

**T.C.
ORDU ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**LİNEER İSTATİSTİKSEL MODELLER, ÇOK DEĞİŞKENLİ
LİNEER REGRESYONDA İSTATİSTİKSEL TESTLERİN
GEOMETRİK YORUMU VE GEOMETRİK OLARAK KOLAY
KORELASYON BİLGİLERİ**

HAKAN GEZGİN

YÜKSEK LİSANS TEZİ

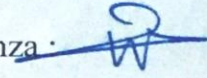
ORDU 2016

TEZ ONAY

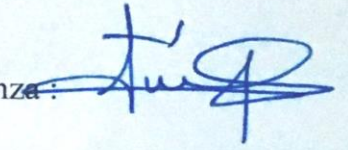
Ordu Üniversitesi Fen Bilimleri Enstitüsü öğrencisi Hakan GEZGİN tarafından hazırlanan ve Prof. Dr. Cemil YAPAR danışmanlığında yürütülen “Lineer İstatistiksel Modeller, Çok Değişkenli Lineer Regresyonda İstatistiksel Testlerin Geometrik Yorumu ve Geometrik Olarak Kolay Korelasyon Bilgileri” adlı bu tez, jürimiz tarafından 23/05/2016 tarihinde oy birliği / ~~oy çokluğu~~ ile Matematik Anabilim Dalında Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman : Prof. Dr. Cemil YAPAR

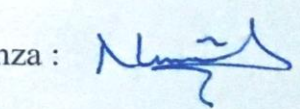
Başkan : Prof. Dr. Cemil YAPAR
Matematik, Ordu Üniversitesi

İmza : 

Üye : Doç. Dr. Selahattin MADEN
Matematik, Ordu Üniversitesi

İmza : 

Üye : Yrd. Doç. Dr. Nurgül OKUR BEKAR
İstatistik, Giresun Üniversitesi

İmza : 

ONAY:

Bu tezin kabulü, Enstitü Yönetim Kurulu'nun 26./05./2016. tarih ve ..2016../261 sayılı kararı ile onaylanmıştır.



TEZ BİLDİRİMİ

Tez yazım kurallarına uygun olarak hazırlanan bu tezin yazılmasında bilimsel ahlak kurallarına uyulduğunu, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, tezin içerdiği yenilik ve sonuçların başka bir yerden alınmadığını, kullanılan verilerde herhangi bir tahrifat yapılmadığını, tezin herhangi bir kısmının bu üniversite veya başka bir üniversitedeki başka bir tez çalışması olarak sunulmadığını beyan ederim.



Hakan GEZGİN

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

ÖZET

LİNEER İSTATİSTİKSEL MODELLER, ÇOK DEĞİŞKENLİ LİNEER REGRESYONDA İSTATİSTİKSEL TESTLERİN GEOMETRİK YORUMU VE GEOMETRİK OLARAK KOLAY KORELASYON BİLGİLERİ

Hakan GEZGİN

Ordu Üniversitesi

Fen Bilimleri Enstitüsü

Matematik Anabilim Dalı, 2016

Yüksek Lisans Tezi, 125s.

Danışman: Prof. Dr. Cemil YAPAR

Bu tez dört bölümden oluşmaktadır. Birinci bölümde tezin amacını belirten bir giriş kısmı verilmiştir. İkinci bölümde lineer istatistiksel modeller genel olarak tanıtılmıştır. Lineer modeller hakkında bazı geometrik yorumlar verilmiş ve konular daha anlaşılabilir bir duruma getirilmiştir. Üçüncü bölümde çoklu regresyon, regresyon analizi, regresyon analizinde singüler değer ayrışımının kullanımı ve çok değişkenli lineer regresyonda istatistiksel testlerin geometrik yorumu verilmiştir. Son bölümde kolay korelasyon bilgileri ve bunların geometrik yorumu sunulmuştur.

Anahtar Kelimeler: Aitken Modeli, En Küçük Kareler, Lineer Yansız Tahmin Edici, Doğrudaşlık, Çoklu Regresyon, Rasgele Dönme, Çok Değişkenli Analiz, Singüler Değer Ayrışımı, Ortogonallik, Korelasyon, Anova.

ABSTRACT

LINEAR MODELS, THE GEOMETRICAL INTERPRETATION OF STATISTICAL TESTS IN MULTIVARIATE LINEAR REGRESSION AND GEOMETRICALLY EASY CORRELATION TRICKS

Hakan GEZGİN

University of Ordu

Institute for Graduate Studies in Science and Technology

Department of Mathematics, 2016

MSc. Thesis, 125p.

Supervisor: Prof. Dr. Cemil YAPAR

This thesis has been consisted of four chapters. In the first chapter the introduction part which states the aim of thesis has been given. In the second chapter generally linear statistical models have been introduced. Some geometrical interpretations about linear models have been given and the matter has been rendered more understandable case. In the third chapter, multiple regression, regression analysis, use of singular value decomposition in regression analysis and the geometrical interpretation of statistical tests in multivariate linear regression have been given. In the last chapter, easy correlation tricks and their geometrical interpretation have been presented.

Key Words: Aitken Model, Least Squares, Linear Unbiased Estimator, Colinearity, Multiple Regression, Random Rotation, Multivariate Analysis, Singular Value Decomposition, Orthogonality, Correlation, Anova.

TEŐEKKÜR

Tüm alıőmalarım boyunca kıymetli zamanını ayırarak alıőmalarıma yardımcı olan, engin bilgi ve deneyimleriyle bana yol gösteren deęerli danıőmanım Prof. Dr. Cemil YAPAR'a en içten teőekkürlerimi sunarım. Ayrıca, her zaman bilgi ve görüşlerine başvurduğum Do. Dr. Selahattin MADEN hocam ve tüm Ordu Üniverstesi Matematik Bölümü öğretim üyelerine teőekkür ederim. Hem bu süreçte hem de hayatım boyunca yanımda olan ve ideallerimi gerçekleőtirmemi saęlayan deęerli aileme yürekten teőekkürü bir bor bilirim. Ayrıca, alıőmalarım boyunca desteklerini esirgemeyen, bana moral veren deęerli dostlarım Ömer Fatih YAZICIOęLU'na, Cem YILMAZ'a ve her zaman yanımda olan deęerli Meltem SÖNMEZ'e teőekkürlerimi sunarım.

İÇİNDEKİLER

TEZ BİLDİRİMİ	I
ÖZET	II
ABSTRACT	III
TEŞEKKÜR	IV
İÇİNDEKİLER	V
ŞEKİLLER LİSTESİ	VI
ÇİZELGELER LİSTESİ	VII
SİMGELER VE KISALTMALAR	VIII
ÖNCEKİ ÇALIŞMALAR	IX
1. GİRİŞ	1
2. LİNEER İSTATİSTİKSEL MODELLER	2
2.1. Lineer Modelin Tanımı	2
2.2. Lineer Modellerin Genel Durumları	2
2.3 Lineer Modellerin Geometrisi	15
3. REGRESYON ANALİZİ	54
3.1 Çoklu Regresyon	54
3.2 Regresyon Analizinde Singüler Değer Ayırışımının Kullanımı	63
3.3 Çok Değişkenli Lineer Regresyonda İstatistiksel Testlerin Geometrik Yorumu	91
EK 3.1 Rasgele dönme matrisleri	103
4. KOLAY KORELASYON BİLGİLERİ	105
4.1 Ortogonalite ve İlişkisizlik	107
4.2 \mathbf{j} Vektörünün Rolü Üzerine	110
4.3 Geometrik Olarak Kısmi Korelasyon	117
5. SONUÇ VE ÖNERİLER	121
KAYNAKLAR	122
ÖZGEÇMİŞ	125

ŞEKİLLER LİSTESİ

<u>Şekil No</u>	<u>Sayfa</u>
Şekil 2.1. İki-bileşenli vektörler	16
Şekil 2.2. Vektörlerin toplamı.....	17
Şekil 2.3. Vektörlerin skalerle çarpımı	18
Şekil 2.4. R^n de bir bulut.....	22
Şekil 2.5. Rasgele vektörler	23
Şekil 2.6. $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_2)$ rasgele vektörü için yoğunluk bulutu	25
Şekil 2.7. Y nin yoğunluk bulutu	26
Şekil 2.8. Bir örnek olarak R^3	28
Şekil 2.9. v_1 ve v_2 vektörü tarafından üretilen vektör uzayı	28
Şekil 2.10. X in sütun uzayları.....	32
Şekil 2.11. Y nin ortogonal izdüşümü	33
Şekil 2.12. Tahmin edilen hata vektörü	34
Şekil 2.13. X in sütun uzayı.....	36
Şekil 2.14. İki vektör arasındaki açı.....	38
Şekil 2.15. u nun v üzerindeki izdüşümü	39
Şekil 2.16. Hata uzayı	43
Şekil 2.17. Y nin μ_Y ye ortogonal olmayan bir altuzay üzerine izdüşümü.....	45
Şekil 2.18. Hata uzayı	47
Şekil 3.1. Regresyon yüzeyi.....	61
Şekil 3.2. Bir regresyon yüzeyinin geometrik gösterimi	67
Şekil 3.3. İki açıklayıcı değişken durumunda singüler değer ayrışımının geometrik yorumu	71
Şekil 3.4. $x_1 = 3$ olduğunda, üç x_1 , x_2 ve x_3 açıklayıcı değişkeni için yakın doğrudaşlık.....	76
Şekil 3.5. Doğrudaşlığın regresyon yüzeyi üzerindeki etkisi	80
Şekil 4.1. Geometrik olarak x ve y arasındaki korelasyon	107
Şekil 4.2. Geometrik olarak $r_{xy \cdot z}$ kısmi korelasyonu.....	110
Şekil 4.3. x ve y arasında kosinüs yüksek fakat korelasyon yok.....	111
Şekil 4.4. $r_{xy \cdot z}$ kısmi korelasyonu 0 a eşit.....	111
Şekil 4.5. $GKT = RKT + HKT$ nin açıklaması	117

ÇİZELGELER LİSTESİ

<u>Çizelge No</u>	<u>Sayfa</u>
Çizelge 2.1. Veri Çizelgesi	19
Çizelge 2.2. Sonda uzunluğu verisi.....	20
Çizelge 3.1. A veri kümesi.....	65
Çizelge 3.2. A veri kümesinin SDA	68
Çizelge 3.3. İki değişkenli bir \mathbf{X} matrisinin SDA	70
Çizelge 3.4. B veri kümesi.....	77
Çizelge 3.5. B veri kümesinin \mathbf{X} matrisin SDA; $\mathbf{X} = \mathbf{U}\mathbf{\Theta}\mathbf{V}'$	77
Çizelge 3.6. $\lambda_1, \dots, \lambda_s$ özdeğerleri için tanımlanan başka matrisler.....	95
Çizelge 3.7. Dört değişmez test istatistiği ve karşılık gelen blok tasarım ölçütü	95



SİMGELER VE KISALTMALAR

ANCOVA	:	Kovaryans analizi
ANOVA	:	Varyans analizi
$\text{boy}(S)$:	S vektör uzayının boyutu
$\mathcal{C}(\mathbf{X})$:	\mathbf{X} matrisinin sütun uzayı
$E()$:	Beklenen değer
GKT	:	Genel kareler toplamı
HKT	:	Hata kareler toplamı
Kov	:	Kovaryans
MANOVA	:	Çok değişkenli varyans analizi
OLSE	:	Alışılmış en küçük kareler tahmini
öbd	:	Özdeş bağımsız dağılan
$\text{Pr}()$:	Olasılık
$\text{rank}(\mathbf{X})$:	\mathbf{X} matrisinin rankı
RKT	:	Regresyon kareler toplamı
SDA	:	Singüler değer ayrışımı
$\ \mathbf{u}\ $:	\mathbf{u} vektörünün normu
Var	:	Varyans
Yan	:	Yanlılık

ÖNCEKİ ÇALIŞMALAR

Tebbs Joshua M. (2010) lineer istatistiksel modellerin genel bir tanımını verdi ve varyans analizi ve kovaryans analizi modellerine lineer model yaklaşımlarını sundu. Miller Arden (2015) lineer modellere geometrik yaklaşımını sundu. Wood Frank (2011) çoklu regresyon analizinin tanımını verdi. Mandel John (1982) regresyon analizinde singüler değer ayrışımının kullanımını ve singüler değer ayrışımının geometrik yorumunu verdi. Langsrud Øyvind (2004) çok değişkenli lineer regresyonda istatistiksel testlerin geometrik yorumunu yaptı. Isolato Jarkko, P.H. Styan George, Puntanen Simo (2011) kolay korelasyon bilgileri ve geometrik olarak kısmi korelasyon üzerine çalışmaları sundu.



1. GİRİŞ

İstatistikte, lineer modeller kavramı geniş bir yer tutar. Özellikle çok deęişkenli lineer modellerde parametre tahminleri, parametrelere ait güven aralıkları ve parametreler üzerine koyulan bazı lineer hipotezlerin testleri ve bunların istatistiksel yorumları bu konularda verilen literatürde geniş bir uygulama alanı bulur. Ekonometride, bu hipotezlerin ekonometrik anlamlarına yer verilerek elde edilen sonuçlar hakkında yorumlar yapılır. Tüm bunların yanında istatistiksel kavramlara geometrik yorumlarla yaklaşım, bu kavramların daha kolay bir şekilde anlaşılmasını sağlar. Okuyucuların matematiksel kavramlarda çekeceęi zorluklar düşünülerek bazı açıklamalar, matematiksel bilgiler, özellikle lineer cebir ve geometri bilgileri, gerekli olduęu yerlerde verilmeye çalışıldı.

2. LINEER İSTATİSTİKSEL MODELLER

Tebbs (2010), lineer istatistiksel modeller adlı çalışmasında, bu modellerin farklı tiplerini sundu ve bunlar hakkında istatistiksel değerlendirmelere geniş bir yer ayırdı. Bu bölümde bu bilgilerin bazılarını sunacağız.

2.1. Lineer Modelin Tanımı

Giriş: **Lineer modeller** parametrelerine göre lineer olan modellerdir. Bir lineer modelin genel biçimi,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.1)$$

ile verilir. Burada \mathbf{Y} gözlenen tepkimelerin bir $n \times 1$ vektörüdür, \mathbf{X} sabit değişmezlerin bir $n \times p$ (tasarım) matrisidir, $\boldsymbol{\beta}$ sabit fakat bilinmeyen parametrelili bir $p \times 1$ vektörüdür ve $\boldsymbol{\varepsilon}$ (gözlenemeyen) rasgele hataların bir $n \times 1$ vektörüdür. \mathbf{Y} tepkime vektörünün ortalaması bilinmeyen $\boldsymbol{\beta}$ parametresine göre lineer olduğundan modele bir lineer model denir.

Kapsam: İstatistikte yaygın olarak kullanılan çeşitli modeller $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ genel lineer modelinin örnekleridir. Bunlar lineer regresyon modelleri ve varyans analizi (ANOVA) modellerini içerir, ancak bunlarla sınırlı değildirler. ANOVA modelleri sıfır ve birlerden ibaret olan \mathbf{X} matrisine başvururken, regresyon modelleri genellikle tam ranklı olan \mathbf{X} matrisine başvururlar.

2.2. Lineer Modellerin Genel Durumları

- **Model 1:** $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ *En küçük kareler modeli:* Bu model $\boldsymbol{\varepsilon}$ üzerinde varsayımlar yapmaz. Parametre uzayı $\Theta = \{\boldsymbol{\beta} : \boldsymbol{\beta} \in R^p\}$ dir.
- **Model 2:** $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ *Gauss Markov modeli:* Burada $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ ve $\text{kov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$ dir. Parametre uzayı $\Theta = \{(\boldsymbol{\beta}, \sigma^2) : (\boldsymbol{\beta}, \sigma^2) \in R^p \times R^+\}$ dir.
- **Model 3:** $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ *Aitken modeli:* Burada $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\text{kov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{V}$ ve \mathbf{V} bilinmeyen bir matristir. Parametre uzayı $\Theta = \{(\boldsymbol{\beta}, \sigma^2) : (\boldsymbol{\beta}, \sigma^2) \in R^p \times R^+\}$ dir.

- **Model 4: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ Genel lineer karma modeli:** Burada $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ ve $\text{kov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ 'dir. $\boldsymbol{\Omega}$, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ nın pozitif tanımlı olduğu $\boldsymbol{\theta}$ nın tüm değerlerinin kümesi olmak üzere, parametre uzayı $\Theta = \{(\boldsymbol{\beta}, \boldsymbol{\theta}) : (\boldsymbol{\beta}, \boldsymbol{\theta}) \in R^p \times \boldsymbol{\Omega}\}$ dir.

Gauss Markov Modeli: $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ ve $\text{kov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$ olmak üzere; $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ lineer modelini göz önüne alınız. Şimdi, bu modelin özel durumlarını belirteceğiz.

Örnek 2.1 (Bir örneklem problemi): Y_1, Y_2, \dots, Y_n nin μ ortalamalı ve σ^2 pozitif varyanslı özdeş bağımsız dağılan (öbd) bir örneklem olduğunu varsayalım. Eğer, $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ ler $E(\varepsilon_i) = 0$ ortalamalı ve σ^2 ortak varyanslı iseler,

$$\mathbf{Y}_{n \times 1} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{X}_{n \times 1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \boldsymbol{\beta}_{1 \times 1} = \mu, \quad \boldsymbol{\varepsilon}_{n \times 1} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (2.2)$$

olmak üzere, Gauss Markov (GM) modelini $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ biçiminde yazabiliriz. $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ ve $\text{kov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$ olduğuna dikkat ediniz.

Örnek 2.2 (Basit lineer regresyon): $i = 1, 2, \dots, n$ için ε_i ler 0 ortalamalı ve σ^2 pozitif ortak varyanslı ilişkisiz rasgele değişkenler olmak üzere, bir Y tepkime değişkeninin,

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (2.3)$$

vasıtasıyla bağımsız bir x değişkeni ile lineer ilişkili olduğu modeli göz önüne alınız. Eğer x_1, x_2, \dots, x_n ler hatasız ölçülen sabit değişmezler iseler, bu takdirde bu,

$$\mathbf{Y}_{n \times 1} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{X}_{n \times 2} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \boldsymbol{\beta}_{2 \times 1} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \boldsymbol{\varepsilon}_{n \times 1} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (2.4)$$

olmak üzere, bir $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ GM modelidir. $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ ve $\text{kov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$ olduğuna dikkat ediniz.

Örnek 2.3 (*Çoklu lineer regresyon*): $i = 1, 2, \dots, n$ için, ε_i ler 0 ortalamalı ve σ^2 pozitif ortak varyanslı ilişkisiz rasgele değişkenler olmak üzere, bir Y tepkime değişkeninin,

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i \quad (2.5)$$

vasıtasıyla x_1, x_2, \dots, x_k diyeceğimiz, birtakım bağımsız değişkenle lineer ilişkili olduğunu farz ediniz. Eğer bağımsız değişkenler hatasız ölçülen sabit değişmezler ise, bu takdirde,

$$\mathbf{Y}_{n \times 1} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{X}_{n \times p} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}, \quad \boldsymbol{\beta}_{p \times 1} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \quad \boldsymbol{\varepsilon}_{n \times 1} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (2.6)$$

ve $p = k + 1$ olmak üzere, bu model özel bir $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ GM modelidir. $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ ve $\text{kov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$ olduğuna dikkat ediniz.

Örnek 2.4 (*Bir yönlü ANOVA*): $a \geq 2$ sayıda işlemi karşılaştırmak için yapılan bir deneyi göz önüne alınız. i -yinci işlem seviyesi için n_i sayıda deneysel birimin rasgele seçildiklerini ve i -yinci işleme tahsis edildiklerini farz ediniz. $i = 1, 2, \dots, a$, $j = 1, 2, \dots, n_i$ ve ε_{ij} rasgele hataları sıfır ortalamalı ve $\sigma^2 > 0$ ortak varyanslı ilişkisiz rasgele değişkenler olmak üzere,

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad (2.7)$$

modelini göz önüne alınız. Eğer a sayıda $\alpha_1, \alpha_2, \dots, \alpha_a$ işlem etkilerine en iyi değişmez sabitler olarak bakılırsa, bu takdirde bu model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ GM modelinin özel bir durumudur. Bunu görmek için, $p = a + 1$ ve $\boldsymbol{\varepsilon}_{n \times 1} = (\varepsilon_{11}, \varepsilon_{11}, \dots, \varepsilon_{an_i})'$, \mathbf{j}_{n_i} birlerin bir $n_i \times 1$ vektörü, $\mathbf{0}_{n_i}$ sıfırların bir $n_i \times 1$ vektörü ve $n = \sum_{i=1}^a n_i$ olmak üzere,

$$\mathbf{Y}_{n \times 1} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{an_a} \end{pmatrix}, \quad \mathbf{X}_{n \times p} = \begin{pmatrix} \mathbf{j}_{n_1} & \mathbf{j}_{n_1} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \mathbf{j}_{n_2} & \mathbf{0}_{n_2} & \mathbf{j}_{n_2} & \cdots & \mathbf{0}_{n_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{j}_{n_a} & \mathbf{0}_{n_a} & \mathbf{0}_{n_a} & \cdots & \mathbf{j}_{n_a} \end{pmatrix}, \quad \boldsymbol{\beta}_{p \times 1} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_a \end{pmatrix} \quad (2.8)$$

olduđuna dikkat ediniz. Yine $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ ve $\text{kov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$ olduđuna dikkat ediniz.

NOT: Örnek 2.4'te \mathbf{X} in birinci sütununun son a sayıda sütunun toplamı olduđuna yani, \mathbf{X} in sütunlarında bir lineer bağımlılıđın olduđuna dikkat ediniz. Lineer cebirdeki sonuçlardan, \mathbf{X} in tam sütun ranklı olmadığını biliyoruz. Gerçekten, \mathbf{X} in rankı $r = a$ dır, yani, sütunların $p = a + 1$ sayısından 1 eksiktir. Bu, ANOVA modellerinin yaygın bir karakteristiđidir; yani, onların \mathbf{X} matrisleri tam sütun ranka sahip deđildir. Öte yandan \mathbf{X} tam sütun ranka sahip olmak üzere, (lineer) regresyon modelleri $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ biçimindedir. Bkz. Örnek 2.2 ve Örnek 2.3.

Örnek 2.5 (İki yönlü yuvalanmış ANOVA): İki faktörlü bir deneyi göz önüne alınız. Burada B faktörü (etmeni) diyeceđimiz bir faktör, A faktörü içinde yuvalanmıştır. Başka bir deyişle B nin her seviyesi kesinlikle A faktörünün bir seviyesinde ortaya çıkar. Bu durum için bir istatistiksel model, $i = 1, 2, \dots, a$, $j = 1, 2, \dots, b_i$ ve $k = 1, 2, \dots, n_{ij}$ için,

$$Y_{ijk} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk} \quad (2.9)$$

dır. Bu modelde μ genel ortalamayı gösterir, α_i A nın i -yinci seviyesine bađlı etkiyi gösterir ve β_{ij} , A nın i -yinci seviyesi içinde yuvalanan, B nin j -yinci seviyesinin etkisini gösterir. Tüm parametreler sabit ise ve ε_{ijk} rasgele hataları sıfır ortalamalı ve sabit $\sigma^2 > 0$ bilinmeyen varyanslı ilişkisiz rasgele deđişkenler ise, bu takdirde bu özel bir $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ GM modelidir. Örneđin, $a = 3$, $b = 2$ ve $n_{ij} = 2$ için,

$$\mathbf{Y} = \begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \\ Y_{311} \\ Y_{312} \\ Y_{321} \\ Y_{322} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{22} \\ \beta_{31} \\ \beta_{32} \end{pmatrix} \quad (2.10)$$

ve $\boldsymbol{\varepsilon} = (\varepsilon_{111}, \varepsilon_{112}, \dots, \varepsilon_{322})'$ yazarız. $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ ve $\text{kov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$ olduğuna dikkat ediniz. \mathbf{X} matrisi tam sütun ranklı değildir. \mathbf{X} in rankı $r=6$ dır ve $p=10$ tane sütun vardır.

Örnek 2.6 (Etkileşimli iki yönlü çaprazlanmış ANOVA): A faktörü a sayıda seviyeye ve B faktörü b sayıda seviyeye sahip olmak üzere, iki faktörlü (A ve B) bir deneyi göz önüne alınız. Genel olarak, eğer A nın her seviyesi B nin her seviyesiyle kombinasyon halinde oluşursa, A ve B faktörlerine çaprazlanmıştır diyeceğiz. $i=1,2,\dots,a$, $j=1,2,\dots,b$ ve $k=1,2,\dots,n_{ij}$ için ε_{ij} rasgele hataları sıfır ortalamalı ve $\sigma^2 > 0$ sabit bilinmeyen varyanslı ilişkisiz rasgele değişkenler olmak üzere,

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \quad (2.11)$$

ile verilen iki faktörlü (çaprazlanmış) ANOVA modelini göz önüne alınız. Eğer tüm parametreler sabit iseler, bu özel bir $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ GM modelidir. Örneğin, $a=3$, $b=2$ ve $n_{ij}=3$ olmak üzere,

$$\mathbf{Y} = \begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{113} \\ Y_{121} \\ Y_{122} \\ Y_{123} \\ Y_{211} \\ Y_{212} \\ Y_{213} \\ Y_{221} \\ Y_{222} \\ Y_{223} \\ Y_{311} \\ Y_{312} \\ Y_{313} \\ Y_{321} \\ Y_{322} \\ Y_{323} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{31} \\ \gamma_{32} \end{pmatrix} \quad (2.12)$$

dır ve $\boldsymbol{\varepsilon} = (\varepsilon_{111}, \varepsilon_{112}, \dots, \varepsilon_{323})'$ yazarız. $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ ve $\text{kov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$ olduğuna dikkat ediniz. \mathbf{X} matrisi tam sütun ranklı değildir. \mathbf{X} in rankı $r=6$ dır ve $p=12$ tane sütun vardır.

Örnek 2.7 (İki yönlü çaprazlanmış ANOVA): A faktörü a sayıda seviyeye ve B faktörü b sayıda seviyeye sahip olmak üzere, iki faktörlü (A ve B) bir deneyi göz önüne alınız. $i=1,2,\dots,a$, $j=1,2,\dots,b$ ve $k=1,2,\dots,n_{ij}$ için ε_{ij} rasgele hataları sıfır ortalamalı ve $\sigma^2 > 0$ ortak varyanslı ilişkisiz rasgele değişkenler olmak üzere, etkileşimsiz iki yönlü çaprazlanmış model,

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \quad (2.13)$$

ile verilir. $H_0 : \gamma_{11} = \gamma_{12} = \dots = \gamma_{32} = 0$ hipotezi doğru olduğunda, etkileşimsiz modelin Örnek 2.6 daki etkileşimli modelin özel bir durumu olduğuna dikkat ediniz.

