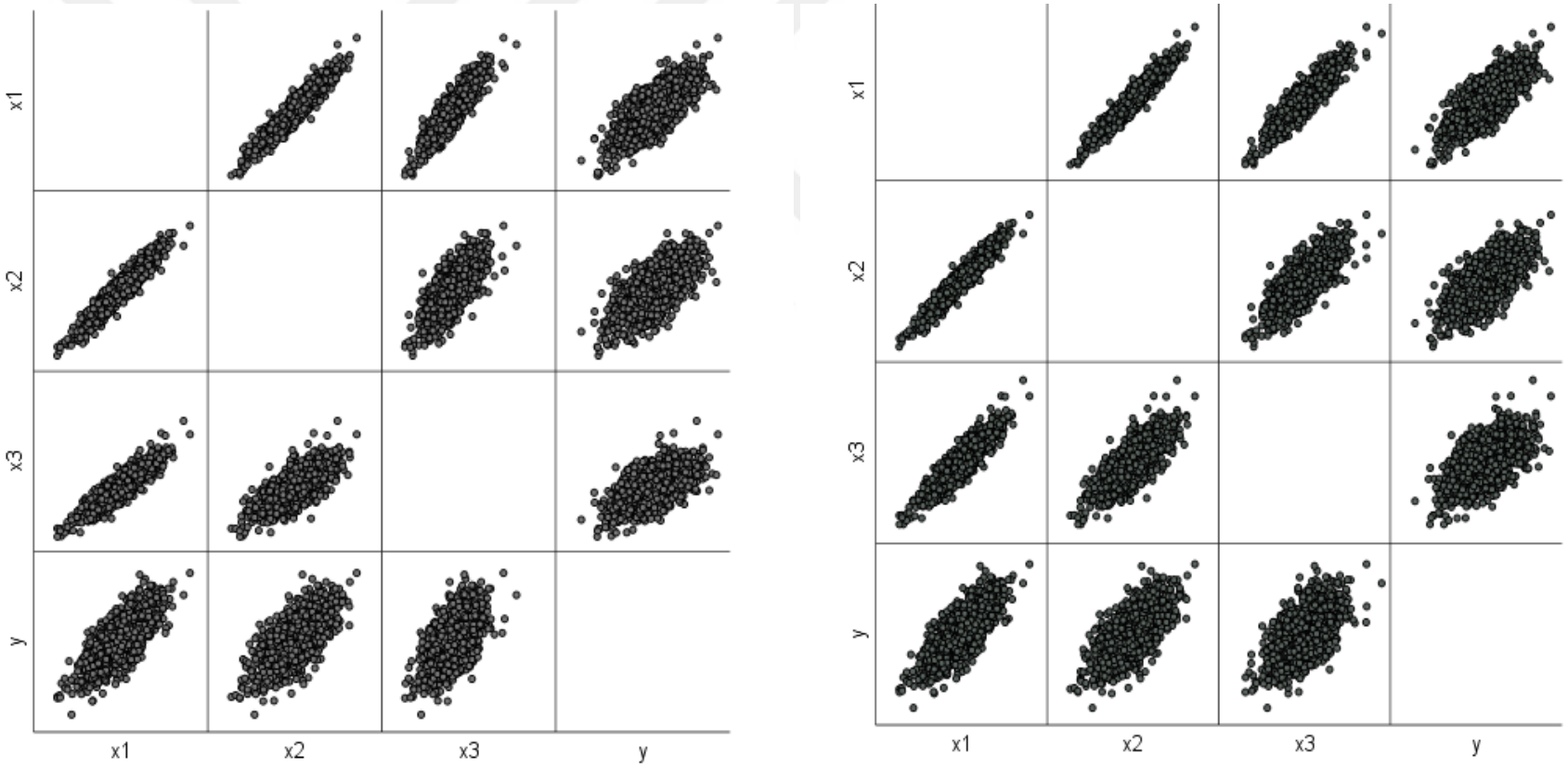
Şekil 4.2: Farklı derecelerde çoklubağlantı içeren veri seti grubu içindeki üçüncü ve dördüncü veri seti için ait saçılım grafikleri



Şekil 4.3: Farklı derecelerde çoklubağlantı içeren veri seti grubu içindeki beşinci ve altıncı veri seti için saçılım grafikleri



Şekil 4.4: Farklı derecelerde çoklubağlantı içeren veri seti grubu içindeki yedinci ve sekizinci veri seti için saçılım grafikleri



Şekil 4.5: Farklı derecelerde çoklubağlantı içeren veri seti grubu içindeki dokuzuncu ve onuncu veri seti için saçılım grafikleri

Tablo 4.1 incelendiğinde VIF değerlerini 10'un üzerinde, çoklu bağlantının derecesi arttığında 30'un üzerinde olduğu gözlenmiştir. Benzer şekilde, tolerans değerlerinin 0'a yaklaşmasıyla ve 0'a yakın özdeğerlerin elde edilmesiyle çoklu bağlantının varlığı ispat edilmiştir. Bir diğer çoklu bağlantı göstergesi olan koşul indeksleri de çoklu bağlantının varlığını desteklemiştir.