Yani; etkileşimsiz model etkileşimli modelin bir **indirgenmiş** uyarlamasıdır. Önceki gibi $a = 3$, $b = 2$ ve $n_{ij} = 3$ olmak üzere,

$$\mathbf{Y} = \begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{113} \\ Y_{121} \\ Y_{122} \\ Y_{123} \\ Y_{211} \\ Y_{212} \\ Y_{213} \\ Y_{221} \\ Y_{222} \\ Y_{223} \\ Y_{311} \\ Y_{312} \\ Y_{313} \\ Y_{321} \\ Y_{322} \\ Y_{323} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \end{pmatrix} \quad (2.14)$$

ve $\boldsymbol{\varepsilon} = (\varepsilon_{111}, \varepsilon_{112}, \dots, \varepsilon_{323})'$ yazarız. $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ ve $\text{kov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$ olduğuna dikkat ediniz.

\mathbf{X} matrisi tam sütun ranklı değildir ve $p = 6$ tane sütun vardır. Etkileşimsiz model için tasarım matrisinin, son 6 sütunun atılması hariç etkileşimli model için tasarım matrisinin aynısı olduğuna da dikkat ediniz.

Örnek 2.8 (Kovaryans analizi): Bir x orta değişkeni için etkileri ayarladıktan sonra $a \geq 2$ sayıda işlemi karşılaştırmak için bir deneyi göz önüne alınız. $i = 1, 2, \dots, a$, $j = 1, 2, \dots, n_i$ için, ε_{ij} rasgele hataları sıfır ortalamalı ve $\sigma^2 > 0$ ortak varyanslı ilişkisiz rasgele değişkenler olmak üzere, kovaryans analizi için bir model:

$$Y_{ij} = \mu + \alpha_i + \beta_j x_{ij} + \varepsilon_{ij} \quad (2.15)$$

ile verilsin. Bu modelde, μ genel ortalamayı gösterir, α_i , i -yinci işlemi almanın (orta değişkenleri dikkate almamanın) (sabit) etkisini gösterir ve β_i ; i -yinci işlem için Y yi x e bağlayan doğrunun eğimini gösterir. Bu modelin işlem eğimlerinin farklı olmasına izin verdiğine dikkat ediniz. x_{ij} lerin hatasız ölçülen sabit değerler olacakları kabul edilir.

NOT: Kovaryans analizi (ANCOVA) modeli özel bir $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ GM modelidir.

Örneğin, $a = 3$ ve $n_1 = n_2 = n_3 = 3$ olmak üzere,

$$\mathbf{Y} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & x_{11} & 0 & 0 \\ 1 & 1 & 0 & 0 & x_{12} & 0 & 0 \\ 1 & 1 & 0 & 0 & x_{13} & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & x_{21} & 0 \\ 1 & 0 & 1 & 0 & 0 & x_{22} & 0 \\ 1 & 0 & 1 & 0 & 0 & x_{23} & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & x_{31} \\ 1 & 0 & 0 & 1 & 0 & 0 & x_{32} \\ 1 & 0 & 0 & 1 & 0 & 0 & x_{33} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \end{pmatrix} \quad (2.16)$$

yazarız. $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ ve $\text{kov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$ olduğuna dikkat ediniz. \mathbf{X} matrisi tam sütun ranklı değildir. Eğer son 3 sütun arasında lineer bağımlılıklar yoksa, \mathbf{X} in rankı $r = 6$ dır ve $p = 7$ sütun vardır.

İndirgenmiş Model: Farklı eğimlere izin veren Örnek 2.8 deki ANCOVA modelini göz önüne alınız. Eğer $\beta_1 = \beta_2 = \dots = \beta_a$ ise, yani, tüm eğimler eşit ise, bu takdirde, ANCOVA modeli,

$$Y_{ij} = \mu + \alpha_i + \beta x_{ij} + \varepsilon_{ij} \quad (2.17)$$

modeline indirgenir. Yani, ortak-eğimler ANCOVA modeli farklı eğimlere izin veren modelin bir indirgenmiş uyarlamasıdır. Aynı hata yapısını kabul ederek, bu indirgeniş ANCOVA modeli de özel bir $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ GM modelidir. Önceden olduğu gibi, $a = 3$ ve $n_1 = n_2 = n_3 = 3$ olmak üzere,

$$\mathbf{Y} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & x_{11} \\ 1 & 1 & 0 & 0 & x_{12} \\ 1 & 1 & 0 & 0 & x_{13} \\ 1 & 0 & 1 & 0 & x_{21} \\ 1 & 0 & 1 & 0 & x_{22} \\ 1 & 0 & 1 & 0 & x_{23} \\ 1 & 0 & 0 & 1 & x_{31} \\ 1 & 0 & 0 & 1 & x_{32} \\ 1 & 0 & 0 & 1 & x_{33} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \end{pmatrix} \quad (2.18)$$

yazarız. x_{ij} lerden en az biri farklı olduğu sürece, \mathbf{X} in rankı $r = 4$ tür ve $p = 5$ sütun vardır.

AMAÇ: Şimdi GM modelleri olmayan $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ biçimindeki lineer modellerin örneklerini vereceğiz.

TERMİNOLOJİ: Sınıflamanın bir faktörüne; eğer o çok büyük sayıda bir seviyeye sahipse ve deneyde içerilen seviyeler mümkün seviyelerin kitlesinden bir rasgele örneklem olarak sayılabilirse, **rasgele** olacaktır denir.

Örnek 2.9 (*Bir-yönlü rasgele etkiler ANOVASI*): $i = 1, 2, \dots, a$ ve $j = 1, 2, \dots, n_i$ için, $\alpha_1, \alpha_2, \dots, \alpha_a$ işlem etkileri rasgele olarak en iyi kabul edilmek üzere, örneğin, ilgili faktörün a seviyeleri mümkün seviyelerin geniş bir kitlesinden çekilir, ε_{ij} rasgele hataları sıfır ortalama ve $\sigma^2 > 0$ ortak varyansına sahip olmak üzere,

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad (2.19)$$

modelini göz önüne alınız. Somutluk için, $a = 4$ ve $n_{ij} = 3$ olsun. $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ modeli,

$$\begin{aligned}
\mathbf{Y} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{31} \\ Y_{32} \\ Y_{33} \\ Y_{41} \\ Y_{42} \\ Y_{43} \end{pmatrix} &= \mathbf{j}_{12}\mu + \underbrace{\begin{pmatrix} \mathbf{j}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{j}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{j}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{j}_3 \end{pmatrix}}_{=\mathbf{Z}_1} \underbrace{\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix}}_{=\boldsymbol{\varepsilon}_1} + \underbrace{\begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \\ \varepsilon_{41} \\ \varepsilon_{42} \\ \varepsilon_{43} \end{pmatrix}}_{=\boldsymbol{\varepsilon}_2} \\
&= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2
\end{aligned} \tag{2.20}$$

gibidir. Burada $\mathbf{X} = \mathbf{j}_{12}$, $\boldsymbol{\beta} = \mu$ ve $\boldsymbol{\varepsilon} = \mathbf{Z}_1\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2$ olduğunu görürüz. Bu bir GM modeli değildir, çünkü, α_i ler ve ε_{ij} hataları ilişkisiz olmaları şartıyla,

$$\begin{aligned}
\text{kov}(\boldsymbol{\varepsilon}) &= \text{kov}(\mathbf{Z}_1\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2) = \mathbf{Z}_1\text{kov}(\boldsymbol{\varepsilon}_1)\mathbf{Z}_1' + \text{kov}(\boldsymbol{\varepsilon}_2) \\
&= \mathbf{Z}_1\text{kov}(\boldsymbol{\varepsilon}_1)\mathbf{Z}_1' + \sigma^2\mathbf{I}
\end{aligned} \tag{2.21}$$

dır. $\text{kov}(\boldsymbol{\varepsilon}) \neq \sigma^2\mathbf{I}$ olduğuna dikkat ediniz.

Örnek 2.10 (*İki faktörlü karma model*): A faktörü sabit ve a sayıda seviyeli ve B faktörü rasgele ve b sayıda seviyeli olmak üzere, iki faktörlü (A ve B) bir deneyi göz önüne alınız. Bu durum için bir istatistiksel model; $i = 1, 2, \dots, a$, $j = 1, 2, \dots, b$ ve $k = 1, 2, \dots, n_{ij}$ için,

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \tag{2.22}$$

ile verilir. α_i ler sabit olarak olarak en iyi ve β_j ler rasgele olarak en iyi kabul edilir.

Bu model etkileşim kabul etmez.

UYGULAMA: Rasgeleleştirilmiş bir blok deneyde, b blokları mevcut blokların büyük bir koleksiyonundan (topluluğundan-biriktirisinden) rasgele seçilmiş olabilir. Eğer amaç blokların büyük kitlesi (deneydeki b sayıda blokların olmadığı)

hakkında bir ifadeyi ortaya koymak ise, bu takdirde bloklar rasgele olarak dikkate alınır. Eğer birileri sadece a sayıda işlemle ilgilenirse, $\alpha_1, \alpha_2, \dots, \alpha_a$ işlem etkileri sabit değişmezler olarak kabul edilir.

NOT: Anlaşılması için, $a=2$, $b=4$ ve $n_{ij}=1$ olduğunu farz ediniz. Yukarıdaki modeli

$$\mathbf{Y} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{24} \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{j}_4 & \mathbf{j}_4 & \mathbf{0}_4 \\ \mathbf{j}_4 & \mathbf{0}_4 & \mathbf{j}_4 \end{pmatrix}}_{=\mathbf{X}\boldsymbol{\beta}} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix} + \underbrace{\begin{pmatrix} \mathbf{I}_4 \\ \mathbf{I}_4 \end{pmatrix}}_{=\mathbf{Z}_1\boldsymbol{\varepsilon}_1} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{24} \end{pmatrix} \quad (2.23)$$

$=\boldsymbol{\varepsilon}_2$

olarak yazabiliriz.

NOT: Eğer α_i ler rasgele olarak en iyi addedilir ise, bu takdirde,

$$\mathbf{Y} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{24} \end{pmatrix} = \mathbf{j}_8\mu + \underbrace{\begin{pmatrix} \mathbf{j}_4 & \mathbf{0}_4 \\ \mathbf{0}_4 & \mathbf{j}_4 \end{pmatrix}}_{=\mathbf{Z}_1\boldsymbol{\varepsilon}_1} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \underbrace{\begin{pmatrix} \mathbf{I}_4 \\ \mathbf{I}_4 \end{pmatrix}}_{=\mathbf{Z}_2\boldsymbol{\varepsilon}_2} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{24} \end{pmatrix} \quad (2.24)$$

$=\boldsymbol{\varepsilon}_3$

yazarız. Bu model bir rasgele etkiler veya varyans bileşenleri modeli olarak da bilinir.

Genel Biçim: Bir lineer karma model genellikle,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\boldsymbol{\varepsilon}_1 + \mathbf{Z}_2\boldsymbol{\varepsilon}_2 + \dots + \mathbf{Z}_k\boldsymbol{\varepsilon}_k \quad (2.25)$$

olarak ifade edilebilir. Burada $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_k$ lar bilinen matrislerdir (genel olarak $\mathbf{Z}_k = \mathbf{I}_k$ dır) ve $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots, \boldsymbol{\varepsilon}_k$ lar ilişkisiz bileşenlere sahip ilişkisiz rasgele vektörlerdir.

Örnek 2.11 (*Zaman serisi modelleri*): Ölçümler zaman üzerinden alındığında, gözlemler büyük olasılıkla ilişkili olduklarından, GM modeli uygun olmayabilir. $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ ve $\text{kov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{V}$, \mathbf{V} biliniyor olmak üzere, $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ biçimindeki bir lineer model daha uygun olabilir. \mathbf{V} nin genel biçimi gözlenen tepkimelerin ilişkisini modellemek için seçilir. Örneğin, $i = 1, 2, \dots, n$ için $\varepsilon_i = \rho \cdot \varepsilon_{i-1} + a_i$, $a_i \sim \mathcal{N}(0, \sigma^2)$ özdeş bağımsız dağılan ve $|\rho| < 1$ (bu bir değişmez şarttır) olmak üzere,

$$Y_i = \beta_0 + \beta_1 t + \varepsilon_i \quad (2.26)$$

istatistiksel modelini göz önüne alınız. $\{\varepsilon_t : t = 1, 2, \dots, n\}$ hata süreci 1 mertebeli bir otoregressiv modeli, yani OR(1) i izlemek üzere, buna basit bir lineer trend (meyil) modeli denir. Her t için $E(\varepsilon_t) = 0$ ve her t ve s için $\text{kov}(\varepsilon_t, \varepsilon_s) = \sigma^2 \rho^{|t-s|}$ olduğunu göstermek kolaydır. Bu nedenle, eğer $n = 5$ ise,

$$\mathbf{V} = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix} \quad (2.27)$$

dir.

Örnek 2.12 (*Rasgele katsayı modelleri*): n birey üzerinde (zaman üzerinde) t sayıda gözlemin, $i = 1, 2, \dots, n$ ve $j = 1, 2, \dots, t$ için

$$Y_{ij} = \mathbf{x}'_{ij} \boldsymbol{\beta}_i + \varepsilon_{ij} \quad (2.28)$$

modelini, yani farklı $n \times 1$ $\boldsymbol{\beta}_i$ regresyon parametresinin "konu-özel" olduklarını göz önüne alınız. Eğer bireyler bir rasgele örneklem olarak düşünülürse, bu takdirde

$\beta_1, \beta_2, \dots, \beta_n$ yi β ortalamalı ve $\Sigma_{\beta\beta}$ diyeceğimiz $n \times n$ kovaryans matrisli bağımsız ve aynı dağılan vektörler olarak ele alabiliriz. Bu modeli

$$\begin{aligned} Y_{ij} &= \mathbf{x}'_{ij} \beta_i + \varepsilon_{ij} \\ &= \underbrace{\mathbf{x}'_{ij} \beta}_{\text{sabit}} + \underbrace{\mathbf{x}'_{ij} (\beta_i - \beta)}_{\text{rasgele}} + \varepsilon_{ij} \end{aligned} \quad (2.29)$$

olarak yazabiliriz. Eğer β_i ler ε_i lerden bağımsız iseler,

$$\text{var}(Y_{ij}) = \mathbf{x}'_{ij} \Sigma_{\beta\beta} \mathbf{x}_{ij} + \sigma^2 \neq \sigma^2 \quad (2.30)$$

olduğuna dikkat ediniz.

Örnek 2.13. Ölçüm hatası modelleri. $\varepsilon_i \sim \text{öbd}\mathcal{N}(0, \sigma_\varepsilon^2)$ olmak üzere,

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (2.31)$$

istatistiksel modelini göz önüne alınız. X_i ler tam olarak gözlenmezler bunun yerine,

$U_i \sim \text{öbd}\mathcal{N}(0, \sigma_U^2)$ özdeş bağımsız dağılan olmak üzere,

$$W_i = X_i + U_i \quad (2.32)$$

olacak şekilde göz ardı edilemez hatalar ile ölçülürler. Burada,

Gözlenen veri: (Y_i, W_i)

Gözlenemeyen: $(X_i, \varepsilon_i, U_i)$

Bilinmeyen parametreler: $(\beta_0, \beta_1, \sigma_\varepsilon^2, \sigma_U^2)$

dır. Bir başvuru koordinat sistemi olarak, Y nin küçük çocuklarda akciğer fonksiyonunun sürekli bir ölçümü olduğunu ve X in NO_2 ye uzun vadeli maruz kalmasını gösterdiğini farz edelim. X in tam olarak ölçülebilmesi beklenmez. Bunun yerine klinik ziyaretinde kaydedilen NO_2 nin miktarı olan W vekilinin gözlenmiş olması daha muhtemeldir. Yukarıdaki modelin

$$\begin{aligned}
Y_i &= \beta_0 + \beta_1(W_i - U_i) + \varepsilon_i \\
&= \beta_0 + \beta_1 W_i + \underbrace{(\varepsilon_i - \beta_1 U_i)}_{\varepsilon_i^*}
\end{aligned} \tag{2.33}$$

olarak yeniden yazılabildiğine dikkat ediniz. W_i ler peşinen sabit olmadıklarından bunun bir GM lineer modeli olması için en azından $E(\varepsilon_i^* | W_i) = 0$ olmasına gereksinim olacaktır. Bununla beraber,

$$\begin{aligned}
E(\varepsilon_i^* | W_i) &= E(\varepsilon_i - \beta_1 U_i | X_i + U_i) \\
&= E(\varepsilon_i | X_i + U_i) - \beta_1 E(U_i | X_i + U_i)
\end{aligned} \tag{2.34}$$

olduğuna dikkat ediniz. Eğer ε_i hem X_i hem de U_i den bağımsız ise, birinci terim sıfırdır. U_i ve $X_i + U_i$ ilişkili olduklarından ikinci terim genel olarak sıfır değildir (şüphesiz $\beta_1 = 0$ olmadıkça). Bu nedenle bu bir GM modeli olamaz.

2.3 Lineer Modellerin Geometrisi

Miller (2015), matris cebiri ve vektörlerden yararlanarak lineer modellere geometrik açıdan yaklaştı ve konunun daha somut bir şekilde anlaşılmasını sağladı. Bu kısımda bu bilgilere yer verilerek geometrik yaklaşımlara girildi.

Lineer modeller ve geometri:

İstatistiksel lineer modellerle ilgili zengin bir geometri vardır. Bu geometriyi anlama regresyon analizi ile ilgili analizin çoğunun iç yüzünü anlamayı sağlayabilir.

- Düşünce, regresyon modelini bir vektör denklemi olarak yazmak ve vektör uzaylarının bir temel anlayışını kullanarak bu denklemin sonuçlarını araştırmaktır.
- Vektörler ve vektör uzaylarının bakış açılarını yeniden gözden geçirmek gerekir.

Vektörlerin temel öğeleri:

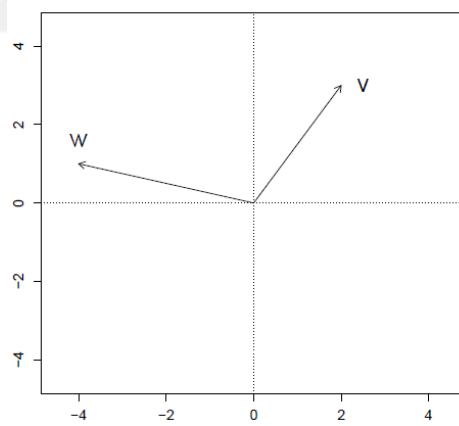
Amacımız için, bir vektör;

$$\mathbf{V} = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{bmatrix} \quad (2.35)$$

olarak göstereceğimiz, reel sayıların bir “ n -lisi ” dir. Koyu harf vektörleri (ve matrisleri göstermek için kullanılacak).

Bir vektörü n – boyutlu uzayda bir noktanın koordinatları olarak düşüneceğiz. Çoğu kez kavramları görselleştirmemize yardımcı olması için orijinden bu koordinatlara uzanan bir yönlü doğru parçasını kullanacağız.

Örnek 2.14: İki-bileşenli vektörler; $\mathbf{V} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ ve $\mathbf{W} = \begin{bmatrix} -4 \\ 1 \end{bmatrix}$ olmak üzere, standart bir dağılım grafiği üzerinde, yönlü doğru parçaları olarak gösterilebilir.



Şekil 2.1. İki-bileşenli vektörler

Vektör toplamı:

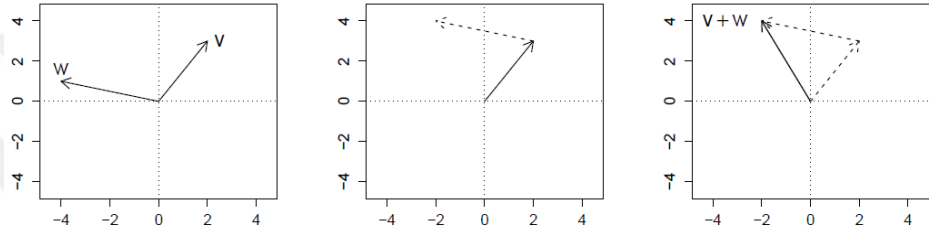
İki vektörün toplamı, aşağıda gösterildiği gibi, karşılık gelen elemanlarını toplayarak elde edilir:

$$\begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{bmatrix} + \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{bmatrix} = \begin{bmatrix} V_1 + W_1 \\ V_2 + W_2 \\ \vdots \\ V_n + W_n \end{bmatrix} \quad (2.36)$$

Örneğin; $\mathbf{V} + \mathbf{W} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} -4 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$ dür.

Vektör toplamını görselleştirme:

Görsel olarak, vektörlerden birinin başlangıç noktasını diğerinin uç noktasına taşıyoruz.



Şekil 2.2. Vektörlerin toplamı

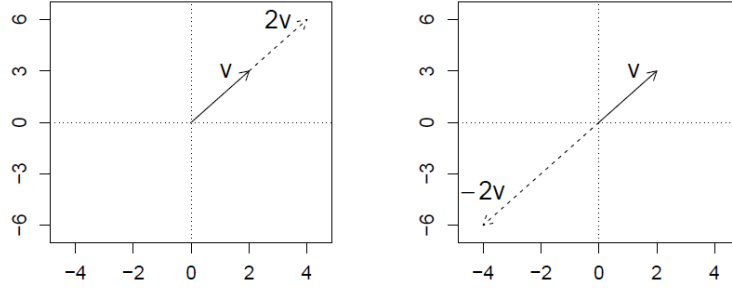
Toplam, orijinden yeni uç noktasına olan vektördür.

Vektörlerin skalerle çarpımı:

Bir vektörü bir sabitle çarpmak için herbir elemanı o sabitle çarpınız:

$$k \times \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{bmatrix} = \begin{bmatrix} k \times V_1 \\ k \times V_2 \\ \vdots \\ k \times V_n \end{bmatrix} \quad (2.37)$$

Skalerle çarpımı görselleştirme:



Şekil 2.3. Vektörlerin skalerle çarpımı

Eğer bir vektörü bir sabitle çarparsak, sonuçta oluşan vektör ilk vektörle aynı yöne sahiptir (veya eğer sabit negatif ise, zıt yöne sahiptir), fakat onun uzunluğu sabitle (sabitin mutlak değeri ile) çarpılmıştır.

Birkaç temel vektör cebiri:

\mathbf{u} , \mathbf{v} ve \mathbf{w} vektörleri ve k_1 ve k_2 skalerleri için:

1. $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$
2. $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
3. $k_1(\mathbf{v} + \mathbf{w}) = k_1\mathbf{v} + k_1\mathbf{w}$
4. $(k_1 + k_2)\mathbf{v} = k_1\mathbf{v} + k_2\mathbf{v}$

dir. Reel sayıların toplam ve çarpımına uygulanan herhangi bir cebirsel özellik büyük ölçüde vektörlerde de uygulanacak.

Lineer regresyon modeli:

Lineer model bir Y tepkime değişkeninin değerini bir veya daha fazla açıklayıcı değişkenin değerine bağlayan bir denklem olarak aşağıdaki gibi yazılabilir:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon \quad (2.38)$$

- β ların tümü değişmez sabitler fakat bilinmeyenlerdir.
- ε ; bir $\mathcal{N}(0, \sigma^2)$ dağılımına sahip olduğu kabul edilen bir rasgele değişkendir.

- Bir sonuç olarak, Y ; $\mu = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$ ortalamalı ve σ^2 varyanslı bir rasgele değişkendir.

Veri:

Tepkime için y_1 den y_n e kadar n tane gözlenen değere sahip olduğumuzu farz ediniz. i gözlemi için açıklayıcı değişkenlerin değerlerini x_{i1}, \dots, x_{ik} olarak gösteriniz ve veriyi aşağıdaki gibi bir tabloda gösteriniz:

Çizelge 2.1. Veri Çizelgesi

Gözlem	Tepkime	X_1	X_2	...	X_k
1	y_1	x_{11}	x_{12}	...	x_{1k}
2	y_2	x_{21}	x_{22}	...	x_{2k}
⋮	⋮	⋮	⋮	⋮	⋮
n	y_n	x_{n1}	x_{n2}	...	x_{nk}

Denklemlerin bir kümesi:

Her bir y_i gözlemi için

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i \quad (2.39)$$

yazabiliriz. Bu denklemlerin istiflenmesi,

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_k x_{1k} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_k x_{2k} + \varepsilon_2 \\ &\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_k x_{nk} + \varepsilon_n \end{aligned} \quad (2.40)$$

denklemler kümesini ortaya koyar.

Bir vektör denklemi olarak lineer model:

Yukarıdaki denklemler kümesi bir vektör denklemi olarak aşağıdaki gibi yazılabilir:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix} + \dots + \beta_k \begin{bmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{nk} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (2.41)$$

Vektörleri göstermek için koyu yazılmış harfleri kullanarak bu;

$$\mathbf{y} = \beta_0 \mathbf{j} + \beta_1 \mathbf{x}_1 + \dots + \beta_k \mathbf{x}_k + \boldsymbol{\varepsilon} \quad (2.42)$$

olur.

Sonda uzunluğu örneği:

12 genç hasta için sondalar kalbin içine doğru bir ana damardan beslendi. Sonda uzunluğu hastaların boy ve ağırlığının o günkü durumuyla ölçüldü. Gerekli sonda uzunluğunu boy ve ağırlığa dayandırarak tahmin etmek mümkün müdür? (Weisberg, 1980).

Sonda uzunluğu verisi:

Çizelge 2.2. Sonda uzunluğu verisi

Hasta	Boy (inc)	Ağırlık (libr)	Sonda (cm)
1	42.8	40.0	37
2	63.5	93.5	50
3	37.5	35.5	34
4	39.5	30.0	36
5	45.5	52.0	43
6	38.5	17.0	28
7	43.0	38.5	37
8	22.5	8.5	20
9	37.0	33.0	34
10	23.5	9.5	30
11	33.0	21.0	38
12	58.0	79.0	47

Sonda regresyon modeli:

Gerekli sonda uzunluğunu hastanın boy ve ağırlığına bağlayan bir

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad (2.43)$$

regresyon modelini kullanmayı araştırabiliriz.

- Y sonda uzunluğudur
- X_1 hasta boyudur
- X_2 hasta ağırlığıdır
- ε hastadan-hastaya değişkenliği gösterir.