Bu veri setlerine uygulanan en küçük kareler regresyonu ve temel bileşenler regresyonu sonuçları Tablo 4.2 ile sunulmuştur. Tüm veri setleri için tümel modeller ve modeldeki katsayılar istatistiksel olarak anlamlı bulunmuştur. Çoklu bağlantı elde edebilmek için yapılan veri türetiminde tüm ilişkiler pozitif yönde tanımlanmıştır. Ancak, En Küçük Kareler çözümlemesinde çoklu bağlantının beklenen etkilerinden biri olarak x_2 ve x_3 bağımsız değişkenleri için regresyon katsayılarının işareti ters (negatif) olacak şekilde elde edilmiştir. Temel Bileşenler Regresyonu çözümlemesinde ise katsayıların işareti doğru yönde (pozitif) olarak bulunmuştur. EKK çözümlemesinde elde edilen katsayılar ile TBR analizi sonucunda elde edilen katsayılar işaretçe farklı olmakla beraber büyüklük olarak da birbirinden farklıdır. Ayrıca, TBR sonuçlarında katsayıların standart hataları EKK sonuçlarına göre daha düşüktür.

Tablo 4.2. Farklı derecelerde çoklu bağlantıya sahip veri setleri için EKK ve TBR ait sonuçlar tablosu

Veri Sayısı	Katsayılar	Ekk Regresyon Katsayıları	Ekk Standart Hata	Ekk R2	Ekk Sigma	Tbr Regresyonu Katsayıları	Tbr Standart Hata	Tbr R2	Tbr Sigma
n=1000	sabit	-0.00586		0.668512711	0.580366307	-0.01716		0.556861024	0.67102504
	x1	1.59867	0.07344			0.26096	0.00896		
	x2	-0.12979	0.03644			0.41149	0.02465		
	x3	-0.83471	0.05551			0.13588	0.01912		
n=1000	sabit	-0.005685		0.744456972	0.507873575	-0.019505		0.546636028	0.676467757
	x1	3.148592	0.103986			0.266733	0.008574		
	x2	-0.962421	0.052201			0.396218	0.024229		
	x3	-1.809292	0.071357			0.125027	0.020592		
n=1000	sabit	-0.005854		0.682443421	0.567366828	-0.01840445		0.541706872	0.681592896
	x1	1.926639	0.079796			0.2570498	0.008701		
	x2	-0.333434	0.041134			0.4041624	0.0257535		
	x3	-1.010515	0.056714			0.1210948	0.021338		
n=1000	sabit	0.00100		0.998867995	0.03305884	-0.02118		0.550601570	0.658687034
	x1	8.44752	0.01303			0.26831	0.00806		
	x2	-3.96968	0.00702			0.37778	0.02287		
	x3	-4.97492	0.00818			0.11844	0.02120		
n=1000	sabit	0.00187		0.998215812	0.041392357	-0.02712		0.502951218	0.690874776
	x1	10.83777	0.02010			0.27357	0.00873		
	x2	-5.81548	0.01182			0.35418	0.02342		
	x3	-5.99535	0.01172			0.12453	0.02341		
n=1000	sabit	-0.0011583		0.998977953	0.031351108	-0.02152536		0.535893557	0.668076745
	x1	8.386543	0.012095			0.2647931	0.007915		
	x2	-4.28833	0.007017			0.3674329	0.023475		
	x3	-4.495257	0.006940			0.108148	0.023453		
n=1000	sabit	0.00114		0.996843317	0.05504378	-0.02112		0.528508731	0.672712648
	x1	13.61870	0.03474			0.26849	0.00804		
	x2	-8.06414	0.02194			0.34123	0.02174		
	x3	-7.07178	0.01881			0.12313	0.02330		
n=1000	sabit	-0.00087		0.998415603	0.038985305	-0.02157		0.525494947	0.674666883
	x1	10.39442	0.01858			0.26555	0.00800		
	x2	-5.98341	0.01169			0.34862	0.02274		
	x3	-5.16923	0.00979			0.11132	0.02453		
n=1000	sabit	0.00134		0.998830334	0.033474948	-0.02214		0.522450897	0.676391255
	x1	8.41439	0.01280			0.26314	0.00798		
	x2	-4.71336	0.00805			0.35727	0.02389		
	x3	-3.99810	0.00656			0.09738	0.02600		
n=1000	sabit	-0.00087		0.999285843	0.026152056	-0.02305		0.519883684	0.678081934
	x1	7.06912	0.00833			0.26190	0.00800		
	x2	-3.86264	0.00527			0.36932	0.02525		
	x3	-3.19804	0.00415			0.07926	0.02780		