Sonda verisi için vektör denklemi:

$$\begin{bmatrix} 37 \\ 50 \\ 34 \\ 36 \\ 43 \\ 28 \\ 37 \\ 20 \\ 34 \\ 30 \\ 38 \\ 47 \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} 42.8 \\ 63.5 \\ 37.5 \\ 39.5 \\ 45.5 \\ 38.5 \\ 43.0 \\ 22.5 \\ 37.0 \\ 23.5 \\ 33.0 \\ 58.0 \end{bmatrix} + \beta_2 \begin{bmatrix} 40.0 \\ 93.5 \\ 35.5 \\ 30.0 \\ 52.0 \\ 17.0 \\ 38.5 \\ 8.5 \\ 33.0 \\ 9.5 \\ 21.0 \\ 79.0 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \\ \varepsilon_{11} \\ \varepsilon_{12} \end{bmatrix} \quad (2.44)$$

y j x_1 x_2 ε

Sabit vektörler ve rasgele vektörler:

Lineer model iki tip vektörü içerir.

1. Sabitlerin vektörü olan sabit vektörler – Bunlar bir matematik dersinde incelediğiniz vektörlerdir.

\mathbf{j} , $\mathbf{x}_1, \dots, \mathbf{x}_k$ ların tümü sabit vektörlerdir.

2. Rasgele değişkenlerin vektörü olan rasgele vektörler. Bu nedenle bir rasgele vektör bir dağılıma sahiptir.

ε bir rasgele vektördür.

Rasgele vektörler hakkında birkaç şey:

V_1, \dots, V_p rasgele değişkenlerini içeren bir \mathbf{V} rasgele vektörü bir yoğunluk fonksiyonuna sahip bir vektör veya rasgele değişkenlerin bir koleksiyonu olarak düşünülebilir.

- \mathbf{V} için dağılım V_1, \dots, V_p nin ortak (birleşik) dağılımıyla belirlenir.
- \mathbf{V} nin beklenen değeri onun “ortalama konumunu” gösterir ve

$$E(\mathbf{V}) = E \begin{bmatrix} V_1 \\ \vdots \\ V_p \end{bmatrix} = \begin{bmatrix} E(V_1) \\ \vdots \\ E(V_p) \end{bmatrix} \quad (2.45)$$

ile verilen bir sabit vektördür.

➤ $E(\mathbf{V})$ yi göstermek için $\mu_{\mathbf{V}}$ notasyonunu kullanacağız.

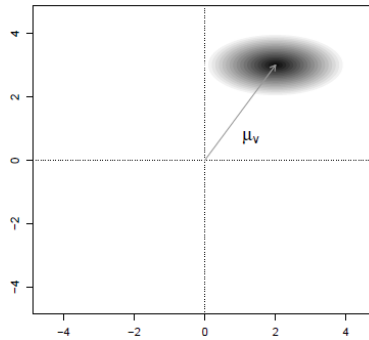
Rasgele vektörler hakkında daha fazla şey:

\mathbf{V} nin $\mu_{\mathbf{V}}$ etrafında nasıl değiştiğini özetlemek için, hem elemanların değişkenliğine hem de onların her birinin diğerine göre nasıl değiştiği göz önüne alınmalı (farklı elemanların varyansları ve eleman çiftleri arasındaki kovaryanslar düşünülmeli). Bu varyans ve kovaryansları $\Sigma_{\mathbf{V}}$ veya $\text{kov}(\mathbf{V})$ diyeceğimiz bir matris içine koymak uygundur.

$$\Sigma_{\mathbf{V}} = \begin{bmatrix} \text{var}(V_1) & \text{kov}(V_1, V_2) & \cdots & \text{kov}(V_1, V_p) \\ \text{kov}(V_2, V_1) & \text{var}(V_2) & \cdots & \text{kov}(V_2, V_p) \\ \vdots & \vdots & \ddots & \vdots \\ \text{kov}(V_p, V_1) & \text{kov}(V_p, V_2) & \cdots & \text{var}(V_p) \end{bmatrix} \quad (2.46)$$

Bir rasgele vektörün yoğunluğu:

Kavramsal olarak, bir rasgele vektör için yoğunluk fonksiyonu, bir rasgele vektör için olası uç noktaları R^n de bir bulut olarak düşünmek yararlıdır. Vektörün bulutun yoğun olmadığı yerden ziyade yoğun olduğu bir bölgede uçlanması daha olasıdır.



Şekil 2.4. R^n de bir bulut

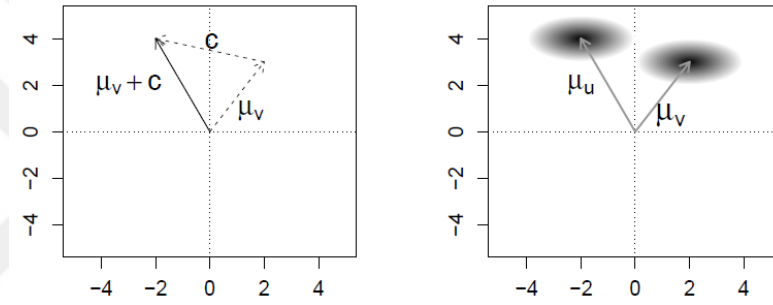
Rasgele vektörlerle çalışma:

Eğer sabit bir \mathbf{C} vektörünü bir rasgele \mathbf{V} vektörüne eklersek, sonuçta oluşan $\mathbf{U} = \mathbf{V} + \mathbf{C}$ vektörü $\boldsymbol{\mu}_U = \boldsymbol{\mu}_V + \mathbf{C}$ ve $\boldsymbol{\Sigma}_U = \boldsymbol{\Sigma}_V$ olmak üzere, bir rasgele vektördür.

Örneğin; eğer $\boldsymbol{\mu}_V = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ ve $\mathbf{C} = \begin{bmatrix} -4 \\ 1 \end{bmatrix}$ ise, bu takdirde

$$\boldsymbol{\mu}_U = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} -4 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$$

dür. Vektörlerin ortalaması değişmiştir fakat vektörün onun ortalaması etrafında nasıl değiştiği aynı kalır.



Şekil 2.5. Rasgele vektörler

Hataların dağılımı:

Lineer model hataların bağımsız $\mathcal{N}(0, \sigma^2)$ gözlemleri olduğunu kabul eder.

- ε_i lerin ortak dağılımı, $E(\varepsilon_i) = 0$, $\text{var}(\varepsilon_i) = \sigma^2$ ve her i ve $j \neq i$ için $\text{kov}(\varepsilon_i, \varepsilon_j) = 0$ olmak üzere, çok değişkenli normal dağılımdır.
- Bu nedenle ortak yoğunluk fonksiyonu

$$f(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{(\varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_n^2)}{2\sigma^2}} \quad (2.47)$$

dir.

$\boldsymbol{\varepsilon}$ un dağılımı:

$\boldsymbol{\varepsilon}$ rasgele vektörünü, ε_i ler bağımsız $\mathcal{N}(0, \sigma^2)$ rasgele değişkenleri olmak üzere,

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (2.48)$$

olarak tanımladık. $\boldsymbol{\varepsilon}$ rasgele vektörü, $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ olarak tasarlanır:

$$\boldsymbol{\mu}_{\boldsymbol{\varepsilon}} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}, \quad \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_n \quad (2.49)$$

$\boldsymbol{\varepsilon}$ için yoğunluk “bulutu”:

$\boldsymbol{\varepsilon}$ için, yoğunluk, $\|\boldsymbol{\varepsilon}\|^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = \varepsilon_1^2 + \dots + \varepsilon_n^2$ olmak üzere

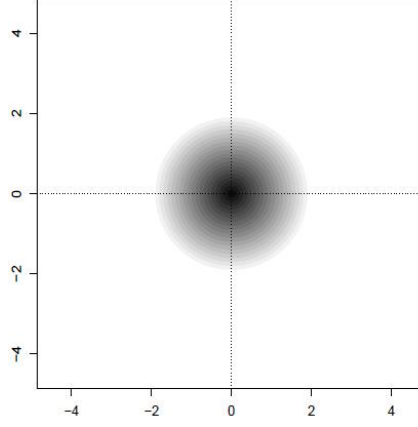
$$f(\boldsymbol{\varepsilon}) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\|\boldsymbol{\varepsilon}\|^2/2\sigma^2} \quad (2.50)$$

olarak yazılabilir.

- $\|\boldsymbol{\varepsilon}\|$, $\boldsymbol{\varepsilon}$ un uzunluğu olduğundan yoğunluk fonksiyonu $\boldsymbol{\varepsilon}$ un uzunluğuna bağlıdır, fakat onun yönüne bağlı değildir.
- $\|\boldsymbol{\varepsilon}\|^2$ artarken yoğunluk azalır.
- Kavramsal olarak, bu yoğunluk orijinde merkezlenen n – boyutlu bir fuzzy topudur.

Bir $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_2)$ vektörü için yoğunluk “bulutu”:

İki boyutlu bir $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_2)$ rasgele vektörü



Şekil 2.6. $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_2)$ rasgele vektörü için yoğunluk bulutu

gibi bir yoğunluk bulutuna sahip olmalıdır.

Tepkime vektörü sabit veya rasgele midir?

Bu aşağıdaki durumlara bağlıdır:

- Lineer modelin özellikleri hakkında konuşuyor olduğumuzda, bu durumda tepkime vektörü \mathbf{Y} olarak göstereceğimiz bir rasgele vektördür.
- Bununla beraber, özel bir veri kümesi hakkında konuştuğumuzda, bu durumda tepkime vektörü, tepkimenin \mathbf{y} ile göstereceğimiz gözlenen değerlerini içerir. Teknik olarak, \mathbf{y} ; \mathbf{Y} rasgele vektörünün özel bir gerçekleşmesini gösteren bir sabit vektördür.

\mathbf{Y} nin dağılımı:

Lineer model \mathbf{Y} yi bir sabit vektör ve bir rasgele vektörün toplamı olarak gösterir, yani, lineer model:

$$\mathbf{Y} = \underbrace{\beta_0 \mathbf{j} + \beta_1 \mathbf{x}_1 + \dots + \beta_k \mathbf{x}_k}_{\text{sabit vektör}} + \underbrace{\boldsymbol{\varepsilon}}_{\text{rasgele vektör}} \quad (2.51)$$

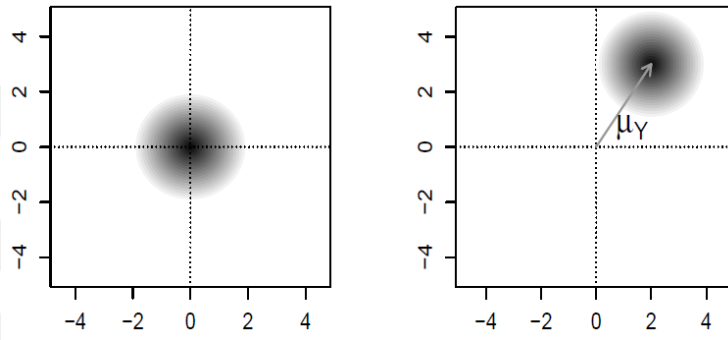
dur.

$$\begin{aligned}
\mu_Y &= \beta_0 \mathbf{j} + \beta_1 \mathbf{x}_1 + \dots + \beta_k \mathbf{x}_k + \mu_\varepsilon \\
&= \beta_0 \mathbf{j} + \beta_1 \mathbf{x}_1 + \dots + \beta_k \mathbf{x}_k \\
\Sigma_Y &= \Sigma_\varepsilon = \sigma^2 \mathbf{I}_n
\end{aligned}$$

olmak üzere, \mathbf{Y} bir rasgele vektördür.

\mathbf{Y} nin yoğunluk bulutu:

\mathbf{Y} , orijinden ziyade μ_Y etrafında merkezleşmesi hariç, ε ile aynı yoğunluğa sahiptir.



Şekil 2.7. \mathbf{Y} nin yoğunluk bulutu

Ortalama:

Lineer model μ_Y için olasılıkları $\mathbf{j}, \mathbf{x}_1, \dots, \mathbf{x}_k$ vektörlerinin lineer kombinasyonlarını olarak oluşturulabilen

$$\mu_Y = \beta_0 \mathbf{j} + \beta_1 \mathbf{x}_1 + \dots + \beta_k \mathbf{x}_k \quad (2.52)$$

vektörlerine kısıtlar.

Sözle, ortalama vektörü $\mathbf{j}, \mathbf{x}_1, \dots, \mathbf{x}_k$ nın bir lineer kombinasyonu olmalı.

Vektör uzayları:

Amaçlarımız için, sadece reel sayıları içeren vektörleri ve vektörlerin toplamı ve skalerle çarpımının alışılâ gelen tanımlarını göz önüne almamız gerekir. Bu durumda, bir vektör uzayı vektörlerin toplama ve skalerle çarpım altında kapalı olan herhangi bir koleksiyonudur.

- Eğer bir vektör uzayından iki \mathbf{u} ve \mathbf{v} vektörünü alırsak, bu takdirde bu; herhangi bir $k_1\mathbf{u} + k_2\mathbf{v}$ lineer kombinasyonunun da o vektör uzayında olması gerektiğini ifade eder.
- Sonuç itibarıyla, sıfır vektörü tüm vektör uzaylarında olmalıdır.

R^n nin tanımı:

R^n , her bir bileşenin bir reel sayı olduğu tüm n – bileşenli vektörlerin kümesi olsun.

- R^n vektör toplamı ve skalerle çarpımın alışlagelen tanımları altında bir vektör uzayıdır.
- Reel sayılar toplama ve çarpım altında kapalıdır.

R^n nin altuzayları:

R^n nin farklı altuzaylarını göz önüne almaya ihtiyacımız olacak.

- Kendisi bir vektör uzayı olan R^n deki vektörlerin herhangi bir alt kümesine R^n nin bir altuzayı denir.
- Toplama ve skalerle çarpım için önceki gibi aynı tanımları kullandığımızdan, gerçekten hepimiz vektörlerin altkümesinin toplama ve skalerle çarpım altında kapalı olduğunu kontrol etmeliyiz.

Bir vektör uzayının tabanı:

Bir altuzayı tanımlamak için alışlagelen yöntem bir taban oluşturan vektörlerin bir kümesinin belirlenmesidir.

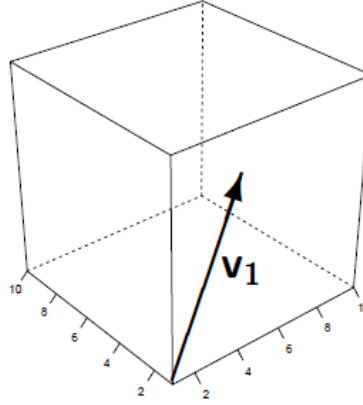
- R^n den vektörlerin herhangi bir sonlu kümesini aldığımızı ve bu vektörlerin olası tüm lineer kombinasyonları ile üretilen vektörlerin kümesini göz önüne aldığımızı farz ediniz. Bu altkümeyi üretme yöntemimiz, onun toplam ve skalerle çarpma altında kapalı olacağını böylece R^n nin bir altuzayı olacağını garanti eder.
- Daha öte, orijinal koleksiyondaki vektörlerden hiçbirinin diğer vektörlerin bir lineer kombinasyonu olarak üretilmediğini, yani, bunların hiçbirinin

gerekmediğini farz ediniz. Bu durumda vektörlerin bu koleksiyonuna onların ürettiği vektör uzayı için bir taban denir.

Bir örnek olarak R^3 :

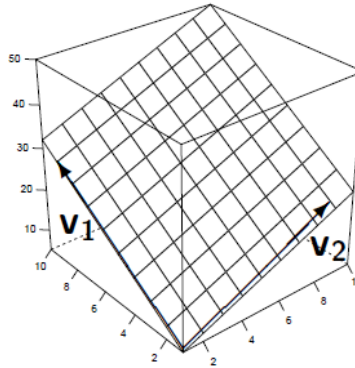
Etrafımızdaki uzayı temsil ettiğini düşünebildiğimizden R^3 yararlı bir örnektir. R^3 de bir tek \mathbf{v}_1 vektörüyle üretilen vektör uzayını; yani, \mathbf{v}_1 ve \mathbf{v}_1 in tüm skalerle çarpımlarından oluşan altuzayı gözönüne alınız.

- Bu altuzay R^3 de sonsuz bir doğru olarak düşünülebilir.



Şekil 2.8. Bir örnek olarak R^3

Şimdi R^3 de iki \mathbf{v}_1 ve \mathbf{v}_2 vektörü tarafından üretilen vektör uzayını göz önüne alınız. $\mathbf{v}_1 \neq k \times \mathbf{v}_2$ şartıyla (yani, onların eş-lineer (doğrudaş) olmamaları şartıyla), \mathbf{v}_1 ve \mathbf{v}_2 tarafından üretilen altuzay bir düzlemdir.



Şekil 2.9. \mathbf{v}_1 ve \mathbf{v}_2 vektörü tarafından üretilen vektör uzayı

R^3 ün altuzayları:

R^3 ün altuzayları aşağıdaki gibi kategorize edilebilir.

1. Orijinin kendisi.
2. Orijinden geçen herhangi bir doğru.
3. Orijinden geçen herhangi bir düzlem.
4. R^3 ün kendisi.

Bu listedeki 1 ve 4 şıkları teknik olarak R^3 ün altuzaylarıdır fakat, çok pratiksel ilgiye sahip değildir – onlara uygun olmayan altuzaylar olarak başvurulur.

Altuzayların boyutları:

Kategorimizin altuzayların boyutlarına dayandırıldığına dikkat ediniz.

- Orijin 0–boyutlu düşünülür.
- Bir tek vektörle tanımlanabildiklerinden, doğrular 1–boyutludurlar.
- Bir düzlemi tanımlamak için 2 (eş doğrultulu olmayan) vektör gerektiğinden, düzlemler 2-boyutludur.
- R^3 3-boyutludur.

Bir altuzayın bir tabanı:

Bir \mathcal{S} altuzayı için, \mathcal{S} deki her vektör V_1, \dots, V_k nın bir lineer kombinasyonu olarak ifade edilebilecek şekilde V_1, \dots, V_k vektörlerine sahip olduğumuzu farz ediniz. Bu durumda V_1, \dots, V_k ya \mathcal{S} nin **gereni** denir. Eğer onlardan herhangi birini, diğerlerinin bir lineer kombinasyonu olarak ifade etmek mümkün değilse, V_1, \dots, V_k vektörlerine **lineer bağımsızdır** denir. Eğer

- (i) V_1, \dots, V_k , \mathcal{S} yi gererse,
- (ii) V_1, \dots, V_k lineer bağımsız iseler

bu takdirde V_1, \dots, V_k vektörlerinin bir kümesi bir \mathcal{S} altuzayı için bir **tabandır**.

Bir altuzayın boyutu:

Herhangi bir \mathcal{S} altuzayı için sonsuz sayıda taban vardır. Bununla beraber, bu tabanların her biri kesinlikle aynı sayıda vektöre sahiptir. \mathcal{S} için bir tabandaki vektörlerin sayısına \mathcal{S} nin **boyutu** denir.

- Örneğin, R^3 deki bir doğru için bir taban o doğru üzerine düşen bir vektörden oluşacaktır – bu nedenle, doğrular 1 – boyutludurlar.
- R^3 deki herhangi bir düzlem için, o düzlem üstündeki iki lineer bağımsız (eş doğrusal olmayan) vektörün herhangi bir kümesi bir tabandır – bu nedenle düzlemler 2 – boyutludur.
- R^3 de 3 lineer bağımsız vektörün herhangi bir kümesi R^3 ün kendisi için bir taban olacak.

R^n ye genişletme:

R^n nin altuzayları onların boyutlarıyla kategorize edilebilir:

- Orijinin kendisi.
- Orijinden geçen herhangi bir doğru (1 – boyutlu).
- Orijinden geçen herhangi bir düzlem (2 – boyutlu).
- Orijinden geçen herhangi bir 3 – boyutlu düzlem.
- ⋮
- Orijinden geçen herhangi bir $(n-1)$ – boyutlu düzlem.
- R^n nin kendisi.

Regresyon modeline geri dönüş:

$\mu_Y = \beta_0 \mathbf{j} + \beta_1 \mathbf{x}_1 + \dots + \beta_k \mathbf{x}_k$ olmak üzere, regresyon modeli için,

$$\mathbf{Y} = \mu_Y + \varepsilon \quad (2.53)$$

yazarız. \mathbf{Y} ve açıklayıcı değişkenler arasındaki ilişki μ_Y sabit vektörü vasıtasıyla modellenir.

- μ_Y ; $\mathbf{j}, \mathbf{x}_1, \dots, \mathbf{x}_k$ vektörlerinin bir lineer kombinasyonu olduğundan, o $\mathbf{j}, \mathbf{x}_1, \dots, \mathbf{x}_k$ tarafından gerilen vektör uzayının bir elemanı olmalı. Buna model uzayı diyeceğiz.
- $\mathbf{j}, \mathbf{x}_1, \dots, \mathbf{x}_k$ nin lineer bağımsız olduklarını kabul ederek model uzayı R^n nin $k+1$ boyutlu bir altuzayıdır.

Regresyon modelinin matris formu:

Bir vektör denklemi olarak regresyon modeli:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} x_{11} \\ \vdots \\ x_{n1} \end{bmatrix} + \dots + \beta_k \begin{bmatrix} x_{1k} \\ \vdots \\ x_{nk} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (2.54)$$

Açıklayıcı değişken vektörlerini bir matris içinde bir araya getirerek bunu daha kapalı bir şekilde yazabiliriz. Böylece,

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{\boldsymbol{\varepsilon}} \quad (2.55)$$

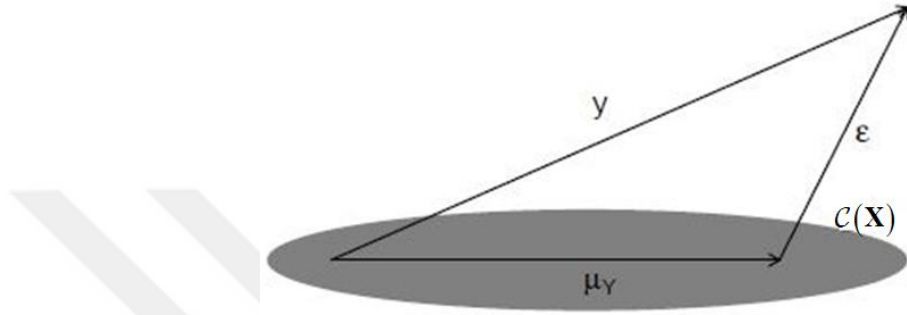
Sonda verisi için matris formu:

$$\begin{bmatrix} 37 \\ 50 \\ 34 \\ 36 \\ 43 \\ 28 \\ 37 \\ 20 \\ 34 \\ 30 \\ 38 \\ 47 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 42.8 & 40.0 \\ 1 & 63.5 & 93.5 \\ 1 & 37.5 & 35.5 \\ 1 & 39.5 & 30.0 \\ 1 & 45.5 & 52.0 \\ 1 & 38.5 & 17.0 \\ 1 & 43.0 & 38.5 \\ 1 & 22.5 & 8.5 \\ 1 & 37.0 & 33.0 \\ 1 & 23.5 & 9.5 \\ 1 & 33.0 & 21.0 \\ 1 & 58.0 & 79.0 \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \\ \varepsilon_{11} \\ \varepsilon_{12} \end{bmatrix}}_{\boldsymbol{\varepsilon}} \quad (2.56)$$

olacaktır.

Geometrik gösterim:

Böylece; $\mu_Y = \mathbf{X}\beta$ olmak üzere, $\mathbf{Y} = \mu_Y + \varepsilon$ elde ederiz. Model uzayını R^n nin \mathbf{X} in sütunları tarafından gerilen bir altuzay olarak tanımladığımızıza dikkat ediniz – bu altuzay için başka bir isim \mathbf{X} in $\mathcal{C}(\mathbf{X})$ olarak gösterilen sütun uzaylarıdır.



Şekil 2.10. \mathbf{X} in sütun uzayları

Model uydurma:

Model uydurma yöntemini iki adıma bölebiliriz:

Adım 1: μ_Y nin, \mathbf{Y} nin dağılımına dayalı bir tahminini bulunuz – bu tahmine $\hat{\mu}_Y$ diyeceğiz.

Adım 2: Regresyon için bilinmeyen parametreleri $(\beta_0, \beta_1, \dots, \beta_k)$ yı tahmin etmek için $\hat{\mu}_Y$ yı kullanınız.

Adım 1: $\hat{\mu}_Y$ yı bulma.

Regresyon modeli, μ_Y yi R^n nin açıklayıcı vektörler tarafından gerilen altuzayına kısıtlar (buna model uzayı dedik). \mathbf{Y} nin dağılımı μ_Y etrafında merkezleştğinde, o $\hat{\mu}_Y$ yı model uzayında \mathbf{Y} ye en yakın nokta olarak tanımlamayı anlamlı kılar.

Bu noktayı bulmak için, \mathbf{Y} nin model uzayı üzerine ortogonal izdüşümünü alırız.

Ortogonal izdüşüm matrisleri:

y , gözlenen tepkime vektörünün $C(\mathbf{X})$ üzerine ortogonal izdüşümünü bulmak için, y yi

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \quad (2.57)$$

ile verilen bir \mathbf{H} izdüşüm matrisiyle önden çarpabiliriz.

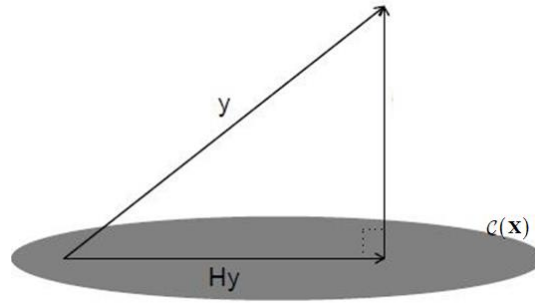
Bir izdüşüm matrisini üretmenin bu yöntemi, lineer bağımsız sütunlara sahip olan herhangi bir \mathbf{X} matrisi için çalışır.

\mathbf{Y} nin ortogonal izdüşümü:

y yi $C(\mathbf{X})$ üzerine izdüşürme, \mathbf{Y} tahmin edilen ortalama vektörümüzü (yani uygun değerleri) verir. Bu nedenle

$$\hat{\boldsymbol{\mu}}_Y = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y} \quad (2.58)$$

dir.



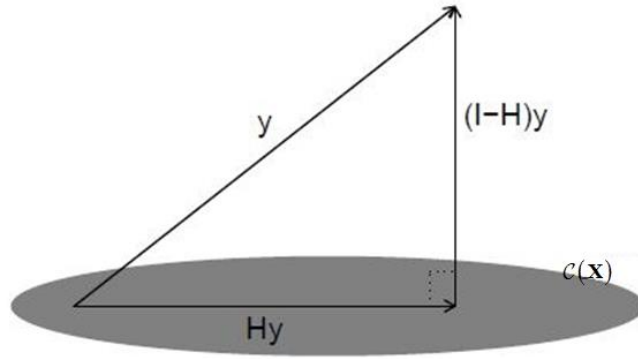
Şekil 2.11. \mathbf{Y} nin ortogonal izdüşümü

Hata (tahmin edilen hata) vektörü:

Hataların vektörü

$$\begin{aligned} \mathbf{r} &= \mathbf{y} - \hat{\boldsymbol{\mu}}_Y \\ &= \mathbf{y} - \mathbf{H}\mathbf{y} \\ &= (\mathbf{I} - \mathbf{H})\mathbf{y} \end{aligned} \quad (2.59)$$

dir.