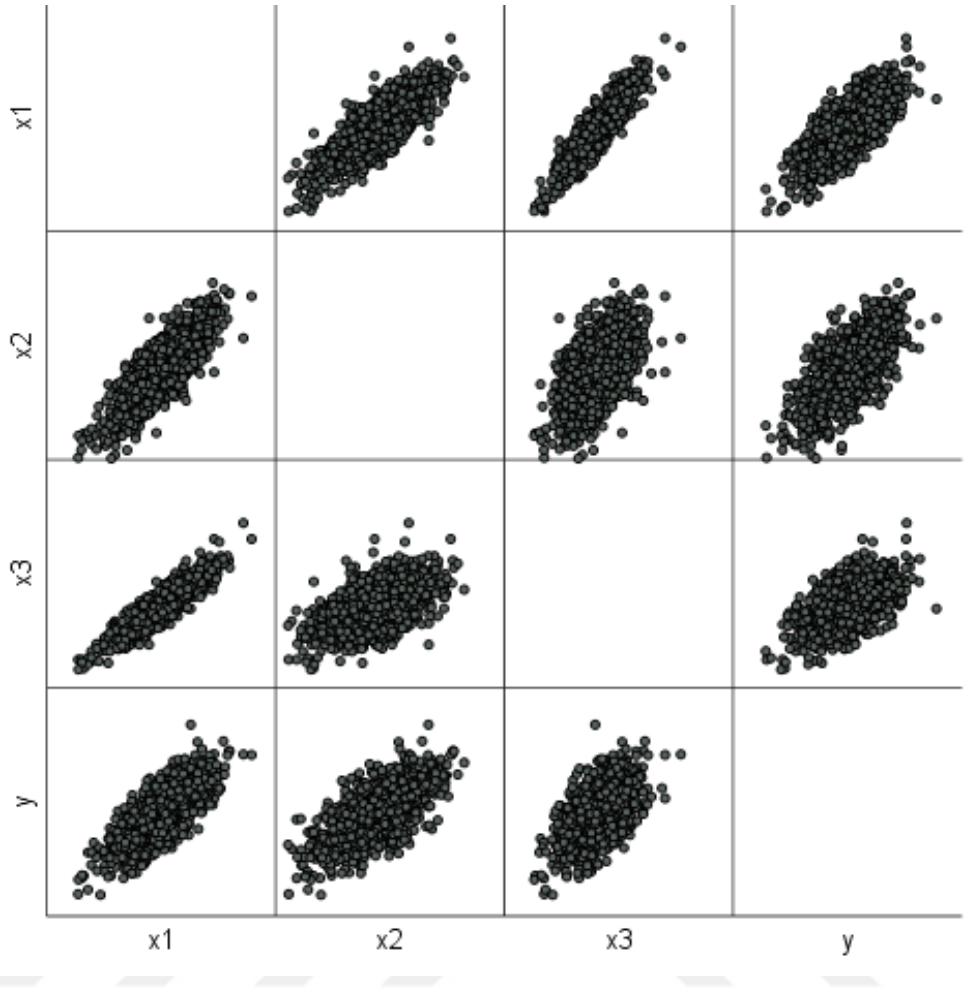
,

Korelasyon yapısı aynı ancak örneklem genişlikleri farklı olan ikinci veri grubu için türetilen 10 veri setine ait çoklu bağlantı göstergesi olan ölçüler Tablo 4.3 ve matris şeklindeki saçılım grafikleri Şekil 4.6 ile verilmiştir.

Tablo 4.3. Farklı örneklem genişliğine sahip ikinci veri grubu için çoklu bağlantı belirleme kriterleri tablosu

Veri Sayısı	Bağımsız Değişkenler	Vıf	Tolerans Değeri	Özdeğer	Koşul İndeksi
n=1000	x1	44.18220	0.02263	2.50335	1.00
	x2	10.89020	0.09183	0.48303	5.18
	x3	20.83430	0.04800	0.01362	183.84
n=2000	x1	39.66840	0.02521	2.49249	1.00
	x2	9.94120	0.10059	0.49237	5.06
	x3	18.86760	0.05300	0.01514	164.62
n=3000	x1	38.09180	0.02625	2.48898	1.00
	x2	9.65420	0.10358	0.49526	5.03
	x3	18.11290	0.05521	0.01576	157.90
n=4000	x1	37.06000	0.02698	2.46872	1.00
	x2	9.53510	0.10488	0.51515	4.79
	x3	17.76440	0.05629	0.01613	153.09
n=5000	x1	37.22420	0.02686	2.46611	1.00
	x2	9.47970	0.10549	0.51785	4.76
	x3	17.98890	0.05559	0.01604	153.78
n=6000	x1	37.10790	0.02695	2.47217	1.00
	x2	9.44470	0.10588	0.51172	4.83
	x3	17.86530	0.05597	0.01611	153.42
n=7000	x1	37.73080	0.02650	2.47252	1.00
	x2	9.57270	0.10446	0.51163	4.83
	x3	18.16440	0.05505	0.01585	156.04
n=8000	x1	37.87070	0.02641	2.47379	1.00
	x2	9.55570	0.10465	0.51043	4.85
	x3	18.27200	0.05473	0.01579	156.68
n=9000	x1	38.45940	0.02600	2.47758	1.00
	x2	9.63510	0.10379	0.50686	4.89
	x3	18.55850	0.05388	0.01556	159.25
n=10000	x1	39.18990	0.02552	2.48121	1.00
	x2	9.82430	0.10179	0.50351	4.93
	x3	18.81460	0.05315	0.01528	162.37

İkinci veri grubuna ait 10 veri seti için çoklu bağlantının göstergesi olan VIF, tolerans değeri, özdeğer ve koşul indeksi incelendiğinde veri setlerinde çoklu bağlantının varlığı gözlenmiştir.



Şekil 4.6: Çoklu bağlantısı olan ve farklı örneklem genişliğindeki veri grubunda yer alan örneklem genişliği 1000 olan veri seti için saçılım grafiği

İkinci veri grubu için de birinci veri grubunda olduğu gibi tüm değişkenler arasındaki korelasyonlar pozitif olacak şekilde veri türetimi yapılmıştır. Tablo 4.4 incelendiğinde EKK çözümlemesinde çoklu bağlantının beklenen etkilerinden biri olarak x_2 ve x_3 değişkeni için regresyon katsayılarının işareti ters (negatif) elde edilmiştir. TBR'de ise katsayılar koşullarımızı sağlayacak şekilde pozitif değerler almıştır. İlk veri grubundakine benzer şekilde iki çözümleme yönteminden elde edilen katsayılar ve standart hataları büyüklük olarak da farklılık göstermiştir.

Tablo 4.4. Çoklubağlantısı olan ve farklı örneklem genişliğine sahip EKK ve TBR'ye ait sonuçlar tablosu