Şekil 2.12. Tahmin edilen hata vektörü

En küçük kareler:

y nin ortogonal izdüşümü, y ve $\hat{\mu}$ arasındaki uzaklığı minimum yapar. Yukarıdaki şekilden aradaki bu uzaklığın $\|\mathbf{r}\|$ olarak göstereceğimiz \mathbf{r} hata vektörünün uzunluğuna eşit olduğu açıktır. Lineer cebirden biliyoruz ki;

$$\|\mathbf{r}\| = \sqrt{\mathbf{r}'\mathbf{r}} = \sqrt{r_1^2 + r_2^2 + \dots + r_n^2} \quad (2.60)$$

dir. Böylece $\|\mathbf{r}\|$ yi minimumlaştırmak için $\hat{\mu}$ nın seçimi, hata kareler toplamını minimumlaştırmayla aynıdır (en küçük kareler yöntemi).

Parametre tahminleri:

\mathbf{X} in sütun vektörleri lineer bağımsız olduklarından, onlar $\mathcal{C}(\mathbf{X})$ için bir taban oluşturur. Bu nedenle \mathbf{X} in sütunlarının, $\hat{\mu}$ yı üreten bir tek lineer kombinasyonu vardır. Bu bağıntı için katsayıları bir $\hat{\beta}$ vektörü içine koyarak $\hat{\mu}_Y = \mathbf{X}\hat{\beta}$ elde ederiz.

$$\hat{\mu}_Y = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \quad (2.61)$$

ile

$$\hat{\mu}_Y = \mathbf{X}\hat{\beta} \quad (2.62)$$

yı bir araya getirmek

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \quad (2.63)$$

olduğunu ortaya koyar.

Yukarıda, lineer modelleri ifade ettik ve onun bir geometrik bakış açısından tahminini tanımladık. Şimdi, burada tahmin edilen lineer modele dayanan istatistiksel sonucu çıkararak işe başlayacağız. Bazı temel matris cebirini yeniden gözden geçirmemiz gerekir.

Matris cebiri:

Uygun boyutlara sahip \mathbf{A} , \mathbf{B} ve \mathbf{C} matrisleri için,

1. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
2. $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
3. $c(\mathbf{A} + \mathbf{B}) = c\mathbf{A} + c\mathbf{B}$
4. $(c + d)\mathbf{A} = c\mathbf{A} + d\mathbf{A}$
5. $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{A}\mathbf{B} + \mathbf{A}\mathbf{C}$
6. $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{A}\mathbf{C} + \mathbf{B}\mathbf{C}$
7. $(\mathbf{A}\mathbf{B})\mathbf{C} = \mathbf{A}(\mathbf{B}\mathbf{C})$
8. $(c\mathbf{A})^{-1} = c^{-1}\mathbf{A}^{-1}$
9. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
10. Eğer \mathbf{A} ve \mathbf{B} nin her ikisi de kare matrisler ise, $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
11. $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$
12. $(c\mathbf{A})' = c\mathbf{A}'$
13. $(\mathbf{A}')' = \mathbf{A}$
14. $(\mathbf{A}\mathbf{B})' = \mathbf{B}'\mathbf{A}'$
15. $(c\mathbf{A} + d\mathbf{B})' = c\mathbf{A}' + d\mathbf{B}'$

Lineer model uydurma:

$\mu_Y = \mathbf{X}\beta$ olmak üzere lineer model: $\mathbf{Y} = \mu_Y + \varepsilon$

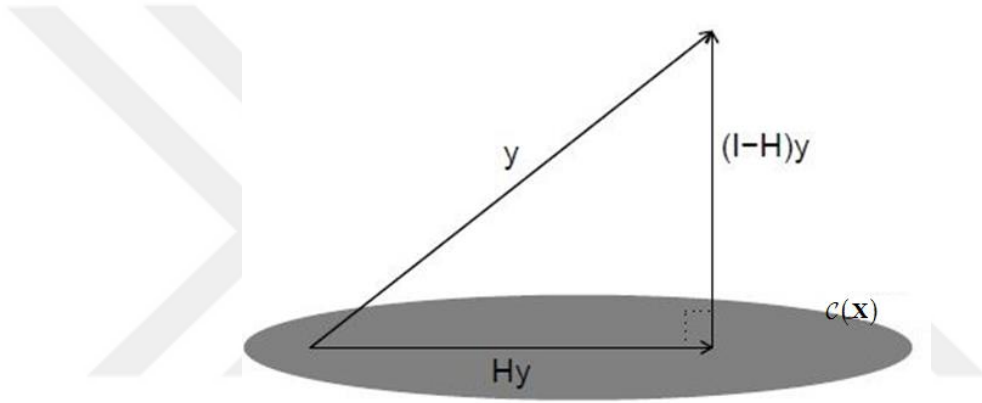
$\hat{\mu}_Y = \mathbf{X}\hat{\beta}$ olmak üzere lineer model: $\mathbf{Y} = \hat{\mu}_Y + \mathbf{r}$ dir. $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ olmak üzere,

$$\hat{\mu}_Y = \mathbf{H}\mathbf{Y} \quad (2.64)$$

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

$$\mathbf{r} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

dir.



Şekil 2.13. \mathbf{X} in sütun uzayı

$\hat{\mu}_Y$, $\hat{\beta}$ ve \mathbf{r} nin dağılımı:

$\hat{\mu}_Y$, $\hat{\beta}$ ve \mathbf{r} nin tümü \mathbf{Y} nin lineer dönüşümleridir, yani, onlar bir matrisle \mathbf{Y} nin çarpımını olarak yazılabilir:

$$\hat{\mu}_Y = [\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y} \quad (2.65)$$

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

$$\mathbf{r} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

Sonuç itibariyle, onların dağılımlarının her birini \mathbf{Y} nin dağılımından elde edebiliriz.

Rasgele vektörlerin lineer dönüşümleri:

Rasgele vektörlerin lineer dönüşümlerinin aşağıdaki örnekleri son derece yararlıdır.

\mathbf{M} ve \mathbf{C} sabit ve \mathbf{V} , $\boldsymbol{\mu}_Y$ ortalamalı ve $\boldsymbol{\Sigma}_Y$ kovaryans matrisli bir rasgele vektör olmak üzere,

$$\mathbf{U} = \mathbf{M}\mathbf{V} + \mathbf{C} \quad (2.66)$$

bağıntısını göz önüne alınız.

$$\begin{array}{c} \left[\begin{array}{c} U_1 \\ \vdots \\ U_p \end{array} \right] \\ \mathbf{U} \end{array} = \underbrace{\left[\begin{array}{cccc} m_{11} & m_{12} & \cdots & m_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ m_{p1} & m_{p2} & \cdots & m_{pk} \end{array} \right]}_{\mathbf{M}} \underbrace{\left[\begin{array}{c} V_1 \\ \vdots \\ V_k \end{array} \right]}_{\mathbf{V}} + \underbrace{\left[\begin{array}{c} c_1 \\ \vdots \\ c_p \end{array} \right]}_{\mathbf{C}}$$

yazabiliriz. Bu takdirde $\boldsymbol{\mu}_U = \mathbf{M}\boldsymbol{\mu}_V + \mathbf{C}$ ve $\boldsymbol{\Sigma}_U = \mathbf{M}\boldsymbol{\Sigma}_V + \mathbf{M}'$ dür. Ayrıca, eğer \mathbf{V} bir normal dağılıma sahipse \mathbf{U} da normal dağılıma sahiptir.

Böylece önceki bilgilerin ışığında, $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}_Y = \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_Y = \sigma^2\mathbf{I})$ verildiğinde $\hat{\boldsymbol{\mu}}_Y$, $\hat{\boldsymbol{\beta}}$ ve \mathbf{r} nin dağılımlarını elde edebiliriz. Böylece,

$$\begin{aligned} \hat{\boldsymbol{\mu}}_Y &\sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\ \hat{\boldsymbol{\beta}} &\sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}) \\ \mathbf{r} &\sim \mathcal{N}(0, \sigma^2(\mathbf{I} - \mathbf{H})) \end{aligned} \quad (2.67)$$

olur.

$\hat{\boldsymbol{\beta}}$ için çıkarım:

$\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$ i regresyon katsayıları hakkındaki çıkarım için taban olarak kullanacağız. $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ in köşegen elemanları $\hat{\boldsymbol{\beta}}_i$ lerin varyansları ve köşegen dışı elemanlar $\hat{\boldsymbol{\beta}}_i$ lerin çiftleri arasında kovaryanslardır. Bunların hepsi σ^2 ye bağlıdır. Eğer σ^2 bilinmiyor ise (hemen hemen daima durum budur) bu takdirde, ilerlemeye geçmeden önce σ^2 yi tahmin etmemiz gerekir.

σ^2 yi tahmin etme:

σ^2 yi tahmin etmenin standart yöntemi hata kareler toplamını hata serbestlik derecesine bölmektir. Böylece

$$S^2 = \hat{\sigma}^2 = \frac{\mathbf{r}'\mathbf{r}}{n-k-1} \quad (2.68)$$

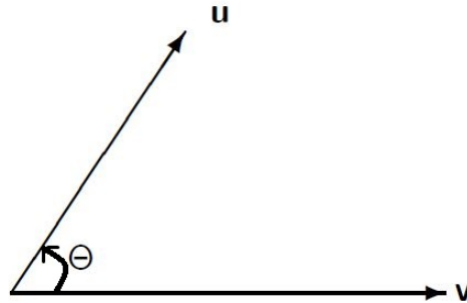
dir. Daha sonra bu tahmin edicinin kullanımını haklı çıkaracağız. Şimdilik sadece onun σ^2 nin bir yansız tahminini bize verdiğini kabul edeceğiz.

Ortogonallik:

İncelememizde, ortogonallik yeniden ortaya çıkan bir kavram olacak. Bu nedenle, vektörler ve vektör uzaylarına ilişkin olarak, ortogonallik hakkındaki ayrıntıyı kısaca ele alacağız.

İki vektör arasındaki açı:

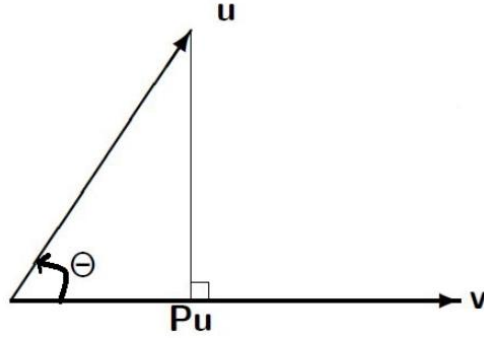
\mathbf{u} ve \mathbf{v} vektörleri arasındaki açıyı bulmak istediğimizi farz ediniz.



Şekil 2.14. İki vektör arasındaki açı

$$\cos \Theta = \frac{\mathbf{u}'\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} \quad (2.69)$$

dir. Bu sonucu elde etmek için \mathbf{u} nun \mathbf{v} üzerindeki izdüşümünü göz önüne alınız.



Şekil 2.15. u nun v üzerindeki izdüşümü

Burada $\mathbf{P} = \mathbf{v}(\mathbf{v}'\mathbf{v})^{-1}\mathbf{v}'$ dür. Tanımdan, $\cos \Theta = \frac{\|\mathbf{Pu}\|}{\|\mathbf{u}\|}$ dur.

$$\begin{aligned}
 \|\mathbf{Pu}\|^2 &= (\mathbf{Pu})' \mathbf{Pu} = \mathbf{u}' \mathbf{P}' \mathbf{Pu} \\
 &= \mathbf{u}' \left(\mathbf{v}(\mathbf{v}'\mathbf{v})^{-1}\mathbf{v}' \right)' \mathbf{v}(\mathbf{v}'\mathbf{v})^{-1}\mathbf{v}' \mathbf{u} \\
 &= \mathbf{u}' \mathbf{v}(\mathbf{v}'\mathbf{v})^{-1}\mathbf{v}' \mathbf{v}(\mathbf{v}'\mathbf{v})^{-1}\mathbf{v}' \mathbf{u} \\
 &= \mathbf{u}' \mathbf{v}(\mathbf{v}'\mathbf{v})^{-1}\mathbf{v}' \mathbf{u} \\
 &= \frac{\mathbf{u}' \mathbf{v} \mathbf{v}' \mathbf{u}}{\mathbf{v}' \mathbf{v}} = \frac{(\mathbf{u}' \mathbf{v})^2}{\|\mathbf{v}\|^2}
 \end{aligned} \tag{2.70}$$

Bu nedenle $\|\mathbf{Pu}\| = \frac{\mathbf{u}' \mathbf{v}}{\|\mathbf{v}\|}$ ve böylece $\cos \Theta = \frac{\|\mathbf{Pu}\|}{\|\mathbf{u}\|} = \frac{\mathbf{u}' \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|}$ dir.

Ortogonal durumlar:

Θ , u ve v vektörleri arasındaki açı olmak üzere, eğer onlar $\cos \Theta = 0$ olduğunu belirten bir dik açı oluşturursa, u ve v vektörleri ortogondur. Eđer bir v vektörü \mathcal{S} altuzayında her vektöre ortogondur ise, v vektörü \mathcal{S} altuzayına ortogondur. Eđer \mathcal{S}_1 altuzayındaki her vektör, \mathcal{S}_2 altuzayındaki her vektöre ortogondur ise \mathcal{S}_1 ve \mathcal{S}_2 altuzayları ortogondur.

u ve v vektörlerinin ortogondur olduklarını göstermek için; $\mathbf{u}' \mathbf{v} = 0$ olduğunu gösteriniz. Bir v vektörünün bir \mathcal{S} altuzayına ortogondur olduğunu göstermek için;

1. v nin \mathcal{S} için bir tabandaki her vektöre ortogondur olduğunu gösteriniz.

2. \mathbf{P} , \mathcal{S} için ortogonal izdüşüm matrisi olmak üzere, $\mathbf{P}\mathbf{v} = \mathbf{0}$ olduğunu gösteriniz.

\mathcal{S}_1 ve \mathcal{S}_2 altuzaylarının ortogonal olduklarını göstermek için;

1. \mathcal{S}_1 için bir tabandaki her bir vektörün, \mathcal{S}_2 için bir tabandaki her vektöre ortogonal olduğunu gösteriniz.
2. \mathbf{P}_1 ve \mathbf{P}_2 , \mathcal{S}_1 ve \mathcal{S}_2 için izdüşüm matrisleri olmak üzere, $\mathbf{P}_1\mathbf{P}_2 = \mathbf{0}$ olduğunu gösteriniz.

R^n nin bir altuzayının ortogonal tümleyeni:

\mathcal{S} , R^n nin herhangi bir altuzayı olsun. Bu takdirde, \mathcal{S} ye ortogonal olan tüm vektörlerin kümesinin kendisi de bir altuzay oluşturur. Bu altuzaya \mathcal{S} nin ortogonal tümleyeni denir ve bunu \mathcal{S}^\perp olarak göstereceğiz.

1. $\text{boy}(\mathcal{S}) + \text{boy}(\mathcal{S}^\perp) = n$
2. Eğer \mathcal{S}^\perp için bir taban ile \mathcal{S} için bir tabanı birleştirirsek R^n için bir taban elde ederiz.
3. Eğer \mathbf{P} , \mathcal{S} için izdüşüm matrisi ise bu takdirde $\mathbf{I} - \mathbf{P}$, \mathcal{S}^\perp için izdüşüm matrisidir.
4. Herhangi bir $\mathbf{u} \in R^n$ vektörü için:

$$\|\mathbf{u}\|^2 = \|\mathbf{P}\mathbf{u}\|^2 + \|(\mathbf{I} - \mathbf{P})\mathbf{u}\|^2 \quad (2.71)$$

dir.

Lineer model uydurma; \mathbf{y} yi \mathbf{X} in sütun uzayı üzerine izdüşürme olarak düşünülebilir bkz. Şekil 2.13. $\mathbf{r} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$ tahmin edilen hata vektörü, \mathbf{y} nin \mathbf{X} in sütun uzayına ortogonal olan bileşenidir. Bu nedenle, \mathbf{X} in sütun uzayının ortogonal tümleyeninin bir elemanıdır. \mathbf{X} in sütun uzayının ortogonal tümleyenine hata uzayı diyeceğiz.

Hata uzayı:

Tanıma göre, hata uzayı \mathbf{X} in sütun uzayına ortogonal olan tüm vektörleri içerir.

1. Hata uzayı $n-k-1$ boyutuna sahiptir.
2. Hata uzayı üzerine izdüşüm matrisleri $\mathbf{I} - \mathbf{H}$ dir.
3. \mathbf{r} hata vektörü, \mathbf{y} nin, hata uzayı üzerine ortogonal izdüşümüdür.

σ^2 yi tahmin etme:

σ^2 yi tahmin etmek için temeli kurmak için, $\boldsymbol{\varepsilon}$ gibi bir $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ vektörünün konuyla ilgili bazı özelliklerini ana hatlarıyla belirteceğiz. $\mathbf{V} = (V_1, \dots, V_n)'$ bir $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ rasgele vektörü olsun.

1. V_1, \dots, V_n bağımsız $\mathcal{N}(0, \sigma^2)$ rasgele değişkenleridir. Onların ortak yoğunluk fonksiyonu:

$$f(v_1, v_2, \dots, v_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(v_1^2 + v_2^2 + \dots + v_n^2)/2\sigma^2} \quad (2.72)$$

2. $V_1/\sigma, \dots, V_n/\sigma$ bağımsız $\mathcal{N}(0,1)$ rasgele değişkenleri olduğundan $\sum_i V_i^2/\sigma^2$ bir χ_n^2 dağılımına sahiptir.
3. \mathbf{V} nin $f(\mathbf{v})$ yoğunluk fonksiyonu $\|\mathbf{v}\|$ nin bir fonksiyonudur. Buna göre

$$f(\mathbf{v}) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\|\mathbf{v}\|^2/2\sigma^2} \quad (2.73)$$

dir.

4. $f(\mathbf{v})$ orijin etrafında tam olarak simetriktir ve $\|\mathbf{v}\|$ artarken $f(\mathbf{v})$ azalır.
5. $\|\mathbf{V}\|^2$, onun elamanlarının karelerinin toplamıdır.

$$\|\mathbf{V}\|^2 = \mathbf{V}'\mathbf{V} = \sum_i V_i^2 \quad (2.74)$$

dir, yani, bu nedenle $\|\mathbf{V}\|^2/\sigma^2$ bir χ_n^2 dağılımına sahiptir.

$\|\mathbf{V}\|^2$ nin dağılımı:

$\|\mathbf{V}\|^2/\sigma^2 \sim \chi_n^2$ ve $E(\chi_n^2) = n$ olduğu verilsin. Bu takdirde,

$$E\left(\frac{\|\mathbf{V}\|^2}{\sigma^2}\right) = n \rightarrow E\left(\frac{\|\mathbf{V}\|^2}{n}\right) = \hat{\sigma}^2$$

dir. Bu nedenle bir \mathbf{v} gözlenen tepkime vektörü için, $\|\mathbf{V}\|^2/n$, σ^2 nin bir yansız tahminidir.

İzdüşümlerin dağılımı:

Şimdi, $\mathbf{V} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ yı, R^n nin bir p -boyutlu altuzayı üzerine izdüşüren bir \mathbf{P} izdüşüm matrisini göz önüne alınız.

1. $f(\mathbf{v})$ yoğunluğunun simetrik doğasından dolayı, $\mathbf{P}\mathbf{V}$ nin dağılımı sadece altuzayın boyunca bağlıdır.
2. $\mathbf{P}\mathbf{V}$ izdüşümü bir p -boyutlu çok değişkenli normal dağılıma sahiptir.
3. $\|\mathbf{P}\mathbf{V}\|^2/\sigma^2$, χ_p^2 dağılımına sahiptir.

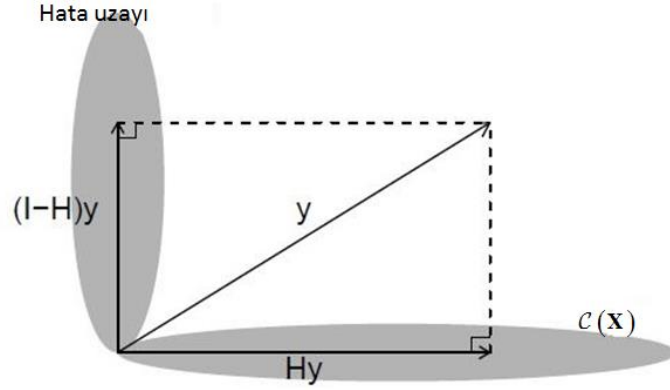
\mathbf{r} hata vektörü \mathbf{y} nin hata uzayı üzerine ortogonal izdüşümüdür. Bu $\boldsymbol{\varepsilon}$ u hata uzayı üzerine izdüşürmeye eşdeğerdir. Sonuç olarak; $\|\mathbf{r}\|^2/\sigma^2$

$$\hat{\sigma}^2 = \frac{\mathbf{r}'\mathbf{r}}{n-k-1}$$

denkleminin, σ^2 nin bir tahmini olarak kullanımını haklı çıkaran bir χ_{n-k-1}^2 dağılımına sahiptir.

Geometrik yorum:

$\hat{\mu}_Y$, y nin \mathbf{X} in sütun uzayı, yani, $\mathcal{C}(\mathbf{X})$ üzerine ortogonal izdüşümüdür ve \mathbf{r} , y nin $\mathcal{C}(\mathbf{X})^\perp$ üzerine, yani, $\mathcal{C}(\mathbf{X})$ in ortogonal tümleyeni üzerine , ortogonal izdüşümüdür.



Şekil 2.16. Hata uzayı

$\hat{\mu}_Y = \mathbf{H}\mathbf{y}$, $\mathbf{r} = (\mathbf{I} - \mathbf{H})\mathbf{y}$ dir. Pisagor bağıntısına göre,

$$\|\mathbf{y}\|^2 = \|\hat{\mu}_Y\|^2 + \|\mathbf{r}\|^2 \rightarrow \mathbf{y}'\mathbf{y} = \hat{\mu}_Y'\hat{\mu}_Y + \mathbf{r}'\mathbf{r} \quad (2.75)$$

dir. σ için tahminimiz

$$\hat{\sigma} = \sqrt{\frac{\mathbf{r}'\mathbf{r}}{n-k-1}} = \sqrt{\frac{\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}}{n-k-1}}$$

dir.

1. Bu tahmin tepkime vektörünü hata uzayı üzerine izdüşürmeye dayanır.
2. Serbestlik derecesi hata uzayının boyutunu gösterir. Tahmin edilen regresyon katsayıları ve onların standart hataları

$$\hat{\beta} = [(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}']\mathbf{y} \text{ ve } \hat{\Sigma}_{\hat{\beta}} = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}$$

formüllerinden elde edilir.

3. Her bir katsayı için $H_0 : \beta_i = 0$ biçimindeki bir sıfır hipotezi için bir t -testi de vardır.

$$t \text{ deęeri} = \frac{\hat{\beta}_i}{\text{se}(\hat{\beta}_i)} \text{ dir.}$$

F – Testleri:

\mathbf{y} nin, R^n nin farklı altuzayları üzerine izdüşümlerini karşılaştıran F testlerine dayalı birçok yararlı çıkarım yapabiliriz.

1. Sıfır hipotezi altında – dağılımın daha yüksek değerlere kaydırıldığı alternatif hipotez altında, bir F – dağılımına sahip olan bir test istatistiğine ihtiyaç duyarız.
2. Bir F – dağılımı χ^2 dağılımlarına sahip olan (ölçeklendirilmiş) iki bağımsız rasgele değişkenin oranından ortaya çıkar.

$$V_1 \sim \chi_{v_1}^2, \quad V_2 \sim \chi_{v_2}^2 \text{ olmak üzere, eęer } W = \frac{V_1/v_1}{V_2/v_2} \text{ ise, bu takdirde } W \sim F_{v_1, v_2}$$

dir. Burada V_1 ve V_2 nin bağımsız olduklarına dikkat ediniz.

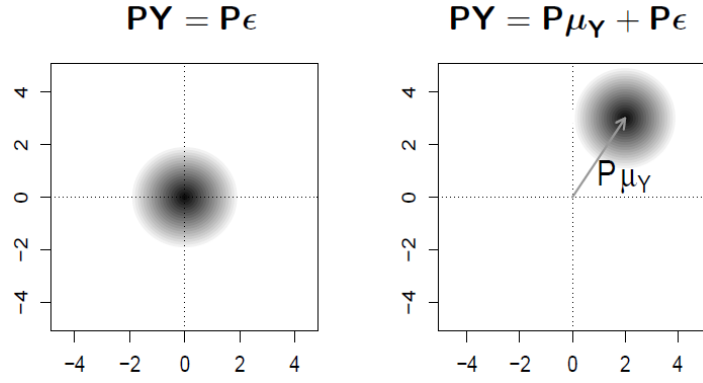
3. H_1 altında $\mathbf{P}_1\mathbf{Y}$ nin dağılımı daha büyük değerlere kaydırılmış olmalı.

olacak şekilde, $\mathbf{P}_1\mathbf{Y}$ ve $\mathbf{P}_2\mathbf{Y}$ diyeceğimiz iki izdüşümüne ihtiyacımız vardır.

Eęer \mathbf{Y} yi özel bir altuzay üzerine izdüşürmenin, $\boldsymbol{\varepsilon}$ u o alt uzay üzerine izdüşürmeye eşdeęer olduğunu önceden biliyorsak, görüntünün uzunluğunun karesi bir (ölçeklendirilmiş) χ^2 dağılımına sahiptir.

1. $\hat{\sigma}^2$ tahminimiz \mathbf{y} nin hata uzayı üzerine ortogonal izdüşümüne dayanır ve gerçekten bu izdüşümün uzunluğunun karesi bir (ölçeklendirilmiş) χ^2 dağılımına sahiptir.
2. $\boldsymbol{\mu}_Y$; hata uzayına ortogonal olduğundan, \mathbf{y} yi hata uzayı üzerine izdüşürme, $\boldsymbol{\varepsilon}$ u hata uzayı üzerine izdüşürmeye denktir.

\mathbf{Y} yi $\boldsymbol{\mu}_Y$ ye ortogonal olmayan bir altuzay üzerine izdüşürsek ne olur?



Şekil 2.17. \mathbf{Y} nin μ_Y ye ortogonal olmayan bir altuzay üzerine izdüşümü

Kavramsal olarak, izdüşümün uzunluğunun karesinin daha büyük olması gerektiğini beklediğimizi görmek kolaydır. Bu nedenle bir F – testini oluşturmak için,

1. Pay için, eğer sıfır hipotezi doğru ise, bu; ϵ u izdüşürmeye eşdeğer, fakat eğer alternatif hipotez doğru ise, ϵ u izdüşürmeye eşdeğer olmayacak şekilde, \mathbf{y} yi bir altuzay üzerine izdüşürmek isteriz.
2. Payda için, ideal olarak hem sıfır hem de alternatif hipotez altında, ϵ u izdüşürmeye eşdeğer olacak şekilde, \mathbf{y} yi bir altuzay üzerine izdüşürmek isteriz.
3. Bu iki izdüşüm bağımsız rasgele vektörler ile sonuçlanmalı.

Şimdi hangi şartlar altında \mathbf{Y} nin iki izdüşümünün bağımsız rasgele vektörleri ortaya koyacağını düşünmemiz gerekir.

Bağımsız rasgele vektörler:

$$\mathbf{V} = \begin{bmatrix} V_1 \\ \vdots \\ V_p \end{bmatrix}, \mathbf{U} = \begin{bmatrix} U_1 \\ \vdots \\ U_q \end{bmatrix}$$

rasgele vektörlerine sahip olduğumuzu farz ediniz. Eğer her bir V_i herbir U_j den bağımsız ise, \mathbf{V} ve \mathbf{U} rasgele vektörleri bağımsızdır.

Eğer her V_i ve U_j normal dağılımlara sahipse, bu takdirde bu bağımsızlık gerektirmesi; her i ve j için, $\text{kov}(V_i, U_j) = 0$ bağıntısına sahip olmayla aynıdır.

Bağımsız normal rasgele vektörler:

\mathbf{V} ve \mathbf{U} nun normal rasgele vektör olduklarını farz ediniz. \mathbf{V} ve \mathbf{U} yu tek bir \mathbf{W} vektörü içinde birleştiriniz ve \mathbf{W} için parçalanmış kovaryans matrisini göz önüne alınız. Yani, $\Sigma_{\mathbf{UV}} = \Sigma'_{\mathbf{VU}}$ olmak üzere,

$$\mathbf{W} = \begin{bmatrix} \mathbf{V} \\ \mathbf{U} \end{bmatrix}, \quad \Sigma_{\mathbf{W}} = \begin{bmatrix} \Sigma_{\mathbf{V}} & \Sigma_{\mathbf{VU}} \\ \Sigma_{\mathbf{UV}} & \Sigma_{\mathbf{U}} \end{bmatrix}$$

dur. Eğer $\Sigma_{\mathbf{VU}} = \mathbf{0}$, yani, $\Sigma_{\mathbf{UV}} = \mathbf{0}$ ise, \mathbf{V} ve \mathbf{U} bağımsızdır.

Ortogonal vektör uzaylarının anımsatmaları:

Ortogonal vektör uzayları hakkında şunları hatırlayınız:

1. Eğer \mathcal{S}_1 deki her vektör \mathcal{S}_2 deki her vektöre ortogonal ise, \mathcal{S}_1 ve \mathcal{S}_2 altuzayları ortogondur.
2. \mathbf{P}_1 ve \mathbf{P}_2 ; \mathcal{S}_1 ve \mathcal{S}_2 için izdüşüm matrisleri olmak üzere, $\mathbf{P}_1\mathbf{P}_2 = \mathbf{0}$ olduğunu göstererek, \mathcal{S}_1 ve \mathcal{S}_2 altuzaylarının ortogonal ($\mathcal{S}_1 \perp \mathcal{S}_2$) olduğunu gösterebiliriz.

İzdüşüm matrislerinin yararlı özellikleri:

Tüm ortogonal izdüşüm matrisleri matematiksel çıkarımlarda çoğu kez faydalı olan iki özelliğe sahiptir. Herhangi bir \mathbf{P} izdüşüm matrisi;

1. İdempotenttir, yani, $\mathbf{PP} = \mathbf{P}$ dir.
2. Simetriktir, yani, $\mathbf{P}' = \mathbf{P}$ dir.

Bir temel sonuç:

F – testlerinin inşası bariz bir şekilde aşağıdaki sonuca bağlı olacak:

\mathbf{V} , $\Sigma_{\mathbf{V}} = \sigma^2\mathbf{I}$ kovaryans matrisine sahip bir normal rasgele vektör olsun ve \mathcal{S}_1 ve \mathcal{S}_2 iki ortogonal altuzay ($\mathcal{S}_1 \perp \mathcal{S}_2$) olsun. Eğer \mathbf{V} yi \mathcal{S}_1 üzerine ve \mathcal{S}_2 üzerine izdüşürsek, sonuçta oluşan $\mathbf{P}_1\mathbf{V}$ ve $\mathbf{P}_2\mathbf{V}$ rasgele vektörleri bağımsız normal rasgele vektörlerdir.

Temel sonucumuzun elde edilmesi:

$\Sigma_V = \sigma^2 \mathbf{I}$ olmak üzere, $\mathbf{W} = \begin{bmatrix} \mathbf{P}_1 \mathbf{V} \\ \mathbf{P}_2 \mathbf{V} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \end{bmatrix} \mathbf{V}$ olsun. Böylece

$$\Sigma_W = \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \end{bmatrix} \sigma^2 \mathbf{I} \begin{bmatrix} \mathbf{P}_1' & \mathbf{P}_2' \end{bmatrix} = \sigma^2 \begin{bmatrix} \mathbf{P}_1 \mathbf{P}_1' & \mathbf{P}_1 \mathbf{P}_2' \\ \mathbf{P}_2 \mathbf{P}_1' & \mathbf{P}_2 \mathbf{P}_2' \end{bmatrix} \quad (2.76)$$

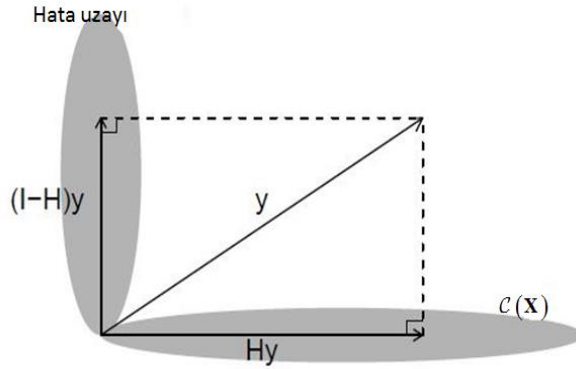
olacaktır. Sadeleştirerek,

$$\Sigma_W = \sigma^2 \begin{bmatrix} \mathbf{P}_1 \mathbf{P}_1' & \mathbf{P}_1 \mathbf{P}_2' \\ \mathbf{P}_2 \mathbf{P}_1' & \mathbf{P}_2 \mathbf{P}_2' \end{bmatrix} = \sigma^2 \begin{bmatrix} \mathbf{P}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2 \end{bmatrix} \quad (2.77)$$

olduğu görülür. Bu nedenle, $\mathbf{P}_1 \mathbf{V}$ ve $\mathbf{P}_2 \mathbf{V}$ bağımsız normal rasgele vektörlerdir.

Bir F – testi:

Aşağıdaki şekle bağlı olarak bir F – testini inşa edebiliriz.



Şekil 2.18. Hata uzayı

$\hat{\mu}_Y = \mathbf{H}\mathbf{Y}$ ve $\mathbf{r} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$ dir. \mathbf{H} , $\mathcal{C}(\mathbf{X})$ için izdüşüm matrisidir ve $(\mathbf{I} - \mathbf{H})$, $\mathcal{C}(\mathbf{X})$ in ortogonal tümleyeni üzerine izdüşüm matrisidir. Sonuç olarak, $\hat{\mu}_Y = \mathbf{H}\mathbf{Y}$ ve $\mathbf{r} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$ bağımsız rasgele vektörlerdir. Bundan önce,

$$\|\mathbf{r}\|^2 / \sigma^2 \sim \chi_{n-k-1}^2 \rightarrow \|\mathbf{r}\|^2 \sim \sigma^2 \times \chi_{n-k-1}^2$$

in, $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ nin, $n-k-1$ boyutlu bir altuzay üzerine ortogonal izdüşümünün aynısı olan \mathbf{r} ye dayandığını ileri sürdük. $\hat{\mu}_Y$; $\mathbf{Y} = \boldsymbol{\mu}_Y + \boldsymbol{\varepsilon}$ nun $k+1$ boyutlu bir

altuzay üzerine bir izdüşümdür. Eğer $\boldsymbol{\mu}_Y = \mathbf{0}$ ise, bu takdirde bu $\boldsymbol{\varepsilon}$ un bir izdüşümüyle eşit sayılır. Bu durumda;

$$\|\hat{\boldsymbol{\mu}}_Y\|^2 / \sigma^2 \sim \sigma^2 \times \chi_{k+1}^2$$

dir. Bir F – dağılımı serbestlik dereceleri ile bölünen iki bağımsız χ^2 dağılımının oranıdır. Eğer $\boldsymbol{\mu}_Y = \mathbf{0}$ ise,

$$\frac{\|\hat{\boldsymbol{\mu}}_Y\|^2 / k + 1}{\|\mathbf{r}\|^2 / (n - k - 1)} \sim F_{k+1, n-k-1}$$

dir. $\boldsymbol{\mu}_Y \neq \mathbf{0}$ ise, bu takdirde, $\|\hat{\boldsymbol{\mu}}_Y\|^2$ nin $\boldsymbol{\mu}_Y = \mathbf{0}$ varsayımından daha büyük olacağını bekleriz. Bu nedenle,

$$F = \frac{\|\hat{\boldsymbol{\mu}}_Y\|^2 / k + 1}{\|\mathbf{r}\|^2 / (n - k - 1)} \sim F_{k+1, n-k-1}$$

istatistiğini kullanarak, $H_1 : \boldsymbol{\mu}_Y \neq \mathbf{0}$ hipotezine karşı $H_0 : \boldsymbol{\mu}_Y = \mathbf{0}$ hipotezini test edebiliriz. (Burada pr - değeri = $\Pr(F_{k+1, n-k-1} \geq F \text{ ist - değeri})$ dir). Bunun $H_1 : \boldsymbol{\beta} \neq \mathbf{0}$ hipotezine (alternatif hipotezine) karşı $H_0 : \boldsymbol{\beta} = \mathbf{0}$ (sıfır hipotezi) hipotezini test etmeye eşdeğer olduğuna dikakt ediniz.

Bu nedenle, $H_0 : \boldsymbol{\mu}_Y = \mathbf{0}$ hipotezine karşı son derece güçlü delile sahibiz. $\boldsymbol{\mu}_Y = \mathbf{X}\boldsymbol{\beta}$ olduğundan bu, $H_0 : \boldsymbol{\beta} = \mathbf{0}$ hipotezini test etmeye eşdeğerdir. Fakat; bu test anlamsızdır. Aslında test etmek istediğimiz şey iki açıklayıcı değişkenin tepkimeyi tahmine yardım edip etmediğidir. Bu nedenle, c belirli olmamak üzere,

$$H_0 : \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} c \\ 0 \\ 0 \end{bmatrix}$$

hipotezini test etmek isteriz.

Ekli deęişken F – testi:

Bu bizi ekli deęişken F – testine getirir. Bizim durumumuzda, iki açıklayıcı deęişkeni içeren bir modelin, tepkimenin tahmin edilmesi işini sadece sabit terime sahip modelden daha iyi yapıp yapmadığını test etmek istiyoruz.

Geometriye dayanarak, $\mathcal{C}(\mathbf{j}) \subset \mathcal{C}(\mathbf{X})$ olmak üzere,

$$H_1 : \boldsymbol{\mu}_Y \in \mathcal{C}(\mathbf{X}) \text{ hipotezine karşı } H_0 : \boldsymbol{\mu}_Y \in \mathcal{C}(\mathbf{j})$$

hipotezini test ediyoruz. H_0 daki vektör uzayı, H_1 deki bir altuzay olmak üzere, herhangi bir durum için yöntem aynı olmalı.

Bunu yapmak için, $\hat{\boldsymbol{\mu}}_Y$ yi iki ortogonal bileşene bölmemiz gerekir: $\hat{\boldsymbol{\mu}}_Y$ nın $\mathcal{C}(\mathbf{j})$ üzerine izdüşümü ve $\mathcal{C}(\mathbf{j})$ nin $\mathcal{C}(\mathbf{X})$ deki ortogonal tümleyeni üzerine izdüşümü.

\mathbf{P}_0 ; $\mathcal{C}(\mathbf{j})$ üzerine izdüşüm matrisi olsun.

$$\mathbf{P}_0 = \mathbf{j}(\mathbf{j}'\mathbf{j})^{-1} \mathbf{j}' = \frac{1}{n} \mathbf{j}\mathbf{j}' = \begin{bmatrix} \frac{1}{n} & \dots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix} \quad (2.78)$$

dir. Bu matris üzerinde işlem yaptığı vektör için, her bir elamanı elemanların ortalaması olan bir vektörü üretir.

$\hat{\boldsymbol{\mu}}_Y$ nın $\mathcal{C}(\mathbf{j})$ üzerine izdüşümü:

$\hat{\boldsymbol{\mu}}_Y$ yi $\mathcal{C}(\mathbf{j})$ üzerine izdüşürme \mathbf{y} yi $\mathcal{C}(\mathbf{j})$ üzerine izdüşürme ile aynıdır:

$$\mathbf{P}_0 \mathbf{y} = \mathbf{P}_0 (\hat{\boldsymbol{\mu}}_Y + \mathbf{r}) = \mathbf{P}_0 \hat{\boldsymbol{\mu}}_Y + \mathbf{P}_0 \mathbf{r} = \mathbf{P}_0 \hat{\boldsymbol{\mu}}_Y \quad (2.79)$$

$\mathbf{r} \perp \mathcal{C}(\mathbf{X})$ olduğundan, $\mathbf{P}_0 \mathbf{r} = \mathbf{0}$ dır ve bu nedenle \mathbf{r} , $\mathcal{C}(\mathbf{X})$ in içerdiği her vektöre diktir.

y yi ayrıştırma:

Daha evvel \mathbf{y} yi $\hat{\boldsymbol{\mu}}_{\mathbf{y}}$ ve $\hat{\boldsymbol{\mu}}_{\mathbf{y}} \perp \mathbf{r}$ olmak üzere \mathbf{r} ye ayırdık. Yani,

$$\mathbf{y} = \hat{\boldsymbol{\mu}}_{\mathbf{y}} + \mathbf{r} = \mathbf{H}\mathbf{y} + (\mathbf{I} - \mathbf{H})\mathbf{y}$$

yazdık. Şimdi, $\hat{\boldsymbol{\mu}}_{\mathbf{y}} = \mathbf{H}\mathbf{y}$ yi daha öte $\mathcal{C}(\mathbf{j})$ üzerine ve onun $\mathcal{C}(\mathbf{X})$ deki ortogonal tümleyeni üzerine izdüşümlerine ayırırız. Böylece,

$$\mathbf{y} = \mathbf{H}\mathbf{y} + (\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{P}_0\mathbf{y} + (\mathbf{H} - \mathbf{P}_0)\mathbf{y} + (\mathbf{I} - \mathbf{H})\mathbf{y} \quad (2.80)$$

yazabiliriz.

y nin bileşenleri:

Gerçek (doğru) modelin $\mathbf{Y} = \mathbf{j}\beta_0 + \boldsymbol{\varepsilon}$ biçimine sahip olduğunu farz ediniz.

Bileşenlerin her birini göz önüne alınız:

$$\begin{aligned} \mathbf{P}_0\mathbf{Y} &= \mathbf{P}_0(\mathbf{j}\beta_0 + \boldsymbol{\varepsilon}) = \mathbf{P}_0(\mathbf{j}\beta_0) + \mathbf{P}_0\boldsymbol{\varepsilon} \\ &= \mathbf{j}\beta_0 + \boldsymbol{\varepsilon} \end{aligned}$$

$$\begin{aligned} (\mathbf{H} - \mathbf{P}_0)\mathbf{Y} &= (\mathbf{H} - \mathbf{P}_0)(\mathbf{j}\beta_0 + \boldsymbol{\varepsilon}) = (\mathbf{H} - \mathbf{P}_0)\mathbf{j}\beta_0 + (\mathbf{H} - \mathbf{P}_0)\boldsymbol{\varepsilon} \\ &= (\mathbf{j}\beta_0 - \mathbf{j}\beta_0) + (\mathbf{H} - \mathbf{P}_0)\boldsymbol{\varepsilon} \\ &= (\mathbf{H} - \mathbf{P}_0)\boldsymbol{\varepsilon} \end{aligned}$$

$$\begin{aligned} (\mathbf{I} - \mathbf{H})\mathbf{Y} &= (\mathbf{I} - \mathbf{H})(\mathbf{j}\beta_0 + \boldsymbol{\varepsilon}) = (\mathbf{I} - \mathbf{H})\mathbf{j}\beta_0 + (\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon} \\ &= (\mathbf{j}\beta_0 - \mathbf{j}\beta_0) + (\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon} \\ &= (\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon} \end{aligned}$$

olur.

Bir F – testinin unsurları:

Eğer gerçek model $\mathbf{Y} = \mathbf{j}\beta_0 + \boldsymbol{\varepsilon}$ ise, $(\mathbf{H} - \mathbf{P}_0)\mathbf{y}$ ve $(\mathbf{I} - \mathbf{H})\mathbf{y}$ nin her biri $\boldsymbol{\varepsilon}$ u bir vektör uzayı üzerine izdüşürmeye eşdeğerdir.

1. Bu nedenle, bu izdüşümlerin σ^2 ile bölünen uzunluğunun karesi bir χ^2 dağılımına sahiptir.

2. $(\mathbf{H} - \mathbf{P}_0)$, k boyutlu bir vektör uzayı üzerine izdüşürür ve $(\mathbf{I} - \mathbf{H})$, $n - k - 1$ boyutlu bir vektör uzayı üzerine izdüşürür.
3. İzdüşümler, vektör uzayları ortogonal iken bağımsızdır.

Eğer gerçek model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ ise (burada \mathbf{X} intercepttir sabit terimin yanı sıra iki açıklayıcı değişken için sütunları içerir) bu takdirde,

$\mathbf{H}(\mathbf{X}\boldsymbol{\beta})$, $\mathbf{P}_0(\mathbf{X}\boldsymbol{\beta})$ ya eşit olmadığından, $(\mathbf{H} - \mathbf{P}_0)\mathbf{y}$ artık $\boldsymbol{\varepsilon}$ u bir vektör uzayı üzerine izdüşürmeye eşdeğer değildir.

F – testimiz:

Şimdi, $H_1 : \boldsymbol{\mu}_Y = \mathbf{X}\boldsymbol{\beta}$ ya karşı $H_0 : \boldsymbol{\mu}_Y = \mathbf{j}\beta_0$ için bir F – testini ortaya koyabiliriz.

$$F\text{-istatistiği} = \frac{\|(\mathbf{H} - \mathbf{P}_0)\mathbf{Y}\|^2 / (k)}{\|(\mathbf{I} - \mathbf{H})\mathbf{Y}\|^2 / (n - k - 1)}$$

$$p\text{-değeri} = \Pr(F_{k, n-k-1} \geq F\text{-ist})$$

Bu, regresyon modelinin toplamayı tahmin etmek için sadece ortalamayı kullanmaktan daha iyi olup olmadığını test eder.

Özet:

“Ekli değişkenler F – testi” için oluş işleyiş biçimini ortaya koyduk. Uygulamamızda sıfır modeli (yalnız sabit terimli) tam modele karşı (açıklayıcıların tümü) test ettik. Modellerden biri diğerinin alt modeli olduğu sürece, yöntemi herhangi iki modeli test etmeye kolayca uyarlayabiliriz.

Bir alt modeli test etme:

Genel olarak, açıklayıcıların bir koleksiyonunun (biriktirisinin) bir altkümesini içeren bir modeli, açıklayıcıların tümünü içeren modelle karşılaştıracak olan bir F – testini inşa edebiliriz.

H_0 : alt model uygundur.

H_1 : Tam model tepkimedeki değişkenliğin daha fazlasını açıklar.

\mathbf{X}_F ve \mathbf{X}_S sırasıyla tam model ve alt model için model matrislerini gösterebilirsin. Bu durumda bu modeller için izdüşüm matrisleri

$$\mathbf{P}_F = \mathbf{X}_F (\mathbf{X}'_F \mathbf{X}_F)^{-1} \mathbf{X}'_F \quad (2.81a)$$

$$\mathbf{P}_S = \mathbf{X}_S (\mathbf{X}'_S \mathbf{X}_S)^{-1} \mathbf{X}'_S \quad (2.81b)$$

olacaktır.

Ekli değişken F – testi:

F – testini inşa etmek için \mathbf{y} tepkime vektörünü $\mathcal{C}(\mathbf{X})$ in $\mathcal{C}(\mathbf{X}_F)$ deki ortogonal tümleyeni üzerine izdüşürmesi gerekir. İzdüşüm matrisi $\mathbf{P}_F - \mathbf{P}_S$ ile verilir.

Eğer tam modeldeki fazladan değişkenlerin hiçbiri tepkimeyi tahmin etmek için ek bilgi sağlamaz ise, bu takdirde \mathbf{y} yi bu altuzaya izdüşürme tam olarak ε u izdüşürmeyle eşdeğerdir.

Bununla beraber, eğer onların bazısı ek bilgi sağlarsa, bu takdirde görüntünün uzunluğunun karesi daha büyük olmaya meyletmeli.

Tam modelin k sayıda açıklayıcıyı ve alt modelin bu açıklayıcılardan p tanesini ($p < k$) içerdiğini farz ediniz. Bu durumda;

$$F - \text{istatistiği} = \frac{\|(\mathbf{P}_F - \mathbf{P}_S) \mathbf{y}\|^2 / (k - p)}{\|(\mathbf{I} - \mathbf{P}_F) \mathbf{y}\|^2 / (n - k - 1)}$$

$$p - \text{değeri} = \Pr(F_{k-p, n-k-1} \geq F - \text{ist})$$

dir.

Eşdeğer modeller:

Aynı model uzayını tanımlayan herhangi iki model uygun değerlerin aynı kümesini üretecek. Aynı model uzayını geren yeni bir tabanı yaratmak için açıklayıcı değişken vektörlerinin lineer kombinasyonlarını alabiliriz.

Eşdeğer modeller genel sonuç:

\mathbf{A} tekil olmayan bir $(k+1) \times (k+1)$ matris olmak üzere, $\mathbf{W} = \mathbf{XA}$ gibi model matrisine sahip olan başka bir lineer modeli uydurmayı göz önüne alınız. Bu alternatif model için uygun değerlerin vektörü orijinal model için olana eşdeğer olacak, yani,

$$\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad (2.82)$$

olacak. Tahmin edilen katsayılar

$$\hat{\boldsymbol{\beta}}_{\mathbf{W}} = \mathbf{A}^{-1}\hat{\boldsymbol{\beta}}_{\mathbf{X}}$$

olarak bağlanır.

3. REGRESYON ANALİZİ

Wood (2011) regresyon ve çoklu regresyon hakkındaki bilgileri sade bir biçimde sunar. Bu bölümde zaman zaman sözünü edeceğimiz bu kavramlar onun çalışmasının ışığı altında verildi.

3.1 Çoklu Regresyon

Regresyon tahminini yeniden gözden geçirme:

Eğer $\mathbf{X}'\mathbf{X}$ in tersi mevcutsa,

$$(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X} = \mathbf{I} \quad (3.1)$$

olduğundan,

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y} \quad (3.2)$$

denklemini aşağıdaki gibi çözeriz,

$$(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \quad (3.3)$$

ve

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \quad (3.4)$$

elde ederiz.

En küçük kareler çözümü:

Matris normal denklemleri

$$\mathbf{Q} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (3.5)$$

nın $\boldsymbol{\beta}$ ya göre minimumlaştırılmasından doğrudan doğruya çıkarılabilir.

Uygun deęerler ve hata tahminleri:

Uygun deęerlerin vektörü, matris formunda

$$\hat{\mathbf{y}} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix} \quad (3.6)$$

olsun. Bu durumda $\hat{\mathbf{y}} = \mathbf{X}\mathbf{b}$ yazarız.

Şapka matrisi - y üzerine şapka koyar:

Uygun deęerleri \mathbf{X} ve \mathbf{y} göre doğrudan doğruya ifade edebiliriz. Böylece

$$\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \quad (3.7)$$

yazarız ve ayrıca \mathbf{H} “şapka matrisini” tanımlarız. $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$ olmak üzere

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y} \quad (3.8)$$

elde ederiz. Şapka matrisleri regresyon analizi için teşhislerde önemli bir rol oynar.

Şapka matrisinin özellikleri:

1. Şapka matrisi simetriktir.
2. Şapka matrisi idempotenttir, yani, $\mathbf{H}\mathbf{H} = \mathbf{H}$ dir.

İdempotent matrisin önemli bir özellięi:

Bir simetrik ve idempotent \mathbf{A} matrisi için, $\text{rank}(\mathbf{A}) = \text{iz}(\mathbf{A})$, yani, \mathbf{A} nın sıfırdan farklı özdeęerlerinin sayısıdır.

Hata tahminleri:

Uygun \hat{y} deęerleri gibi hata tahminleri Y_i tepkime deęişkeni gözlemlerinin bir lineer kombinasyonu olarak ifade edilebilir. Buna göre,

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y} \quad (3.9)$$

aynı zamanda,

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\mathbf{b} \quad (3.10)$$

olduğunu da hatırlayalım. Bunlar eşdeğerdirler.

Hata tahminlerinin kovaryansı:

$$\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{y} \quad (3.11)$$

ile başlayarak,

$$\sigma^2 \{\mathbf{e}\} = (\mathbf{I} - \mathbf{H})\sigma^2 \{\mathbf{y}\}(\mathbf{I} - \mathbf{H})' \quad (3.12)$$

olduğunu görürüz, fakat

$$\sigma^2 \{\mathbf{y}\} = \sigma^2 (\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I} \quad (3.13)$$

olması,

$$\sigma^2 \{\mathbf{e}\} = \sigma^2 (\mathbf{I} - \mathbf{H})\mathbf{I}(\mathbf{I} - \mathbf{H}) = \sigma^2 (\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H}) \quad (3.14)$$

olduğunu ifade eder ve $\mathbf{I} - \mathbf{H}$ idempotent olduğundan,

$$\sigma^2 \{\mathbf{e}\} = \sigma^2 (\mathbf{I} - \mathbf{H}) \quad (3.15)$$

elde ederiz.