Veri Sayısı	Katsayılar	Ekk Regresyon Katsayıları	Ekk Standart Hata	Ekk R2	Ekk Sigma	Tbr Regresyonu Katsayıları	Tbr Standart Hata	Tbr R2	Tbr Sigma
n=1000	sabit	-0.00569		0.744456972	0.507873575	-0.01951		0.546636028	0.676467757
	x1	3.14859	0.10399			0.26673	0.00857		
	x2	-0.96242	0.05220			0.39622	0.02423		
	x3	-1.80929	0.07136			0.12503	0.02059		
n=2000	sabit	-0.01994		0.732571571	0.522596495	-0.01916		0.540177528	0.685263818
	x1	3.03051	0.07289			0.27418	0.00626		
	x2	-0.90207	0.03660			0.39092	0.01736		
	x3	-1.73171	0.05063			0.13744	0.01499		
n=3000	sabit	-0.01014		0.733644104	0.520883395	-0.00930		0.542665242	0.68253741
	x1	2.98513	0.05854			0.27787	0.00513		
	x2	-0.88138	0.02941			0.38739	0.01410		
	x3	-1.69720	0.04081			0.14372	0.01230		
n=4000	sabit	-0.00420		0.727618272	0.519913905	-0.00803		0.534574178	0.679622399
	x1	2.96420	0.05058			0.27852	0.00450		
	x2	-0.87659	0.02532			0.37964	0.01198		
	x3	-1.67952	0.03518			0.14300	0.01050		
n=5000	sabit	-0.00111		0.728081641	0.516559944	-0.00295		0.536691116	0.674274561
	x1	2.94464	0.04507			0.27796	0.00401		
	x2	-0.86939	0.02259			0.37888	0.01069		
	x3	-1.66457	0.03132			0.14445	0.00923		
n=6000	sabit	-0.00039		0.729904603	0.517374668	0.00276		0.537915304	0.676717428
	x1	2.94726	0.04098			0.27831	0.00365		
	x2	-0.87114	0.02054			0.37711	0.00981		
	x3	-1.66311	0.02848			0.14735	0.00849		
n=7000	sabit	0.00093		0.731531880	0.519518133	0.00557		0.541426645	0.678982777
	x1	2.96416	0.03822			0.28026	0.00337		
	x2	-0.87890	0.01914			0.37647	0.00907		
	x3	-1.67239	0.02660			0.15118	0.00786		
n=8000	sabit	0.00219		0.731582743	0.519253345	0.00976		0.542527418	0.677885931
	x1	2.95177	0.03569			0.27929	0.00314		
	x2	-0.87293	0.01788			0.37687	0.00849		
	x3	-1.66557	0.02484			0.15072	0.00731		
n=9000	sabit	0.00204		0.731192541	0.519017008	0.00872		0.542756222	0.676916085
	x1	2.95467	0.03378			0.27782	0.00295		
	x2	-0.87693	0.01692			0.37491	0.00802		
	x3	-1.66618	0.02349			0.15152	0.00686		
n=10000	sabit	0.00280		0.732722038	0.51890498	0.00591		0.545106062	0.676957396
	x1	2.97171	0.03223			0.27779	0.00278		
	x2	-0.88486	0.01614			0.37573	0.00759		
	x3	-1.67833	0.02241			0.15109	0.00653		

5.TARTIŞMA

Çoklu doğrusal regresyon analizinin varsayımlarından biri olan bağımsız değişkenlerin birbirleriyle ilişkisinin olmaması varsayımı yerine getirilmediği takdirde çoklu doğrusal bağlantı problemi ile karşılaşılır. Bu durum ise kestirilmek istenen parametrelerin gerçek değerlerinin elde edilememesine, kestirimlerin mutlak değerlerinin büyük olmasına ve kestirimlerin işaretlerin değişmesine neden olabilecektir. Verideki çoklu bağlantı durumunun elde edilen regresyon modeli üzerine yapacağı olumsuz etkiler nedeniyle, bu durumun ortadan kaldırması ya da etkisinin indirgemesi yoluna gidilmelidir (13).

Çoklu bağlantı durumunun ortadan kaldırılması için önerilen bazı yöntemler vardır. İlk yapılması gereken regresyon modelinin oluşturulması sürecinde değişken seçiminin uygun bir şekilde yapılmasıdır. Ayrıca, veriye yeni gözlem eklenmesi, modelin yeniden oluşturulması ya da bazı yanlı kestirim yöntemlerinin kullanılması da çoklu bağlantının giderilmesi sürecinde kullanılan yöntemlerdir. Her bir yöntemin kendine göre uygulama alanı ve sakıncalı yönleri de var olabilir. Örneğin, oluşturulan örneklemin seçildiği evreni çok iyi temsil etmemesi sebebiyle ortaya çıkan bir çoklu bağlantı durumda veriye uygun yeni gözlemlerin eklenmesi tavsiye edilir. Ancak örnekleme birim ilave etmek her zaman mümkün olmayabilir. Bir veya birden fazla bağımsız değişkenin modelden atılması gerekebilir. Bu işlem modelin yeniden tanımlanması olarak adlandırılır. Bu süreçte de hangi değişkenlerin modelden çıkarılacağı bir sorun olabilir ve bu yaklaşım modeli yanlış tanımlamamıza neden olabilir (12, 26).

Çoklu bağlantının modelin üzerindeki olumsuz etkileriyle başa çıkabilmek için kullanılabilecek bir diğer yaklaşım yanlı kestirim yöntemlerinin kullanılmasıdır. Yanlı kestirim sonuçlarını kullanan yöntemlerden en çok tercih edilenleri Temel Bileşenler Regresyonu, Ridge Regresyon ve Kısmi En Küçük Kareler Regresyonudur.