ANOVA (Varyans analizi):

$$\mathbf{y}'\mathbf{y} = \sum Y_i^2, \quad \frac{(\sum Y_i)^2}{n} = \frac{1}{n} \mathbf{y}'\mathbf{J}\mathbf{y}, \quad \mathbf{j} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \text{ ve } \mathbf{J} = \mathbf{j}\mathbf{j}' \text{ olmak üzere,}$$

$$\text{GKT} = \mathbf{y}'\mathbf{y} - \frac{1}{n} \mathbf{y}'\mathbf{J}\mathbf{y} \quad (3.16)$$

ye giden,

$$\text{GKT} = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - \frac{\sum (Y_i)^2}{n} \quad (3.17)$$

ile başlayarak, ayrıca ANOVA y1 ifade edebiliriz.

Hata kareler toplamı:

$$\text{HKT} = \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 \quad (3.18)$$

olduğunu hatırlayınız. Matris formunda bu

$$\begin{aligned} \text{HKT} &= \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{X}\mathbf{b})' (\mathbf{y} - \mathbf{X}\mathbf{b}) \\ &= \mathbf{y}'\mathbf{y} - 2\mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} \\ &= \mathbf{y}'\mathbf{y} - 2\mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{y}'\mathbf{y} - 2\mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{I}\mathbf{X}'\mathbf{y} \end{aligned} \quad (3.19)$$

dir. Bu sadeleştirildiğinde, $\text{HKT} = \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y}$ olur veya $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ olduğunu hatırlama,

$$\begin{aligned} \text{HKT} &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y} \\ &= \mathbf{y}'[\mathbf{I} - \mathbf{H}]\mathbf{y} \end{aligned} \quad (3.20)$$

olduğunu ortaya koyar.

RKT:

$$\text{GKT} = \mathbf{y}'\mathbf{y} - \frac{1}{n}\mathbf{y}'\mathbf{J}\mathbf{y} \quad \text{ve} \quad \text{HKT} = \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y}$$

olmak üzere,

$$\text{RKT} = \text{GKT} - \text{HKT} \quad (3.21)$$

olduğunu biliyoruz. Buradan,

$$\text{RKT} = \mathbf{b}'\mathbf{X}'\mathbf{y} - \frac{1}{n}\mathbf{y}'\mathbf{J}\mathbf{y} \quad (3.22)$$

elde ederiz ve önceden olduğu gibi \mathbf{b} yi yerine koyarak,

$$\text{RKT} = \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \frac{1}{n}\mathbf{y}'\mathbf{J}\mathbf{y} \quad (3.23)$$

olduğunu ortaya koyarız.

Kuadratik formlar:

- ANOVA kareler toplamları kuadratik formlar olarak yorumlanabilir.

Kuadratik formun bir örneği

$$5Y_1^2 + 6Y_1Y_2 + 4Y_2^2$$

ile verilir.

- Bunun matris notasyonu ile \mathbf{A} daima (bir kuadratik form durumunda) bir simetrik matris olmak üzere

$$\begin{pmatrix} Y_1 & Y_2 \end{pmatrix} \begin{pmatrix} 5 & 3 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \mathbf{y}'\mathbf{A}\mathbf{y}$$

olarak ifade edilebildiğine dikkat ediniz.

- Çarpım asosyatif olduğundan köşegen dışı terimler çapraz-çarpımın katsayısının yarısına eşit olmalı.
- Genel olarak, \mathbf{A} kuadratik formun matrisi olmak üzere; bir kuadratik form

$$\mathbf{y}'\mathbf{A}\mathbf{y} = \sum_i \sum_j a_{ij} Y_i Y_j \quad (3.24)$$

ile tanımlanır. Burada $a_{ij} = a_{ji}$ dir.

- GKT, HKT ve RKT ANOVA toplamlarının tümü kuadratik formlar biçiminde düzenlenebilir. Buna göre,

$$\text{GKT} = \mathbf{y}' \left(\mathbf{I} - \frac{1}{n} \mathbf{J} \right) \mathbf{y}$$

$$\text{HKT} = \mathbf{y}' (\mathbf{I} - \mathbf{H}) \mathbf{y} \quad (3.25)$$

$$\text{RKT} = \mathbf{y}' \left(\mathbf{H} - \frac{1}{n} \mathbf{J} \right) \mathbf{y}$$

dir.

Testler ve sonuç çıkarma:

- ANOVA testleri ve sonuç çıkarmalar önceki gibi aynı yürütülebilir.
- Sadece nicelikleri elde etmenin cebirsel yöntemi değişir.
- Matris notasyonu bir kısa kesme yazmadır, bir hesapsal kısaltma değildir.

Sonuç çıkarma:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \mathbf{A}\mathbf{y} \quad (3.26)$$

olduğunu hatırlatarak β vektör tahmin edicisinin örneklem varyansını elde edebiliriz.

Burada \mathbf{A} bir sabit

$$\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}', \quad \mathbf{A}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \quad (3.27)$$

matrisidir. Standart matris kovaryans operatörünü kullanarak,

$$\sigma^2 \{\mathbf{b}\} = \mathbf{A} \sigma^2 \{\mathbf{y}\} \mathbf{A}' \quad (3.28)$$

olduğunu görürüz.

b nin varyansı:

$\sigma^2 \{\mathbf{y}\} = \sigma^2 \mathbf{I}$ olduğundan,

$$\begin{aligned} \sigma^2 \{\mathbf{b}\} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \sigma^2 \mathbf{I} \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{I} \\ &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \end{aligned} \quad (3.29)$$

yazabiliriz. Şüphesiz

$$\begin{aligned} E(\mathbf{b}) &= E\left((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}\right) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' E(\mathbf{y}) \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}\beta = \beta \end{aligned} \quad (3.30)$$

dir. Şüphesiz bu, σ^2 nin bilindiğini kabul eder. Eğer bilinmezse, genel olarak onun yerine HKO (hata kareler ortalamasını) koyunuz.

$$\sigma^2 \{\mathbf{b}\} = \begin{pmatrix} \frac{\sigma^2}{n} + \frac{\sigma^2 \bar{X}^2}{\sum (X_i - \bar{X})^2} & \frac{-\bar{X} \sigma^2}{\sum (X_i - \bar{X})^2} \\ \frac{-\bar{X} \sigma^2}{\sum (X_i - \bar{X})^2} & \frac{\sigma^2}{\sum (X_i - \bar{X})^2} \end{pmatrix} \quad (3.31)$$

$$s^2 \{b\} = \text{HKO}(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} \frac{\text{HKO}}{n} + \frac{\bar{X}^2 \text{HKO}}{\sum (X_i - \bar{X})^2} & \frac{-\bar{X} \text{HKO}}{\sum (X_i - \bar{X})^2} \\ \frac{-\bar{X} \text{HKO}}{\sum (X_i - \bar{X})^2} & \frac{\text{HKO}}{\sum (X_i - \bar{X})^2} \end{pmatrix} \quad (3.32)$$

Ortalama tepkime:

- Ortalama tepkimeyi tahmin etmek için aşağıdaki matrisi ortaya koyabiliriz:

$$X_h = (\mathbf{j} \quad X_h)$$

- Bu durumda tahmin $\hat{Y}_h = X_h \mathbf{b}$ bir, yani,

$$\hat{Y}_h = X_h' \mathbf{b} = (\mathbf{j} \quad X_h) \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = (b_0 + b_1 X_h)$$

dir.

Ortalama tepkimenin varyansı:

-

$$\sigma^2 \{\hat{Y}_h\} = \sigma^2 X_h' (\mathbf{X}'\mathbf{X})^{-1} X_h \quad (3.33)$$

ile verilir ve β nın varyansı ile aynı yolla ulaşılır.

- Benzer şekilde matris notasyonu ile tahmin edilen varyans

$$s^2 \{\hat{Y}_h\} = \text{HKO} \left(X_h' (\mathbf{X}'\mathbf{X})^{-1} X_h \right) \quad (3.34)$$

ile verilir.

Sonuç:

- Rasgele vektör ve matrislerin beklenen değeri ve varyansı
- Matris formunda basit lineer regresyon
- Sonra: Çoklu regresyon

Çoklu regresyon:

- İstatistiksel analizde en yaygın olarak kullanılanlardan biridir.
- Çoklu regresyon için matris ifadeleri basit lineer regresyon için olanın aynısıdır.

Çoğu kez tepkime çoklu girdi niceliklerinin bir fonksiyonu olarak anlaşılır.

İki açıklayıcı değişken ile birinci-mertebe:

- İki X_1 ve X_2 açıklayıcı değişkeni var olduğunda

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \quad (3.35)$$

regresyon modeline iki açıklayıcı değişkenli birinci mertebeden bir model denir.

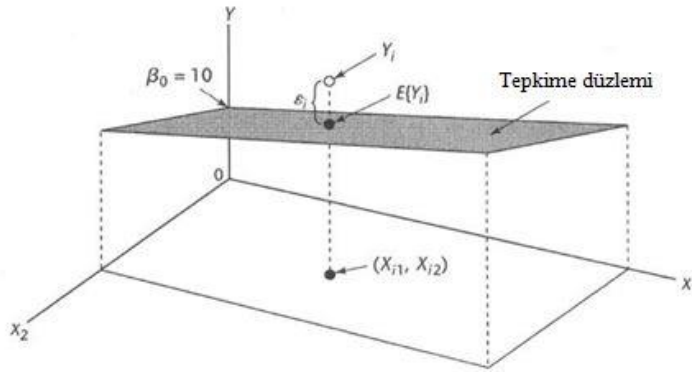
- Bir birinci mertebeden model açıklayıcı değişkenlere göre lineerdir.
- X_{i1} ve X_{i2} ; iki açıklayıcı değişkenin i – yinci denemedeği değeridir.

Regresyon yüzeyinin fonksiyonel biçimi:

- $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ beklenen değerinde hatayı sıfır kabul ederek
- Bu regresyon fonksiyonunun formu bir düzlemdir.

Örneğin, $E(Y) = 10 + 2X_1 + 5X_2$

Gri örnek;



Şekil 3.1. Regresyon yüzeyi

Regresyon katsayılarının anlamı:

- X_1 ve X_2 sıfır olduğunda, β_0 sabit terim (intercept) dir.
- β_1 ; X_2 sabit tutulduğunda, X_1 de birim artış başına $E(Y)$ ortalama tepkimesindeki değişimi gösterir.
- β_2 tam tersini gösterir.
- Örnek: $X_2 = 2$ alınız.

$$E(Y) = 10 + 2X_1 + 5(2) = 20 + 2X_1, \quad X_2 = 2$$

intercepti değiştirir fakat lineerlik açıktır.

- Başka bir deyişle, regresyon yüzeyinin tüm bir boyutlu kısıtlamaları doğrudur.

Terminoloji:

1. X_1 in ortalama tepkime üzerindeki etkisi X_2 nin düzeyine bağlı değildir (ve tersi de). İki açıklayıcı değişkene, toplamsal etkilere sahiptir veya etkileşimsizdir denir.
2. β_1 ve β_2 parametrelerine bazen kısmi regresyon katsayıları denir.

Yorumlar:

1. Bir düzlemsel tepkime yüzeyi daima uygun olmayabilir, fakat, olmadığında bile o, ekseriyeye girdi uzayının “yerel” bölgelerindeki regresyon fonksiyonunun iyi bir yaklaşım açıklayıcısıdır.
2. Parametrelerin anlamı herbirine göre regresyon fonksiyonunun kısımlerini (parçalarını) olarak belirlenebilir.

2 den fazla açıklayıcı değişkene sahip birinci mertebeden model:

$p - 1$ sayıda açıklayıcı değişken var olsun. Bu durumda

$$Y_i = \beta_0 + \sum_{k=1}^{p-1} \beta_k X_{ik} + \varepsilon_i \quad (3.36)$$

olarak ve eğer $X_{i0}=1$ ise; $X_{i0}=1$ olmak üzere, $Y_i = \sum_{k=1}^{p-1} \beta_k X_{ik} + \varepsilon_i$ olarak da yazılabilen

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad (3.37)$$

dir.

Tepkime yüzeyinin geometrisi:

- Bu kurgulamada tepkime yüzeyi bir hiperdüzlemdir.
- Görselleştirmek için bu zordur fakat aynı sezgiler gerçekleşir.
- Girdi değişkenlerinin biri hariç hepsini sabitleyerek her bir β_p , tepkime değişkeninin bir girdi değişkenine göre ne kadar artacağını veya azalacağını söyler.

Genel lineer regresyon modeli:

Genel regresyon modeline ulaştık. Genel olarak, regresyon modelinde X_1, \dots, X_{p-1} değişkenleri farklı açıklayıcı değişkenleri göstermemeli, ve onların tümü sayısal (sürekli) olmamalıdır. Genel model, $E(\varepsilon_i) = 0$ olduğunda, tepkime fonksiyonu

$$E(Y) = \beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1} \quad (3.38)$$

dir.

3.2 Regresyon Analizinde Singüler Değer Ayrışımının Kullanımı

Matrisler hakkında singüler (tekil)-değer ayrışımı lineer cebirde önemli bir yer tutar. Mandel (1982) regresyon analizinde singüler-değer ayrışımının kullanımını örnekleriyle beraber ortaya koydu. Bu kısımda, bu bilgilere yer verilmiştir.

Model:

Modelin bilindiğini ve

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (3.39)$$

biçimine sahip olduğunu kabul ederiz. Burada \mathbf{Y} ve \mathbf{e} her biri N elemanlı vektörlerdir, \mathbf{X} ; x_{ij} elemanlarının bir $N \times p$ matrisidir ve $\boldsymbol{\beta}$, p elemanın bir

vektörüdür. Stokastik olmayan elamanlardan oluşan \mathbf{X} matrisi verilir. \mathbf{Y} vektörü her biri iki terimin toplamı, yani,

$$E(y_i) = \sum_j x_{ij} \beta_j \quad (3.40)$$

beklenen değeri ve e_i hata terimlerinin toplamı olan y_i ölçümlerinden oluşur. e_i hataların sıfır ortalaması ve değeri bilinmeyen σ^2 sabit varyansı ile ilişkisiz oldukları kabul edilir. e_i nin vektörü (3.39) da \mathbf{e} terimiyle temsil edilir. Bu çalışmadaki genel düşünceler Çizelge 3.1 de gösterilen yapay veri için açıklanacak. Bu veride $N=8$ ve $p=3$ tür. Bu durumda her i için ilki 1 e eşit olan üç açıklayıcı değişken, yani, x_1 , x_2 ve x_3 vardır. Regresyon denklemi

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i \quad (3.41)$$

dir fakat $x_{i1} = 1$ olduğundan, bir β_1 “bağımsız terimi” ile

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i \quad (3.42)$$

olur. Böyle bir bağımsız terimin dahil edilmesi regresyon çalışmasında yaygın bir uygulamadır. Onun yararı, biri; lineer ilişki ile ifade edilebilen fakat orantılı olmayan birimler olan açıklayıcıları göz önüne aldığında ortaya çıkar. Örneğin, eğer x_2 açıklayıcısı sıcaklık ise, Celsiustan Fahrenheit birimlerine bir dönüşüm, bir bağımsız terime izin verilmediği takdirde kabul edilen model içinde mümkün olacaktı.

Regresyon analizinin bir çok uygulayıcısı regresyon analizinden önce, bağımsız terim hariç tüm açıklayıcılar üzerinde bir standartlaştırma yapar. x_j açıklayıcısının standartlaştırılması regresyon denkleminde onun yerine

$$x_j = \bar{x}_j + s_j t_j \quad (3.43)$$

koymaktan ibarettir. Burada \bar{x}_j ortalamadır ve s_j ; x_j sütunundaki x_{ij} elemanlarının standart sapmasıdır. Şimdi regresyon y nin t_j açıklayıcıları üzerindedir ve sonraki her bir j için $\bar{t}_j = 0$ (merkezleme) ve $s_{t_j} = 1$ (ölçekleme) olacak şekildedir. Merkezleme ve ölçeklemenin kullanımları ve yararları Draper ve Smith (1981) de

bazı ayrıntıları ile ele alınır. Sunumun basitliği (sadeligi) için bu adımı başından sonuna kadar es geçeceğiz.

Regresyon analizinin amacı σ^2 nin yanı sıra β_j ($j=1, \dots, p$) katsayılarını tahmin etmek, $\mathbf{x} = (x_1, x_2, \dots, x_p)$ regressörlerinin (açıklayıcılarının) herhangi bir “gelecek” vektörü için y nin değerini “öngörmek” ve böyle bir öngörülen değerin, \hat{y} diyeceğimiz hatasını tahmin etmektir. Karışıklıktan kaçınmak için, bulunacak böyle bir y -değeri için, (x_1, x_2, \dots, x_p) değerlerinin bir kümesine; bundan böyle “ X uzayında bir nokta” veya sadece bir vektörden ziyade bir ”nokta” olarak başvurulur.

Regresyon probleminin başka görünümleri incelememiz sırasında ortaya çıkar.

Çizelge 3.1. A veri kümesi

Nokta	x_1	x_2	x_3	y
1	1	16.85	1.46	41.38
2	1	24.81	- 4.61	31.01
3	1	18.85	-.21	37.41
4	1	12.63	4.93	50.05
5	1	21.38	- 1.36	39.17
6	1	18.78	- .08	38.86
7	1	15.58	2.98	46.14
8	1	16.30	1.73	44.47

Regresyonun geometrik gösterimi:

Sadece iki x_1 ve x_2 açıklayıcı değişkeninin var olduğu bir durumu kabul ediniz. Bu durumda (x_1, x_2) “tasarım noktaları” Şekil 3.2 de gösterildiği gibi bir D düzleminde yerleştirilebilir. Tasarım noktalarının her birinde bir doğru parçası, y ; (x_1, x_2) noktasında “tepkime” değişkeninin değeri olmak üzere, D düzlemine dik bir doğrultuda ve y yüksekliğine sahip olacak şekilde ikame edilir. (3.39) model denkleminde göre, bu doğru parçalarının uç noktaları bir düzleme yakın olmalı. Eğer y tepkime değişkeni tamamen deneysel hatadan bağımsız olsa idi, onlar kesinlikle bir düzlemde olacaktı.

P (gerçek) tepkime düzlemini belirtsin. y deki hatalardan dolayı, P tam olarak belirlenemez, fakat ona Şekil (3.2) de gösterildiği gibi, P_f diyeceğimiz, uygun bir düzlemle yaklaşılabılır. p sayıda açıklayıcı değişkenli daha genel durumda D düzlemi p -boyutlu bir hiper düzlem olur ve P ve P_f de öyle olur.

X in singüler (tekil) değer ayrışımı:

Bir \mathbf{X} $N \times p$ matrisi verildiğinde \mathbf{X} in her bir x_{ij} elemanını aşağıdaki biçimde ifade etmek mümkündür:

$$x_{ij} = \theta_1 u_{1i} v_{1j} + \theta_2 u_{2i} v_{2j} + \dots + \theta_r u_{ri} v_{rj} \quad (3.44)$$

veya $\theta_1 \geq \theta_2 \geq \dots \geq \theta_r$ olmak üzere, daha özlü bir biçimde

$$x_{ij} = \sum_{k=1}^r \theta_k u_{ki} v_{kj} \quad (3.45)$$

bağıntısıyla ifade etmek mümkündür. Bu; \mathbf{X} matrisinin singüler değer ayrışımı (SDA) olarak bilinir. (3.44) deki terimlerin sayısı r dir, yani, \mathbf{X} matrisinin rankıdır; r , N veya p yi aşamaz, yani, onlardan daha küçük olanını geçemez.

Daima $N \geq p$ olduğunu kabul edeceğiz ve $r \leq p$ olduğu görülür. r bileşenli \mathbf{u} ve r bileşenli \mathbf{v} vektörlerinin her birisi diğerine ortogondur. Bundan başka, bu vektörlerin her biri birim uzunluğa sahiptir, bu nedenle,

$$\text{her } k \text{ için, } \sum_i u_{ki}^2 = \sum_j v_{kj}^2 = 1 \quad (3.46)$$

dir. Matris notasyonu ile,

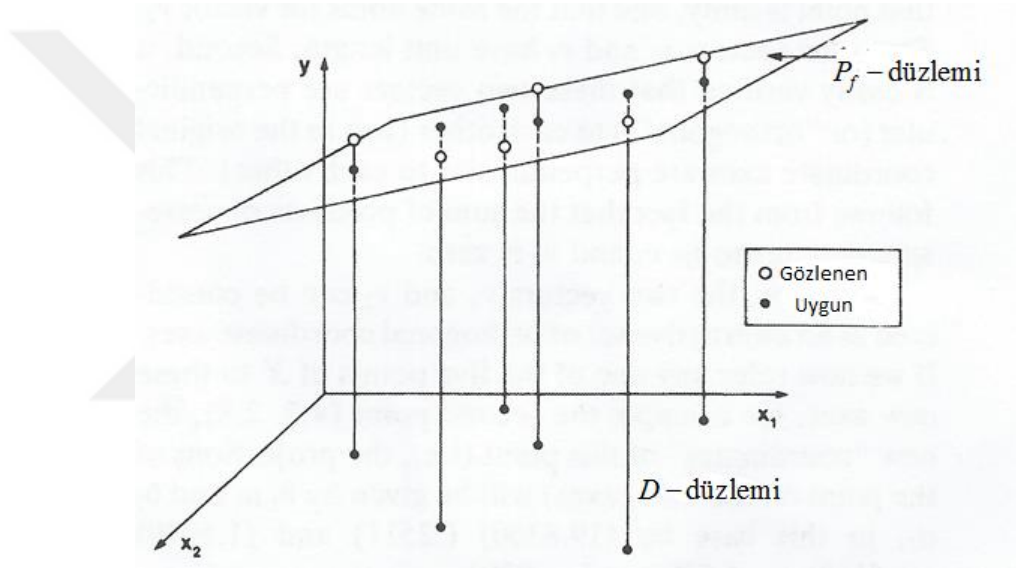
$$\mathbf{X}_{N \times p} = \mathbf{U}_{N \times r} \boldsymbol{\theta}_{r \times r} \mathbf{V}_{r \times p} \quad (3.47)$$

yazarız. $\boldsymbol{\theta}$ matrisi köşegendir ve her θ_k pozitiftir. \mathbf{U} matrisinin sütunları \mathbf{u} vektörleri ve \mathbf{V}' nün satırları (3.44) ün \mathbf{v} vektörleridir. \mathbf{u} ve \mathbf{v} nin ortogonallığı ve onların uzunluğu, $(\cdot)'$ bir matrisin transpozisini göstermek üzere

$$\mathbf{U}'\mathbf{U} = \mathbf{I} \quad (3.48)$$

$$\mathbf{V}'\mathbf{V} = \mathbf{I} \quad (3.49)$$

şartlarını ifade eder. θ_k ; \mathbf{XX}' kare matrisinin yanı sıra $\mathbf{X}'\mathbf{X}$ kare matrisinin sıfırdan farklı özdeğerlerinin karekökleri olarak gösterilebilir. \mathbf{U} nun sütunları \mathbf{XX}' nün özvektörleridir ve \mathbf{V}' nün satırları $\mathbf{X}'\mathbf{X}$ in özvektörleridir. Bir matrisin SDA sını elde etmek için mükemmel bir algoritma mevcuttur (Chambers, 1977). Çizelge 3.2, onun SDA ile birlikte, A veri kümesinin \mathbf{X} matrisini gösterir. \mathbf{U} , $\boldsymbol{\theta}$ ve \mathbf{V}' matrisleri onların \mathbf{X} matrisi ile boyutsal ilişkilerini göstermek için sergilenirler. Bu durumda, r rankı 3 tür, yani, $r = p$ dir. Bu; tam-rank durumu olarak bilinir. \mathbf{X} in her bir elemanı \mathbf{U} , $\boldsymbol{\theta}$ ve \mathbf{V}' matrislerinin karşılık gelen elamanlarını çarparak ve terimleri toplayarak kolaylıkla yeniden inşa edilir.



Şekil 3.2. Bir regresyon yüzeyinin geometrik gösterimi

Çizelge 3.2. A veri kümesinin SDA

X				U		
				u_1	u_2	u_3
	1	16.85	1.46	.322575	.176104	.193765
	1	24.81	- 4.61	.473864	- .603455	.049884
	1	18.85	- .21	.360574	- .038169	.333727
	1	12.63	4.93	.242392	.621398	- .036214
	1	21.38	- 1.36	.408731	- .186949	- .659192
	1	18.78	- .08	.359251	- .021580	.241838
	1	15.58	2.98	.298488	.370545	- .453637
	1	16.30	1.73	.312108	.210988	.385321
V'				θ		
v_1	.053067	.998579	.004786	52.347807	0	0
v_2	.067340	- .008360	.997695	0	7.853868	0
v_3	.996317	- .052622	- .067688	0	0	.055690

Örneğin, dördüncü satır ve üçüncü sütundaki 4.93 elemanı

$$\begin{aligned}
 & [.004786 \times 52.347807 \times .242392] \\
 & + [.997695 \times 7.853868 \times .621398] \\
 & + [(-.067688) \times .055690 \times (-.036214)]
 \end{aligned}$$

toplamına eşittir.

SDA nın geometrik yorumu:

Yorumu basitleştirmek için sadece iki açıklayıcı değişkenli bir örneği göz önüne alacağız. Çizelge 3.3, iki \mathbf{u} vektörü, iki \mathbf{v} vektörü ve köşegen $\boldsymbol{\theta}$ matrisinin yanı sıra beş noktadan oluşan, \mathbf{X} matrisini gösterir. \mathbf{X} matrisinin her bir satırı iki x_1 ve x_2 sayısından oluşur. Bunları, 2-boyutlu uzayda x_1 ve x_2 koordinatlı bir nokta olarak yorumlayabiliriz (bkz. Şekil 3.3). Orijini bu noktaya birleştiren vektör de noktayı temsil etmek için kullanılabilir ve böylece iki sayının herhangi bir kümesine bir noktanın yanı sıra bir vektör olarak bakmak için tereddüt etmeyeceğiz. Bu nedenle, \mathbf{X} matrisi beş nokta veya beş vektör ile temsil edilir.

Benzer şekilde, \mathbf{v}_1 ve \mathbf{v}_2 ile adlandırılan satırlar da her biri bir noktayı veya her biri bir vektörü gösterir. Örneğin; \mathbf{v}_1 noktasının koordinatları .7309 ve .6825 sayılarıdır.

İlk olarak, orijinin o noktaya uzaklığının 1 olduğuna ve aynısının \mathbf{v}_2 için gerçekleştiğine dikkat ederiz. İkinci olarak, bu iki vektörün her birinin diğerine dik (veya ortogonal) olduğu kolayca gerçekleşir (tam orijinal koordinat eksenlerinin her birinin diğerine dik olduğu gibi). Bu; \mathbf{v}_1 ve \mathbf{v}_2 deki karşılık gelen terimlerin çarpımlarının toplamının sıfır olduğu gerçeğinden görülür.

Bu nedenle, iki \mathbf{v}_1 ve \mathbf{v}_2 vektörü ortogonal koordinat eksenlerinin başka bir kümesi olarak gösterilebilir. Eğer şimdi \mathbf{X} in beş noktasından herhangi birine bu yeni eksenlere göre bakarsak, örneğin ikinci nokta (4.2, 2.8), bu noktanın yeni koordinatları (yani, noktanın v_1, v_2 eksenleri üzerindeki izdüşümü), bu durumda $\theta_1 u_1$ ve $\theta_2 u_2$ ile, yani, $(19.8360) \times (.2511)$ ve $(1.6040) \times (-.5113)$ veya 4.9808 ve -0.8201 ile verilecek.

Bu iki sayının göreceli büyüklükleri rastlantısal değildir, noktaların v_1 – eksenindeki izdüşümleri v_2 – eksenindeki izdüşümlerinden daha geniş bir aralığı kapsar. Başka bir deyişle, tasarım matrisinin beş noktası büyük bir çoğunlukla v_1 – eksenini boyunca düşer ve v_2 – eksenini boyunca daha az düşer. Eğer $\theta_2 = 0$ olsaydı, beş noktanın her birinin v_2 – eksenindeki koordinatı sıfır olacaktı ki; bu durumda, beş noktanın tümü v_1 – doğrusu (orijinde v_2 ye dik olan doğru) üzerinde olacaktı. SDA ile yerine getirilen amacın koordinatlara, onları \mathbf{X} matrisinin kendisinin noktalarıyla yapılan kalıbı (yapıyı) daha yakından izletecek bir şekilde, yeniden yön vermek olduğunu görürüz. SDA bize \mathbf{X} matrisinin yapısını anlamaya yardımcı olur.

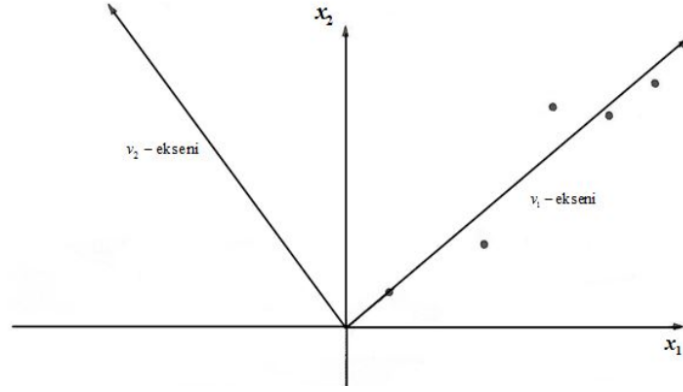
Çizelge 3.3. İki değişkenli bir \mathbf{X} matrisinin SDA

\mathbf{X} – matrisi			\mathbf{U} – matrisi	
	x_1	x_2	u_1	u_2
	1.3	1.2	.1202	.4037
	4.2	2.8	.2511	-.5113
	6.3	7.4	.4867	.6912
	8.0	7.1	.5391	-.1689
	9.4	8.2	.6285	-.2634
\mathbf{V}'			$\boldsymbol{\theta}$ – matrisi	
v_1	.7309	.6825	19.8360	0
v_2	-.6825	.7309	0	1.6040

Çizelge 3.2 için bütün yanlarıyla benzer bir yorum gerçekleşir, fakat burada tasarım değişkenlerinin vektör uzayı üç – boyutludur. Eğer Çizelge 3.2 de, θ_3 kesin olarak sıfır olsaydı tüm noktalar v_1, v_2 düzleminde, yani, orijinde v_3 e dik olan düzlemde olacaktı. A veri kümesi için, θ_3 sifira yakın olduğundan, noktalar bu düzlemde olmaktan ziyade v_1, v_2 ye yakın olur.

Temel bileşenler regresyonu:

Şimdi bu çalışmanın asıl amacı daha kesin sözlerle ifade edilebilir. \mathbf{X} matrisinin SDA sınıfın tanıtımına sahip olarak, şimdi \mathbf{Y} nin \mathbf{X} üzerindeki regresyonunu yürütmeye \mathbf{X} in yerine onun SDA sınıfını koymanın avantajlarını göstermeye çalışıyoruz.



Şekil 3.3. İki açıklayıcı değişken durumunda singüler değer ayrışımının geometrik yorumu

Bu yönteme temel bileşenler regresyonu denir. Bu teknik yukarıda tanımlandığı gibi (3.39) ile kapsanan her regresyon durumunda kullanılabilirken, onun özellikle doğrudanlık ve yakın doğrudanlık durumunda aydınlatıcı olduğunu göreceğiz. Bu terimler (sözler) aşağıda açıklanacak.

(3.47) bağıntısını (3.39) bağıntısı içine yerleştirerek,

$$\mathbf{Y} = \mathbf{U}\boldsymbol{\theta}\mathbf{V}'\boldsymbol{\beta} + \mathbf{e} \quad (3.50)$$

elde ederiz. Bu biçimde yazılan modele temel bileşenler regresyon modeli olarak başvurulur.

$\boldsymbol{\theta}\mathbf{V}'\boldsymbol{\beta}$ bir $r \times 1$ matris, yani, r sayıda elamanın bir vektörü olmak üzere, (3.50) bağıntısı

$$\mathbf{Y} = \mathbf{U}(\boldsymbol{\theta}\mathbf{V}'\boldsymbol{\beta}) + \mathbf{e} \quad (3.51)$$

olarak yazılabilir. Bu vektörü $\boldsymbol{\alpha}$ ile gösterelim. Bu durumda

$$\boldsymbol{\alpha} = \boldsymbol{\theta}\mathbf{V}'\boldsymbol{\beta} \quad (3.52)$$

ve

$$\mathbf{Y} = \mathbf{U}\boldsymbol{\alpha} + \mathbf{e} \quad (3.53)$$

dir. \mathbf{Y} vektörü ve \mathbf{U} matrisi biliniyor. Bilinmeyen $\boldsymbol{\alpha}$ katsayıları için en küçük kareler çözümü alışlagelen matris çözümüyle elde edilir, yani,

$$\hat{\alpha} = (\mathbf{U}'\mathbf{U})^{-1} \mathbf{U}'\mathbf{Y} \quad (3.54)$$

dir ki bu (3.48) in bir sonucu olarak,

$$\hat{\alpha} = \mathbf{U}'\mathbf{Y} \quad (3.55)$$

olur. j -yinci \mathbf{u}_j vektörüyle \mathbf{Y} vektörünün iç çarpımı olduğundan, bu denklem kolayca çözülür.

(3.55) in uygulanmasıyla Çizelge 3.2 de gösterilen \mathbf{U} matrisini kullanarak A veri kümesi için,

$$\hat{\alpha} = \mathbf{U}'\mathbf{Y} = \begin{pmatrix} 111.285635 \\ 36.565303 \\ .018803 \end{pmatrix}$$

elde ederiz.

(3.53) ve (3.55) den

$$\hat{\alpha} = \mathbf{U}'(\mathbf{U}\alpha + \mathbf{e}) = \alpha + \mathbf{U}'\mathbf{e} \quad (3.56)$$

veya

$$\hat{\alpha} - \alpha = \mathbf{U}'\mathbf{e} \quad (3.57)$$

olduğu görülür.

$$E(\hat{\alpha}_j - \alpha_j) = 0 \quad (3.58)$$

veya

$$E(\hat{\alpha}_j) = \alpha_j \quad (3.59)$$

olduğu görülür. Bu nedenle, $\hat{\alpha}_j$ yansızdır. Bundan başka, $\hat{\alpha}_j$ nin varyansı

$$\begin{aligned} \text{var}(\hat{\alpha}_j) &= E(\hat{\alpha}_j - \alpha_j)^2 \\ &= E\left[\sum_l \sum_t u_{lj} u_{lj} e_l e_t\right] \\ &= \left(\sum_l u_{lj}^2\right) \sigma^2 = \sigma^2 \end{aligned}$$

dir. Bundan dolayı

$$\text{var}(\hat{\alpha}_j) = \sigma^2 \quad (3.60)$$

dir. α_j lerin karşılıklı olarak ilişkisiz oldukları da kolayca gösterilir. β nın elemanlarının sayısı p iken, α nın elemanlarının sayısının r olduğuna, yani, p den az olabildiğine dikkat ediniz. Şimdi β ve α arasındaki ilişki, aynı zamanda α ve β nın en küçük kareler tahminleri için de gerçekleşen, yani,

$$\hat{\alpha} = \theta \mathbf{V}' \hat{\beta} \quad (3.61)$$

olan (3.52) bağıntısıyla verilir. (3.61) bağıntısında, $\theta \mathbf{V}'$ $r \times p$ boyutuna sahiptir. Bu nedenle, $\hat{\alpha}$ nın r tane değeri verildiğinde, (3.61) matris bağıntısı p tane bilinmeyen $\hat{\beta}$ parametre tahminlerine göre r tane denklemi gösterir. Eğer $r = p$ ise (tam-rank durumu) çözüm mümkündür ve tektir.

Bu durumda, \mathbf{V}' bir $p \times p$ ortogonal matristir (bkz. (3.49)). Bu nedenle çözüm

$$\hat{\beta} = \mathbf{V}'^{-1} \theta^{-1} \hat{\alpha} = \mathbf{V} \theta^{-1} \hat{\alpha} \quad (3.62)$$

ile verilir. $\theta^{-1} \hat{\alpha}$ nın, her biri $\hat{\alpha}_j$ yi karşılık gelen θ_j ile bölerek, elde edilen bir $p \times 1$ vektör olduğuna dikkat ediniz. A veri kümesine (3.62) yi uygulayarak, Çizelge 3.2 nin \mathbf{X} inin SDA nı kullanarak, hemen

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} .053067 & .067340 & .996317 \\ .998579 & -.008360 & -.052622 \\ .004786 & .997695 & -.067688 \end{pmatrix} \times \begin{pmatrix} \hat{\alpha}_1/52.347807 \\ \hat{\alpha}_2/7.853868 \\ \hat{\alpha}_3/.055690 \end{pmatrix} \quad (3.63)$$

elde ederiz. β nın tahminlerini $\hat{\alpha}$ nın fonksiyonları olarak vermesinin dışında, (3.62) nin önemli bir kullanımı; $\hat{\beta}_j$ nın varyanslarının kolay hesaplanmasıdır. Bu hesaplamayı (3.63) sayısal bağıntılarına dayanarak açıklayacağız. Böylece, (3.63) bağıntısından

$$\hat{\beta}_1 = (.053067) \frac{\hat{\alpha}_1}{52.347807} + (.067340) \frac{\hat{\alpha}_2}{7.853868} + (.996317) \frac{\hat{\alpha}_3}{.055690}$$

elde ederiz. Genel notasyona göre bu denklem

$$\hat{\beta}_j = \sum_{k=1}^p v_{jk} \frac{\hat{\alpha}_k}{\theta_k} \quad (3.64)$$

olarak yazılır. $\hat{\alpha}_j$ ler karşılıklı olarak ortogonal olduklarından ve tümü σ^2 varyansına sahip olduğundan,

$$\text{var}(\hat{\beta}_j) = \left(\sum_{k=1}^p \frac{v_{jk}^2}{\theta_k^2} \right) \sigma^2 \quad (3.65)$$

olduğunu görürüz. β_1 için, verimize uygulanan (3.65)

$$\text{var}(\hat{\beta}_1) = \left(\frac{(.053067)^2}{(52.347807)^2} + \frac{(.067340)^2}{(7.853868)^2} + \frac{(.996317)^2}{(.055690)^2} \right) \times \sigma^2 \quad (3.65a)$$

olur. Her bir terimdeki paylar \mathbf{V} matrisinin birinci satırındaki elemanların, yani, 0 ve 1 arasındaki değerlerin kareleridir. Fakat paydalar θ_j nin kareleridir. Şimdi, θ_3 ün θ_1 ve θ_2 den daha küçük olduğunun, bu nedenle (3.65) deki üçüncü terimin $\hat{\beta}_1$ nin varyansına (ve $\hat{\beta}_2$ ve $\hat{\beta}_3$ ün varyanslarına da) aşırı büyük bir oranda katkı yaptığının farkına varırız. Gerçekten, (3.65a) dan

$$\text{var}(\hat{\beta}_1) = \left[(1.03 \times 10^{-6}) + (74.4 \times 10^{-6}) + (320) \right] \cdot \sigma^2 \quad (3.66)$$

buluruz. Gidişatın bu olumsuz ifadesinin nedeni θ_3 ün (θ_1 ve θ_2 ye göre) çok küçük değerleridir. SDA tekniğinin kullanımının bize bazı katsayılar için bulunan büyük varyanslar için gerçek neden veya nedenleri bulma veya tanımlama fırsatı verdiğini görürüz. Örneğimizde θ nın değerinin, pratik amaçlar için sifıra eşit olacağı düşünülebilir. Fakat sifıra eşit bir θ -değeri, birazdan göreceğimiz gibi, regresyon sonuçlarının yorumu için önemli sonuçlara sahiptir. Bu nedenle, θ nın sifır veya sifıra yakın bir değerinin sonuçlarını incelemek için sayısal örneğimizin tartışmasını yarıda bırakırız.

Doğrudaşlık ve onun regresyon üzerindeki etkileri:

θ lar $\mathbf{X}'\mathbf{X}$ matrisinin özdeğerlerinin kare kökleri olduklarından, bir sıfır θ değeri bir sıfır özdeğeri belirtir. O halde bu, \mathbf{X} matrisinin p sayıda \mathbf{x} – vektörünün en azından bazıları arasında bir lineer ilişkinin varlığına işaret eder. Sonuç olarak sıfıra yakın bir θ değeri bu p sayıda \mathbf{x} – vektörünün en azından bazıları arasında yaklaşık bir lineer ilişkiyi belirtir.

Şekil 3.4 bir yaklaşık lineer ilişkiyi, yani, her i için \mathbf{X} in sütunları ($x_{i1} \equiv 1$ olduğunu hatırlayınız) arasında,

$$15x_{i1} - .75x_{i2} - x_{i3} \cong 0 \quad (3.67)$$

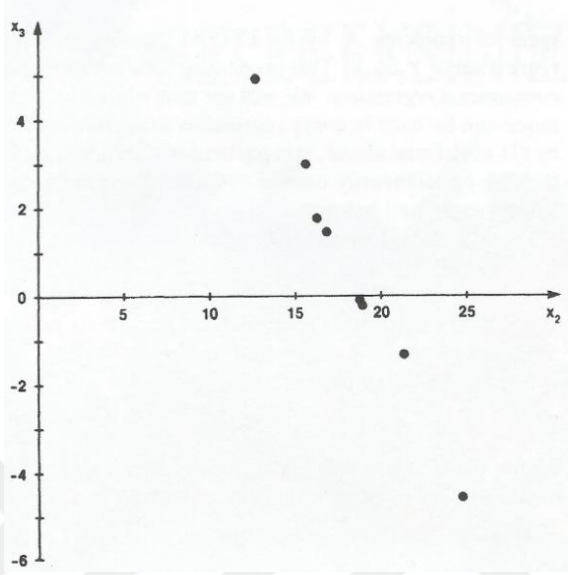
bağıntısının gerçekleştiğini gösterir.

Bu nedenle θ_3 için çok küçük değere neden olan ilişkiyi ortaya koyduk. Bununla beraber, bu; grafikleri yapmaksızın da başarılabilir.

İlk olarak, eksenler x_1 , x_2 ve x_3 ile adlandırılmak üzere, üç boyuta sahip bir Öklid uzayını göz önüne alırsak, (3.67) nin anlamının daha kolayca kavrandığını görürüz. Her bir i için, (x_1, x_2, x_3) üçlüsü, bu uzayda bir noktayı gösterir (iki boyutlu bir benzeri için bkz. Şekil 3.3). (3.67) bağıntısı basit olarak, N noktanın tümünün bir tek düzlemde olduğunu söyler (daha evvel, iki x değişkeni arasındaki bir lineer ilişkinin bir doğru üzerinde bulunan tüm noktaları gösterdiği gibi). Bu nedenle onlar düzlemdaştır, düzlemdaşlık; doğrudaşlığın (noktalar aynı doğru üzerinde) bir genişlemesidir. Bununla beraber, (3.67) gibi bir lineer ilişkinin varlığına daima doğrudaşlık olarak başvurulur (bu nedenle o, aynı doğru üzerindeki noktaların daha fazla sınırlanan durumunun bir genelleştirmesini içeren, bir geometrik terim olarak kullanılır).

Doğrudaşlığın regresyon analizi üzerindeki etkilerini incelemek için; Çizelge 3.4 te gösterilen, B veri kümesi olarak adlandırılan ikinci bir veri kümesini ortaya koyarız. B veri kümesi sadece A kümesinin (Çizelge 3.1) bir değişik biçimidir. x_2 ve x_3 arasındaki ilişki şimdi kesin olarak bir düz doğrudur.

B veri kümesinin \mathbf{X} matrisinin SDA, Çizelge 3.5 te gösterilmiştir. Şimdi θ_3 ün değeri kesin olarak sıfırdır, bu nedenle \mathbf{X} matrisinin rankı 3 ten ziyade 2 dir. Şimdi $r < p$ durumuna sahibiz.



Şekil 3.4. $x_1 = 3$ olduğunda, üç x_1 , x_2 ve x_3 açıklayıcı değişkeni için yakın doğrudanlık (x_2, x_3) noktaları düz bir doğruya yakın düşer

(3.55) bağıntısını kullanarak, sadece iki değerden oluşan $\hat{\alpha}$ yı hesaplayabiliriz. Buna göre,

$$\hat{\alpha} = \begin{pmatrix} 111.524562 \\ 35.628363 \end{pmatrix} \quad (3.68)$$

dür. Sadece iki α – değeri var olsa bile, üç \mathbf{x} – vektörünün her biri için bir tane olmak üzere yine üç β – değeri vardır. (3.52) bağıntısını kullanarak, bu β – değerlerini elde etmeye çalışırız.

$$(\theta \mathbf{V}') \hat{\beta} = \hat{\alpha} \quad (3.69)$$

yazarak,

Çizelge 3.4. B veri kümesi

Nokta	x_1	x_2	x_3	y
1	1	16.85	2.3625	41.38
2	1	24.81	-3.6075	31.01
3	1	18.85	.8625	37.41
4	1	12.63	5.5275	50.05
5	1	21.38	-1.0350	39.17
6	1	18.78	.9150	38.86
7	1	15.58	3.3150	46.14
8	1	16.30	2.7750	44.47
$x_3 = 15x_1 - .75x_2$				

Çizelge 3.5. B veri kümesinin \mathbf{X} matrisin SDA; $\mathbf{X} = \mathbf{U}\mathbf{\Theta}\mathbf{V}'$

$\mathbf{X} =$	$\begin{pmatrix} .323879 & .193143 \\ .469906 & -.598138 \\ .360569 & -.005670 \\ .246463 & .612642 \\ .406982 & -.257171 \\ .359285 & .001287 \\ .300581 & .319391 \\ .313789 & .247817 \end{pmatrix}$					
		52.406330	0	.053074	.997433	.048047
		0	8.036461	.063871	-.051408	.996633

$$\begin{pmatrix} 2.781414 & 52.271803 & 2.517967 \\ .513297 & -.413138 & 8.009402 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} 111.524562 \\ 35.628363 \end{pmatrix} \quad (3.70)$$

elde ederiz. (3.70) bağıntısı üç bilinmeyenli iki denklemi gösterir ve β için bir tek çözümü göstermez. Bununla beraber, o, bize üç $\hat{\beta}$ -değerinin herhangi ikisini üçüncüsünün bir fonksiyonu olarak ifade etme imkanı verir.

Genel durumu ele almak için, $(\mathbf{\Theta}\mathbf{V}')$ $r \times p$ matrisini \mathbf{Z} ile gösterelim, yani,

$$\mathbf{Z} = (\mathbf{\Theta}\mathbf{V}') \quad (3.71)$$

olsun ve \mathbf{Z} yi sırasıyla boyutları $r \times r$ ve $r \times (p-r)$ olan \mathbf{Z}_A ve \mathbf{Z}_B matrislerine parçalayalım. $\boldsymbol{\beta}$ vektörünü buna uygun olarak $r \times 1$ ve $(p-r) \times 1$ alt vektörlerine parçalayarak,

$$(\mathbf{Z}_A \quad \mathbf{Z}_B) \begin{pmatrix} \hat{\boldsymbol{\beta}}_A \\ \hat{\boldsymbol{\beta}}_B \end{pmatrix} = (\hat{\boldsymbol{\alpha}}) \quad (3.72)$$

elde ederiz ki bu;

$$\mathbf{Z}_A \hat{\boldsymbol{\beta}}_A + \mathbf{Z}_B \hat{\boldsymbol{\beta}}_B = (\hat{\boldsymbol{\alpha}}) \quad (3.73)$$

olarak yazılabilir. (3.70) deki veri için, bu denklem

$$\begin{pmatrix} 2.781414 & 52.271803 \\ .513297 & -.413138 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} + \begin{pmatrix} 2.517967 \\ 8.009402 \end{pmatrix} \hat{\beta}_3 = \begin{pmatrix} 111.524562 \\ 35.628363 \end{pmatrix} \quad (3.74)$$

olur. (3.73) ün her iki yanını \mathbf{Z}_A^{-1} ile (bir $r \times r$ tekil olmayan matrisin tersi ile) önden çarparak,

$$\hat{\boldsymbol{\beta}}_A + \mathbf{Z}_A^{-1} \mathbf{Z}_B \hat{\boldsymbol{\beta}}_B = \mathbf{Z}_A^{-1} \hat{\boldsymbol{\alpha}} \quad (3.75)$$

elde ederiz. (3.75) bağıntısı $\hat{\boldsymbol{\beta}}_B$ için bir değer bir kez keyfi olarak seçildiğinde, $\hat{\boldsymbol{\beta}}_A$; bu seçim için tek bir şekilde belirlenir. Bu nedenle, (3.74) de $\hat{\beta}_1$ ve $\hat{\beta}_2$, $\hat{\beta}_3$ nin keyfi olarak seçilen herhangi bir değeri için bir tek olarak belirlenir. Şüphesiz $\hat{\beta}_1$ ve $\hat{\beta}_2$ dan birini de keyfi olarak seçilen parametre olarak alabilirdik.

Doğrudaşlık durumunda tahmin (öngörü):

Doğrudaşlığın sonuçları için arayışımızla devam ederek, “yeni” bir x noktasını göz önüne alalım, bu nokta için \hat{y} yi tahmin etmeyi isteyelim.

$$\hat{y} = \mathbf{x} \hat{\boldsymbol{\beta}} \quad (3.76)$$

veya $\hat{\boldsymbol{\beta}}$ nin parçalanmasını ortaya koyarak,

$$\hat{y} = \mathbf{x} \begin{pmatrix} \hat{\boldsymbol{\beta}}_A \\ \hat{\boldsymbol{\beta}}_B \end{pmatrix} \quad (3.77)$$

yi elde ederiz. $\hat{\beta}_A$ $r \times 1$ boyutlu ve \mathbf{x} $1 \times p$ boyutlu olduğundan, \mathbf{x} i ona göre \mathbf{x}_A ($1 \times r$ boyutlu) ve \mathbf{x}_B ($1 \times (p-r)$ boyutlu) ye parçalarız. Böylece,

$$\hat{y} = (\mathbf{x}_A \quad \mathbf{x}_B) \begin{pmatrix} \hat{\beta}_A \\ \hat{\beta}_B \end{pmatrix}$$

$$\hat{y} = \mathbf{x}_A \hat{\beta}_A + \mathbf{x}_B \hat{\beta}_B \quad (3.78)$$

olur. (3.75) bağıntısını (3.78) bağıntısının içine yerleştirerek,

$$\hat{y} = \mathbf{x}_A \left[\mathbf{Z}_A^{-1} \hat{\alpha} - \mathbf{Z}_A^{-1} \mathbf{Z}_B \hat{\beta}_B \right] + \mathbf{x}_B \hat{\beta}_B$$

veya

$$\hat{y} = \mathbf{x}_A \left(\mathbf{Z}_A^{-1} \hat{\alpha} \right) + \left(\mathbf{x}_B - \mathbf{x}_A \mathbf{Z}_A^{-1} \mathbf{Z}_B \right) \hat{\beta}_B \quad (3.79)$$

elde ederiz. $\hat{\beta}_B$ nin veriyle belirlenemediğini, fakat keyfi olduğunu hatırlayınız. Bu durumda (3.79) u anlamlı kılmak için, \hat{y} nin değerinin $\hat{\beta}$ nin herhangi bir keyfi değeri için değişmeyeceği üzerinde durmalıyız. Bu,

$$\mathbf{x}_B - \mathbf{x}_A \mathbf{Z}_A^{-1} \mathbf{Z}_B = 0 \quad (3.80)$$

olduğunu belirtir. \mathbf{V}' ; \mathbf{Z} ye benzer şekilde parçalanmak üzere, $\mathbf{Z}_A^{-1} \mathbf{Z}_B = (\mathbf{V}'_A)^{-1} \mathbf{V}_B$ olduğu kolayca görülür (\mathbf{V}_A nın bir ortogonal matris olmadığına dikkat ediniz). Bu nedenle, çarpım θ matrisini içermez ve (3.80)

$$\mathbf{x}_B = \mathbf{x}_A \left[\mathbf{V}'_A^{-1} \mathbf{V}_B \right] \quad (3.80a)$$

olarak yazılabilir. Eğer (3.80) gerçekleşirse, çözüm

$$\hat{y} = \mathbf{x}_A \left(\mathbf{Z}_A^{-1} \hat{\alpha} \right) \quad (3.81)$$

dır. (3.79) un iki bağıntıyı, yani, (3.80) şartı ve (3.81) çözümünü ürettiğini gerçeklemek önemlidir. Fakat çözüm sadece şart gerçekleştiğinde geçerlidir.

B veri kümesi için,

$$\mathbf{Z}_A^{-1} = \begin{pmatrix} 68.206904 \\ -1.495779 \end{pmatrix} \quad (3.82)$$

$$\mathbf{Z}_A^{-1} \mathbf{Z}_B = \begin{pmatrix} 15.000 \\ -.750 \end{pmatrix} \quad (3.83)$$

elde ederiz. Bu nedenle, şart

$$x_3 = 15x_1 - 0.75x_2 \quad (3.84)$$

olur ve (3.84) e bağlı çözüm

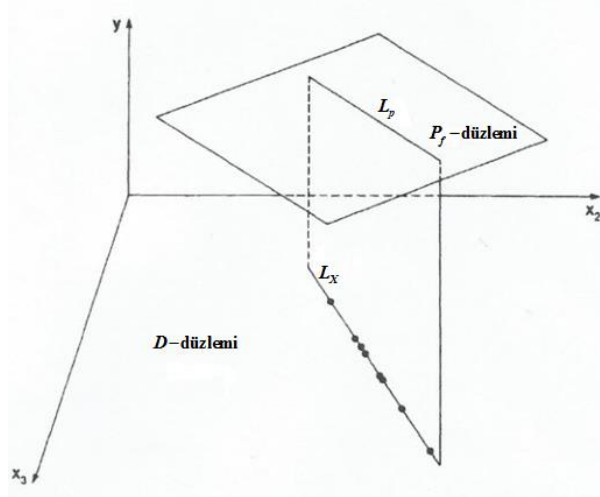
$$\hat{y} = 68.207x_1 - 1.496x_2$$

dir ki bu (3.84) ün bir sonucu olarak,

$$\hat{y} = 1.9144x_2 + 4.5471x_3 \quad (3.85)$$

olarak yazılabilir.