Bu çalışmada ise çoklu bağlantının etkilerinin gözlendiği veri setlerine TBR uygulanarak çoklu bağlantı nedeniyle oluşan etkilerinin kaldırılması ve hangi durumlarda EKK regresyonu yerine TBR'nin kullanılabileceğinin belirlenmesi amaçlanmıştır. Bu nedenle, amacımız doğrultusunda farklı derecelerde çoklu bağlantıya sahip olacak şekilde türetilen ilk veri grubundaki 10 veri setine her iki yöntem de uygulanmıştır. Veri türetilirken tüm değişkenler arasındaki ilişkiler pozitif olacak şekilde tanımlandıkları halde çoklu bağlantının beklenen etkilerinden biri olarak EKK çözümlemesinde x_2 ve x_3 değişkeni için regresyon katsayılarının işareti negatif olarak elde edilmiştir. TBR sonuçlarında ise

katsayıların işaretleri pozitiftir. EKK regresyon katsayıları büyüklük olarak da TBR sonuçlarından farklıdır ve çoklu bağlantı arttıkça katsayı değeri de büyüme eğilimdedir. Aynı zamanda EKK katsayılarının standart hataları da TBR sonuçlarına göre büyüktür. Her ne kadar EKK çözümlemesi için açıklayıcılık daha fazla gözükse de varsayımlar sağlanmadığı için bu model yardımıyla yapılacak kestirimler doğru olmayacaktır.

Çoklu bağlantının etkisinin örneklem genişliği ile nasıl değiştiğinin incelenmesi için türetilen ikinci veri grubundaki 10 veri setine de hem EKK hem de TBR uygulanmıştır. Veri türetilirken sadece örneklem genişliğinin arttırıldığı ve korelasyon yapısının sabit tutulduğu bu veri setlerinde de çoklu bağlantının etkisi birinci veri grubundakine benzer şekilde gözlenmiştir. EKK yönteminde x_2 ve x_3 bağımsız değişkenlerinin regresyon katsayıları ters işaretli ve TBR'ye göre standart hatası büyük olacak şekilde elde edilmiştir. Örneklem genişliği arttıkça beklendiği üzere her iki yöntemde de standart hatalar küçülmüştür, ancak katsayı kestirimlerinde pek bir değişiklik gözlenmemiştir.

6. SONUÇ VE ÖNERİLER

Bu çalışma ile çoklu doğrusal regresyon çözümlemesinin varsayımlarından biri olan veride çoklu bağlantı olmaması koşulunun yerine getirilemediği durumlarda, bu varsayımın bozulmasının çözümleme sonuçları üzerine etkileri simülasyon ile türetilmiş verilerde gözlenmiştir. Regresyon çözümlemelerindeki temel amaçlardan biri olan kestirim yapma amacıyla elde edilen modelin kullanılabilirliği bu varsayımın bozulması ile tehlikeye girmektedir.

Çok değişkenli modellemeler yapılırken çoklu bağlantının varlığı incelenmeli ve bu duruma çözüm olabilecek regresyon yöntemlerinden biri kullanılmalıdır. Aksi takdirde yapılacak kestirimler yanlış ve yanlı sonuçlara götürebilecektir.

Çoklu doğrusal regresyon çözümlemesi yapılırken veride çoklu bağlantı olduğu belirlendiği taktirde, yapılan bu çalışmanın sonuçları doğrultusunda en küçük kareler regresyonu yerine temel bileşenler regresyonunun kullanılması önerilmektedir.

Buna benzer sonraki yapılacak çalışmalarda daha az varsayım gerektirmesi nedeniyle veri bilimi yöntemlerinin kullanılması ve daha iyi sonuçlar elde edilmesi beklenmektedir.

KAYNAKLAR

1. Alpar R. *Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş*, Detay Yayıncılık 2013.

2. Özdamar K. *Paket Programlar İle İstatistiksel Veri Analizi*. Eskişehir, Kaan Kitabevi 2004.

3. Şahinler S. En küçük kareler yöntemi ile doğrusal regresyon modeli oluşturmanın temel prensipleri. *Mustafa Kemal Üniversitesi Ziraat Fakültesi Dergisi* 2000; 5(1-2): 57-73.

4. Galton F. On the anthropometric laboratory at the late International Health Exhibition. *MAN* 1885; 14205-21.

5. Galton F, Farr W. Considerations Adverse to the Maintenance of Section F (Economic Science and Statistics). *MAN* 1877; 14205-21.

6. Akdeniz F. *Olasılık ve istatistik*, Akademisyen Kitabevi 2015.

7. Seber GA, Lee AJ. *Linear regression analysis*, John Wiley & Sons 2012.

8. Ünver Ö, Gamgam H. *Uygulamalı İstatistik Yöntemler*. Ankara, Siyasal Kitabevi 1996.

9. Karagöz M. *İstatistik yöntemleri*, Ekin Kitabevi 2006.

10. Alpar R. *Spor, sağlık ve eğitim bilimlerinden örneklerle uygulamalı istatistik ve geçerlik-güvenirlik*, Detay Yayıncılık 2010.

11. Özdamar K. *SPSS ile Biyoistatistik.* . Eskişehir, Kaan Kitabevi 2001: 315-68.

12. Montgomery DC, Peck EA, Vining GG. *Introduction to linear regression analysis*, John Wiley & Sons 2012.

13. Albayrak AS. Çoklu Doğrusal Bağlantı Halinde En Küçük Kareler Tekniğinin Alternatifi Yanlı Tahmin Teknikleri ve Bir Uygulama. *Uluslararası Yönetim İktisad ve İşletme Dergisi* 2012; 1(1): 105-26.

14. Çekerol G, Nalçakan M. Lojistik Sektörü İçerisinde Türkiye Demiryolu Yurtiçi Yük Taşıma Talebinin Ridge Regresyonla Analizi. *MU Iktisadi ve Idari Bilimler Dergisi* 2011; 31(2): 321-44.

15. Pamukçu E, Çolak C, Çalık S, Kuzu Z. Sistolik Kan Basıncının Tahmininde Yanlı Regresyon Yöntemlerinin Kullanılması. *Turgut Özal Tıp Merkezi Dergisi* 2010; 17(4).

16. Orhunbilge N. *Uygulamalı Regresyon ve Korelasyon analizi*. Nobel Yayın 2017.

17. Kalaycı Ş. *SPSS uygulamalı çok değişkenli istatistik teknikleri*, Asil Yayın Dağıtım Ankara, Turkey 2010.

18. Türkay GS. Ridge Regresyon Yöntemiyle Tofaş Firmasının (1975-1994) Yılları Arası Otomobil Talep Miktarı Analizi. Sosyal Bilimler Enstitüsü, İşletme Anabilim Dalı. Yüksek Lisans, Eskişehir: Anadolu Üniversitesi 1996.

19. İmir E. Çoklu Bağlantılı Doğrusal Modellerde Ridge Regresyon Yöntemiyle Parametre Kestirimi. *Anadolu Üniversitesi Yayınları* 1986; (212).

20. Uslu VR. Ridge Regresyon ve Öğrenci Başarısı Üzerine Bir Uygulama. Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı. Yüksek Lisans, Samsun: Ondokuz Mayıs Üniversitesi 1991.

21. Sümbüloğlu K, Sümbüloğlu V. *Biyoistatistik,*. Ankara, Hatipoğlu Yayıncılık 2010.

22. Akdeniz F, Çabuk A. Ridge regresyon teorisinde 1970-2001 arasındaki gelişmeler. *V Ulusal Ekonometri ve İstatistik Sempozyumu 20-22 Eylül, Çukurova Üniversitesi, Adana* 2001.

23. Wetherill GB. *Regression analysis with application*, Chapman & Hall, Ltd. 1987.

24. Özkale R. Çoklu İç İlişki ile İlgili Yöntemler. Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı. Doktora Tezi, Adana: Çukurova Üniversitesi 2007.

25. Marquaridt DW. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics* 1970; 12(3): 591-612.

26. Orçanlı K, Birgören B, Oktay E. Çok Değişkenli Kalite Kontrolünde Süreç Tabanlı Temel Gösterimleri Yönteminin Hata Teriminde Kovaryansın Etkileri. *Sosyal Bilimler Araştırma Dergisi* 2017; 6(2): 20-40.

27. AKTAŞ C. Çoklu bağıntı ve Liu kestiricisiyle enflasyon modeli için bir uygulama. *Uluslararası Yönetim İktisat ve İşletme Dergisi* 2012; 3(6): 67-80.

28. Gujarati DN, Porter DC, Şenesen Ü, Günlük-Şenesen G. *Temel ekonometri*, Literatür Yayıncılık 2012.

29. Freund E. Matematiksel İstatistik. *Literatür Yayıncılık, İstanbul* 2002.

30. Corp I. IBM SPSS statistics for windows, version 25.0. *Armonk, NY: IBM Corp* 2017.

31. Faraway JJ. *Extending the linear model with R (Texts in Statistical Science)*, Chapman & Hall 2005.

32. Cankaya S. A comparative study of some estimation methods for parameters and effects of outliers in simple regression model for research on small ruminants. *Trop Anim Health Prod* 2009; 41(1): 35-41.

33. Özgül V, Alma ÖG. Regresyon analizinde kullanılan en küçük kareler ve en küçük medyan kareler yöntemlerinin karşılaştırılması. *Süleyman Demirel Üniversitesi Fen Edebiyat Fakültesi Fen Dergisi* 3(2): 219-29.

34. İşi A. Yanlı Tahmin Ediciler ve Kombinasyonları. Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı. Yüksek Lisans Tezi, Ankara: Gazi Üniversitesi 2002.

35. Göktaş A, Öznur İ. Türkiye'de İşsizlik Oranının Temel Bileşenli Reegresyon Analizi ile Belirlenmesi *Sosyal Ekonomik Araştırmalar Dergisi* 2010; 10(20): 279-94.

36. Aşkın F. Ortalama Artelyel Kan Basıncını Etkileyen Faktörlerin Temel Bileşenler Regresyonu İle Belirlenmesi. Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı. Yüksek Lisans, Elazığ: Fırat Üniversitesi 2011.

37. Ortabaş N. Principal components in the problem of multicollineartity. Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı. Yüksek Lisans Tezi, İzmir: Dokuz Eylül Üniversitesi 2001.

38. biostatapps.inonu.edu.tr/NAT/ 2018

39. Hintze J. NCSS 2007. *Statistical analysis and graphics, user's guide* 2007.

EKLER

EK-1. Özgeçmiş

ÖZGEÇMİŞ

Adı Soyadı: Zeynep TUNÇ

Doğum Tarihi: 1988

Öğrenim Durumu: Yüksek Lisans

Derece	Bölüm/Program	Üniversite	Yıl
Lisans	Matematik	Çukurova Üniversitesi	2010
Y. Lisans	Biyoistatistik ve Tıp Bilişimi AD	İnönü Üniversitesi	2018

Yüksek Lisans Tez Başlığı ve Tez Danışman(lar)ı:

En küçük kareler ve temel bileşenler regresyon analizlerinin karşılaştırılması, Dr. Öğretim Üyesi Harika Gözde GÖZÜKARA BAĞ

Görevler:

Görev Unvanı	Görev Yeri	Yıl
Arş. Gör.	İnönü Üniversitesi	2014 – Devam ediyor

EK-2. Etik Kurul Almama Gerekçesi

Evrak Tarih ve Sayısı: 05/01/2017-E.1344

T.C.
İNÖNÜ ÜNİVERSİTESİ REKTÖRLÜĞÜ

Tıp Fakültesi Dekanlığı
Biyoistatistik Anabilim Dalı Başkanlığı

Sayı : 58137357-020
Konu : Zeynep TUNÇ Tez Öneri Formu

SAĞLIK BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Enstitünüz 38140042004 numaralı öğrencisi Zeynep TUNÇ'un tez öneri formu ektedir.
Gereğini arz ederim.

e-imzalıdır
Prof.Dr. Saim YOLOĞLU
Anabilim Dalı Başkanı

Ek:
1- 1 Adet Tez Öneri Formu
2- 1 Adet Etik Kurul Almama Gerekçesi

05/01/2017 Öğretim Elemanı : Ar.Gör. Şeyma YAŞAR

Tıp Fakültesi Dekanlığı
Telefon No: 3410660 Faks No: 3410036
E-Posta: biyoistatistik@inonu.edu.tr İnternet Adresi: https://www.inonu.edu.tr/tr/cms/biyoistatistik
Bilgi İçin: Şeyma YAŞAR
Unvan: Öğretim Elemanı
Telefon No: 3410660

EK-2. Etik Kurul Almama Gerekçesi (Devamı)

SAĞLIK BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

13 Nisan 2013 tarih ve 28617 sayı ile T.C. Resmi Gazetede yayınlanan "Klinik Araştırmalar Hakkında Yönetmelik"in birinci bölümünün 2.maddesinin 1.fıkrası (Bu yönetmelik biyoyararlanım ve biyoeşdeğerlik çalışmaları dahil, ruhsat veya izin alınmış olsa dahi insanlar üzerinde yapılacak olan ilaç, tıbbi ve biyolojik ürünler ile bitkisel ürünlerin klinik araştırmaları, klinik araştırma yerlerini ve bu araştırmaları gerçekleştirecek gerçek veya tüzel kişileri kapsar.) gereğince yüksek lisans öğrencisi Zeynep TUNÇ'un tezinin klinik bir çalışma olmaması, kullanılacak verinin bilgisayar ortamında benzetim teknikleriyle türetilecek olması sebebiyle Etik Kurul kararı alınmamıştır.