$x_1 = 1$ olduğundan, (3.84); x_2 ve x_3 arasında bir düz doğru ilişkisinin grafiğidir. Çizelge 3.4 ün verisini kullanarak, x_2 ye karşı x_3 ün bir şematik grafiği Şekil 3.5 te gösterilir ve bu ilişki sergilenir. Fakat (3.84) türetimiz bu ilişkinin, bir \hat{y} tahmininin gelecek herhangi bir (x_1, x_2, x_3) noktası için de sağlanması gerektiğini gösterir.



Şekil 3.5. Doğrudaşlığın regresyon yüzeyi üzerindeki etkisi

Bu nedenle, $\mathbf{x} = (1 \ 15 \ 3.75)$ için gerçek bir tahmini, yani, $\hat{y} = 45.767$ yi elde edebiliriz, fakat $\mathbf{x} = (1 \ 20 \ 3.75)$ için elde edemeyiz. Nedeni Şekil 3.5 e dayanarak

kolayca anlaşılır. Dayandırılan tahmin işlemimiz üzerindeki \mathbf{X} matrisi (x_2, x_3) noktalarının D düzleminde düz bir doğruyla temsil edilen, x_2 ve x_3 arasında kesin bir ilişki var olacak şekildedir. Bu doğruya L_x deyiniz. Regresyon problemi (x_2, x_3, y) uzayında sıfırdan farklı bir intercept (sabit terim) ile gözlenen y değerine en iyi uyan mümkün olabilir bir P_f –düzlemi bulmaktır. Fakat y lerin tümü L_x doğrusu boyunca düşer. (x_2, x_3) noktalarında ortaya çıktığından, istenen düzlem sadece L_x üzerinde dikilen dik düzlemlerle onun kesişimiyle tanımlanır. Bu kesişim Şekil 3.5 te L_p ile gösterilir ve bulunacak düzlemin L_x üzerinde konumlanmayan herhangi bir nokta için \hat{y} nın bir tek değerini üretmeyerek, serbestçe bu kesişim doğrusuna dönürebildiği görünür.

Kısaca, eğer $r < p$ ise, bu x noktasının sıfır θ değerinden ortaya çıkan doğrudakiği yerine getirdiği durumdaki hariç, bir x noktası için \hat{y} için tek bir çözüm yoktur. Bu şart (3.80) ile verilir ve eğer (3.80) şartı gerçekleşirse, çözüm (3.81) ile verilir.

Regresyon ve tahmin (öngörü):

Şimdi \mathbf{X} matrisinin tam rank olduğu fakat çok küçük θ –değerine sahip olduğu, A veri kümemize (Çizelge 3.1) döneriz. Burada, önceki kısımda ele alınan L_x doğrusu kesin surette mevcut değildir, fakat \mathbf{X} matrisinin sekiz noktasının hepsi böyle bir doğruya çok yakın düşer. Tam anlamıyla konuşarak, bulunacak düzlem şimdi serbestçe döndürülemez ve herhangi bir x noktası için \hat{y} için tahminler elde edilemez. Bununla beraber, \mathbf{X} matrisini tanımlayan noktaların şanssız seçiminden dolayı istenen düzlemin kesin durumu, yaklaşık L_x doğrusu üzerinde yükselen dik düzlemlerle onun kesişiminin yakın komşuluğundaki hariç, kötü bir şekilde (zayıf olarak) bilinir.

Bu iddiayı ispatlamak için, “yeni” bir \mathbf{x} $(1 \times p)$ noktasını göz önüne alınız. İlk olarak basit

$$\mathbf{X} = \mathbf{U}\mathbf{\Theta}\mathbf{V}'$$

SDA denklemini kullanarak, bu noktayı u – koordinatları cinsinden ifade ederiz. Bir tek x noktasına (örneğin, \mathbf{X} deki bir tek satıra, fakat aynı zamanda p elemanlı herhangi bir “yeni” noktaya) uygulandığında, \mathbf{x} ve \mathbf{u} nun her ikisi de p elemanlı olmak üzere, bu denklem

$$\mathbf{x} = \mathbf{u}\boldsymbol{\theta}\mathbf{V}' \quad (3.86)$$

değerini üretir. (3.86) dan (tam-rank durumu için, $\mathbf{V}^{-1} = \mathbf{V}'$ olduğundan)

$$\mathbf{u} = \mathbf{x}\mathbf{V}\boldsymbol{\theta}^{-1} \quad (3.87)$$

elde ederiz. Bu bağıntıyı ayrıca aşağıda ele alacağız, fakat ilk olarak \hat{y} nin x noktasındaki tahmini ile ilgileneceğiz. Noktayı, (3.87) ile verilen u – koordinatları cinsinden ifade ederek,

$$\hat{y} = \mathbf{u}\hat{\boldsymbol{\alpha}} \quad (3.88)$$

ve sonuç olarak, (bkz. 3.60)

$$\text{var}(\hat{y}) = \sigma^2 \sum_{j=1}^p u_j^2 \quad (3.89)$$

elde ederiz. Bu denklem, \mathbf{u} – vektörünün sayısal olarak büyük bir tek bileşeninin bile aslında \hat{y} nin varyanslarını artırabildiğini gösterir.

Şimdi \mathbf{u} vektörünün büyük bileşenlere sahip olabilmesi şartları altında inceleme yapacağız.

$\mathbf{X} = \mathbf{U}\boldsymbol{\theta}\mathbf{V}'$ temel bağıntısından,

$$\mathbf{U}\boldsymbol{\theta} = \mathbf{X}\mathbf{V} \quad (3.90)$$

elde ederiz. $\boldsymbol{\theta}$ ları sonraki (ikinci) değerlerinden önemli derecede daha küçük düşünülen tüm $\boldsymbol{\theta}$ değerlerini içerecek şekilde, $\boldsymbol{\theta}_A$ ve $\boldsymbol{\theta}_B$ gibi iki gruba ayıralım.

Örneğin Çizelge 3.2 de,

$$\boldsymbol{\theta}_A = \begin{pmatrix} 52.347807 & 0 \\ 0 & 7.853868 \end{pmatrix} \quad (3.91)$$

ve

$$\theta_B = .055690 \quad (3.92)$$

yapacaktık. Bu durumda θ_B bir tek değerden oluşur, fakat genel olarak, o; diyelim ki l sayıda değeri içerir. Bu takdirde, θ_A $p-1$ değere sahip bir köşegen matristir.

$$p-1 = t \quad (3.93)$$

olsun. Bu durumda θ_A bir kare, yani, $t \times t$ köşegen matristir ve θ_B bir kare, yani, $l \times l$ köşegen matristir. (3.90) daki θ ve \mathbf{V} nin her ikisini de parçalayarak, \mathbf{V}_A $p \times t$ ve \mathbf{V}_B $p \times 1$ olmak üzere,

$$\mathbf{U} \begin{pmatrix} \theta_A & \mathbf{0} \\ \mathbf{0} & \theta_B \end{pmatrix} = \mathbf{X} (\mathbf{V}_A \quad \mathbf{V}_B) \quad (3.94)$$

elde ederiz. Bu denklem,

$$\left(\mathbf{U} \begin{pmatrix} \mathbf{0} \\ \theta_A \end{pmatrix} \quad \mathbf{U} \begin{pmatrix} \mathbf{0} \\ \theta_B \end{pmatrix} \right) = (\mathbf{XV}_A \quad \mathbf{XV}_B) \quad (3.95)$$

olarak yazılabilir. Şimdi B grubundaki tüm θ -değerleri çok küçük olduğundan, $\mathbf{U} \begin{pmatrix} \mathbf{0} \\ \theta_B \end{pmatrix}$ ile gösterilen sütunlar tümü çok küçük olan elemanları içerir. Bu durumda aynısı ister istemez \mathbf{XV}_B deki elemanlar için de geçerlidir. Bu nedenle, B grubunda θ -değerlerinin küçüklüğü " \cong " yaklaşık eşitliği göstermek üzere,

$$\mathbf{XV}_B \cong 0 \quad (3.96)$$

olduğunu belirtir. (3.96) bağıntısı B grubunda θ ların küçüklüğü ile \mathbf{X} matrisi üzerine koyulan şarttır ve \mathbf{X} in sütunları arasında bir veya daha fazla lineer ilişkiyi gösterir. Örneğin, Çizelge 3.2 için,

$$\mathbf{XV}_B = \mathbf{X} \begin{pmatrix} .996317 \\ -.052622 \\ -.067688 \end{pmatrix} \quad (3.97)$$

elde ederiz ki bu, \mathbf{X} deki her satır için

$$.996317x_1 - .052622x_2 - .067688x_3 \cong 0 \quad (3.98)$$

olduğunu belirtir.

Bunun \mathbf{X} matrisi için sağlandığı gerçektebilir ve esasen o, (x_2, x_3) noktalarına ($x_1 \equiv 1$ olduğunu hatırlayınız) bir lineer uyumu gösteren bir L_x in denklemdir. Gerçekten, (3.98) in her iki yanını x_3 ün katsayısı ile bölerek, (3.98) in aslında (3.84) ün aynısı olduğu görülür.

Yeni bir x noktası için y yi tahmin etme problemine geri dönerek (3.87) denklemini

$$\mathbf{u} = (\mathbf{xV}_A \quad \mathbf{xV}_B) \boldsymbol{\theta}^{-1} \quad (3.99)$$

veya

$$\mathbf{u} = (\mathbf{xV}_A \quad \mathbf{xV}_B) \begin{pmatrix} \boldsymbol{\theta}_A^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\theta}_B^{-1} \end{pmatrix} \quad (3.100)$$

biçiminde yazarız. \mathbf{u} nun son l sayıda elemanı

$$(\mathbf{xV}_B) (\boldsymbol{\theta}_B^{-1})$$

çarpımından ortaya çıkar ve $\boldsymbol{\theta}_B$ deki her bir eleman çok küçük ve sonuç olarak $\boldsymbol{\theta}^{-1}$ tersi çok büyük olduğundan, \mathbf{u} nun elemanları çok büyük olacak, $\mathbf{XV}_B \cong 0$ olmadıkça, yani, yeni x noktası \mathbf{X} matrisinin yakın doğrudanlık şartını (3.97) sağlamadıkça, \hat{y} nin varyansının büyük olmasına neden olacak yeni x noktası L_x doğrusundan daha ötede olacak, yani, \hat{y} nin varyansı daha büyük olacak, yani, tahmin edilen değer hassasiyeti zayıf olacak. Tahmin etme amaçları için tahmin edilen regresyonun kullanılabilirliği bakış açısından bakıldığında, bu gerçekten doğrudanlık probleminin kalbi (özü) dir.

Eğer x noktası için yakın-doğrudanlık kesin olarak gerçekleşirse, (3.100) den

$$\mathbf{xV}_B = 0 \text{ ve } \mathbf{u} = \mathbf{xV}_A (\boldsymbol{\theta}_A^{-1}) \quad (3.101)$$

elde ederiz ki bu \mathbf{u} nun son l sayıda elemanının sıfır ve

$$\hat{y} = u_1 \hat{\alpha}_1 + \dots + u_t \hat{\alpha}_t \quad (3.102)$$

olduğunu belirtir. Bu denkleminde sadece ilk t sayıda $\hat{\alpha}$ değerleri içerilir Diğerleri sıfır çarpanlara sahip olmaktadır.

A veri kümesi (Çizelge 3.1), yakın doğrudanlık şartı (3.98) bağıntısıyla verilir ve (3.100) (bir eşitlik olarak göz önüne alınan) ü sağlayan x için tahmin edilen \hat{y}

$$\begin{aligned}\hat{y} &= \mathbf{u}\hat{\boldsymbol{\alpha}} = \mathbf{x}(\mathbf{V}_A)(\boldsymbol{\theta}_A^{-1})(\hat{\boldsymbol{\alpha}}) \\ &= \mathbf{x} \begin{pmatrix} .053067 & .067340 \\ .998579 & -.008360 \\ .004786 & .997695 \end{pmatrix} \begin{pmatrix} \frac{1}{52.347807} & 0 \\ 0 & \frac{1}{7.853868} \end{pmatrix} \begin{pmatrix} 111.524562 \\ 36.628363 \end{pmatrix} \\ &= \mathbf{x} \begin{pmatrix} .418452 \\ 2.089422 \\ 4.536091 \end{pmatrix}\end{aligned}\quad (3.103)$$

Özet olarak: Yakın doğrudanlık durumu bir veya daha fazla küçük (sıfır olmadığı halde) $\boldsymbol{\theta}$ değerleriyle karakterize edilir. \mathbf{X} in rankı p dir, bu nedenle tahminler kural olarak geçerli olacak temel modelin bilindiği herhangi bir x noktası için yapılabilir. Bununla beraber, tahmin edilen değerlerin hassasiyeti, x noktasının

$$\mathbf{x}\mathbf{V}_B \cong 0 \quad (3.104)$$

bağıntısıyla verilen yakın-doğrudanlık şartından gitgide ayrılırken gitgide zayıflar. Eşitlik işareti kesin olarak sağlandığında bu şartı sağlayan bir x noktası için tahmin edilen \hat{y} :

$$\hat{y} = \mathbf{x}\mathbf{V}_A(\boldsymbol{\theta}_A^{-1})\hat{\boldsymbol{\alpha}} \quad (3.105)$$

dır. (3.104) deki daha zayıf yaklaşım o noktada \hat{y} nın daha zayıf hassasiyetidir.

Bu noktada önemli bir yorum yapılabilir. Eğer B grubundaki $\boldsymbol{\theta}$ -değerleri çok küçük iseler \mathbf{V} matrisinin karşılık gelen parçalanması, yani, \mathbf{V}_B çok zayıf sayısal doğrulukta bilinecek ve bu durumda (3.104) çok büyük yuvarlatma hatalarını içerebilir. O durumda, $\boldsymbol{\theta}_B$ tam sıfır gibi çok daha iyi rol oynamaktadır ve yakın-doğrudanlık şartı (3.80) vasıtasıyla ifade edilmektedir. A veri kümesini bir açıklama olarak kullanarak (bu durumda (3.104)) uygun olduğu halde θ_3 ü ve \mathbf{V}' nün üçüncü satırını yok sayarak,

$$\mathbf{Z}_A^{-1}\mathbf{Z}_B = \begin{pmatrix} 14.7193 \\ -.7774 \end{pmatrix}$$

elde ederiz ve sonuç olarak, (3.80) yi uygulayarak (3.98) e denk olan

$$x_3 = 14.7193x_1 - .7774x_2$$

doğrudaşlık şartını elde ederiz.

Yanlı tahmin:

Yakın doğrudaşlık durumunda, esasen tasarım matrisinin noktalarının yer aldığı \mathbf{X} in altuzayından (doğrudaşlık şartıyla tanımlanan altuzay) gözle görülür derecede uzaklaşan x -noktaları için \hat{y} nin varyansının şiddetli bir şekilde arttığını gördük. Bu şart \hat{y} tahmin edicisinin yansızlığının şartını gözden çıkarmak suretiyle daha kesin tahminleri elde etmek için çeşitli teşebbüslere sevk etti. Bu nedenle, önerilen yöntemler “yanlı tahmin” olarak bilinir. Önerilen yöntemlerin sadece biri ile, yani, doğrudan doğruya “temel bileşenler regresyon tekniği” ile ilgili olanla ilgileneceğiz. Bununla beraber irdelememiz problemin temel yapısını ve önerilen çözümleri açığa çıkarmak için yeterli olacak.

Temel bileşenler regresyonu:

Aşağıda, bir yakın-doğrudaş \mathbf{X} matrisine sahip olduğumuzu kabul ederiz. (3.101) ve (3.102) ye göre eğer bir x -noktası $\mathbf{xV}_B = 0$ şartını sağlarsa, y nin bu noktaya karşılık gelen tahmini

$$\mathbf{u} = \mathbf{xV}\boldsymbol{\theta}^{-1}$$

olmak üzere, $\hat{y} = \sum_j u_j \hat{\alpha}_j$ daki ilk t sayıda terimle verilir. y nin bu tahmini $\mathbf{xV}_B = 0$ şartını sağlayan noktalar için doğru en küçük kareler çözümünü üretirken, bu şartı sağlamayan x -noktaları için, yani, doğrudaşlık şartı ile tanımlanan altuzayda olmayan noktalar için o en küçük kareler çözümü değildir.

Bununla beraber $\mathbf{xV}_B = 0$ şartının sağlanıp sağlanmadığına bakmaksızın

$$\tilde{y} = \sum_{j=1}^t u_j \hat{\alpha}_j \tag{3.106}$$

niceliğini \tilde{y} ile gösterelim. (3.106) bağıntısına çoğu kez yanlı temel bileşen tahmin denklemini denir. Böyle bir tahmini göz önüne alma nedeni şüphesiz onun, varyansın indirgenmesini başarmasıdır.

$$\text{var}(\hat{y}) = \text{var}\left(\sum_{j=1}^p u_j \hat{\alpha}_j\right) = \sigma^2 \sum_{j=1}^p u_j^2 \quad (3.107)$$

$$\text{var}(\tilde{y}) = \text{var}\left(\sum_{j=1}^t u_j \hat{\alpha}_j\right) = \sigma^2 \sum_{j=1}^t u_j^2 \quad (3.108)$$

elde ederiz. A veri kümesi için $t=2$ alarak ve bir örnek olarak $\mathbf{x} = (1 \ 23 \ -6)$ noktasını göz önüne alarak

$$\mathbf{u} = \mathbf{x}(\mathbf{V}\boldsymbol{\theta}^{-1}) = (.439 \ -0.778 \ 3.450)$$

ve sonuç olarak,

$$\begin{aligned} \text{var}(\tilde{y}) &= \left[(.439)^2 + (-0.778)^2 + (3.45)^2 \right] \sigma^2 \\ &= 12.701 \sigma^2 \end{aligned}$$

$$\text{var}(\hat{y}) = \left[(.439)^2 + (-0.778)^2 \right] \sigma^2 = .7980 \sigma^2$$

elde ederiz. Bu takdirde, bu durumda y nin tahmininin varyansındaki azalma çok büyüktür.

Öte yandan, \tilde{y} bir yanlı tahmin edici iken, \hat{y} tahmin edicisi yansızdır.

$$E(\hat{y}) = E(y) \quad (3.109)$$

$$E(\tilde{y}) = E(y) - \sum_{j=t+1}^p u_j \alpha_j \quad (3.110)$$

elde ederiz. Bu nedenle,

$$\text{yan}(\tilde{y}) = E(\tilde{y}) - E(y) = - \sum_{j=t+1}^p u_j \alpha_j \quad (3.111)$$

dir.

$$\text{HKO} = \text{Varyans} + \text{yan}^2 \quad (3.112)$$

ile tanımlanan, hata kareler toplamını HKO ile göstererek,

$$\text{HKO}(\hat{y}) = \sigma^2 \sum_{j=1}^p u_j^2 \quad (3.113)$$

$$\text{HKO}(\tilde{y}) = \sigma^2 \sum_{j=1}^p u_j^2 + \left(-\sum_{j=t+1}^p u_j \alpha_j \right)^2 \quad (3.114)$$

elde ederiz. \hat{y} nin yerine \tilde{y} yı kullanarak, aslında varyanstaki $\sigma^2 \sum_{j=t+1}^p u_j^2$ ye eşit olan azalmayı, bir yan^2 lik giriş için $\left(-\sum_{j=t+1}^p u_j \alpha_j \right)^2$ ye eşit olan azalma ile takas ederiz.

$t = p-1$ olduğu (sadece bir θ -değeri küçük olduğunda ortaya çıkan) durumu inceleyelim. Bu durumda varyanstaki azalma $\sigma^2 u_p^2$ dir ve $(\text{yan})^2 = u_p^2 \alpha_p^2$ dir. Bu nedenle, \tilde{y} nin \hat{y} dan daha küçük bir HKO ya sahip olması için gerek ve yeter şart, $(\text{yan})^2 < \text{varyans azalması}$, yani,

$$u_p^2 \alpha_p^2 < \sigma^2 u_p^2$$

veya

$$\alpha_p^2 < \sigma^2 \quad (3.115)$$

olmasıdır. Maalesef, hem α_p ve hem de σ bilinmeyen parametrelerdir, onlar için sadece tahminler elde mevcuttur. α_p için tahmin şüphesiz $\hat{\alpha}_p$ dir ve σ^2 için tahmin $\hat{\sigma}^2$, $\hat{\alpha}_1$, $\hat{\alpha}_2$ ve $\hat{\alpha}_3$ ü uydurduktan sonra hata varyansıdır. En küçük kareler teorisinden

$$\hat{\sigma}^2 = \frac{\left(\sum_{i=1}^N (y_i - \hat{y}_i)^2 \right)}{(N-p)} \quad (3.116)$$

elde ederiz ki bu $\hat{\alpha}$ tahmin edicileri kullanıldığında,

$$\hat{\sigma}^2 = \frac{\left(\sum_{i=1}^N (y_i - \hat{y}_i)^2 \right)}{(N-p)} = \frac{\left(\sum_{i=1}^N y_i^2 - \sum_{i=1}^p \hat{\alpha}_i^2 \right)}{(N-p)} \quad (3.117)$$

olur. 1 veri kümesi için, ($p = 3$), $\hat{\alpha}_3 = .018803$,

$$\hat{\sigma} = \sqrt{\frac{13729.3041 - 13721.5143}{8-3}} = 1.248$$

elde ederiz. Hiç şüphesiz, bu durumda $\hat{\alpha}_3 < \hat{\sigma}$ dır fakat bu ille de $\alpha_3 < \sigma$ olduğunu belirtmez. Gerçekten $\hat{\alpha}_3$ nin standart hatası σ dır (3.60 ile karşılaştırınız); eğer $\hat{\alpha}_3$, σ ile aynı büyüklük mertebesine sahipse, (3.115) şartının sağlanıp sağlanmadığına, sadece veriye dayanarak karar vermek esas itibariyle mümkün olmayacaktır. \tilde{y} nin \hat{y} ya göre tercih edilebilir olup olmadığına karar vermek için, yani, $\text{HKO}(\tilde{y}) < \text{HKO}(\hat{y})$ olup olmadığına karar vermek için, ekseriyeye daha öte varsayımlar yapmak gerekir. Örneğin, eğer $\alpha_3 = 0$ olduğunu kabul edersek, bu takdirde \tilde{y} nin kullanılması açıktır. $\alpha_3 = 0$ hipotezi,

$$t = \frac{\hat{\alpha}_3 - \alpha_3}{S_{\hat{\alpha}_3}} = \frac{\hat{\alpha}_3 - \alpha_3}{\hat{\sigma}} \quad (3.118)$$

Student t testini kullanarak temel bileşen regresyonunda test edilebilir ki o, $\alpha_3 = 0$ hipotezi için,

$$t = \frac{\hat{\alpha}_3}{\hat{\sigma}} \quad (3.119)$$

olur. 1 veri kümesi için,

$$t = \frac{.088}{1.248} = .0705$$

elde ederiz. $\alpha_3 = 0$ hipotezi reddedilemez, fakat bu bize (3.115) şartının sağlandığı sonucunu çıkarma izni vermez.

Deney yapma felsefesi açısından, önemli başka bir mesele düşünmeyi hak eden problemimiz hem bir lineer model varsayımı ve hem de bir veri kümesini içerir. “İyi”

bir denemede, verinin kendisi genellikle modeli test etmek için bazı tanısai (teşhisle ilgili) araçları sağlar. Yakın-doğrudaşlık durumunda, bu tanısai araçlar tamamıyla yakın-doğrudaşlık tanımlanan altuzayın yakın komşuluğuna toplanır. Bu nedenle (3.115) şartının yerine getirildiğini bilseydik bile, büyük u_3 – değerli noktalar için, yani, (v_1, v_2) düzleminden v_3 doğrultusunda uzaklaşan noktalar için, modelin geçerliliği hakkında yine tereddütümüz olacaktı.

Analizimiz bizi yakın-doğrudaş X matrislerine dayanan lineer model çıkarımlarının büyük dikkatle yapılması gerektiği sonucuna götürdü.

SDA tekniğinin diğer uygulamaları:

Doğrudaşlık ve yakın-doğrudaşlığın tespit ve iyileştirmesinin singüler değer ayrışımının kullanımıyla geniş ölçüde hafifletildiğini gördük. Bu ilginç teknik veri analizinde iki-yönlü tablolar (Mandel, 1971; Bradu ve Gabriel, 1978) regresyonda deneysel tasarımların değerlendirilmesi (Hahn ve ark., 1976) ve iki veya daha fazla argümanın fonksiyonlarının deneysel uyumu (Mandel, 1981) gibi diğer önemli kullanımlara sahiptir.

Özet ve sonuç:

Öncelikle, singüler değer ayrışımı tekniğine dayanarak, çoklu lineer regresyonun cebirsel ve geometrik görünümünü sunduk ve açıklayıcı değişkenler arasında kısıtlamalar (kesin ve yaklaşık lineer kısıtlamalar gibi) mevcut olduğunda yorumun zorluklarının ortaya çıktığını gösterdik. Lineer kısıtlamalar doğrudaşlık olarak bilinirler. Yakın doğrudaşlık bir veya daha fazla çok küçük singüler değerli biçimde kendisini ortaya koyar.

Lineer kısıtlamalar altında tepkime ve açıklayıcı değişkenler arasındaki ilişkinin gerçek katsayıları ek varsayımları ortaya koymaksızın açık bir biçimde tahmin edilemez. Bununla beraber, tahminin yapıldığı noktanın üzerinde regresyon hesaplarının dayandırıldığı noktalarla aynı altuzayda olması şartıyla tepkimenin geçerli tahminini yapmak mümkündür.

En küçük kareler tekniği doğrudalık veya yakın – doğrudalık durumuna uygulandığında, regresyon denkleminin katsayıları kesin olarak tahmin edilemese bile, katsayıların belirli lineer kombinasyonları belli bir güvenle tahmin edilebilir.

3.3 Çok Değişkenli Lineer Regresyonda İstatistiksel Testlerin Geometrik Yorumu

Langsrud (2004) çok değişkenli lineer regresyonda istatistiksel testlerin geometrik yorumuna yer verdi. Bu tezde hedeflenen çalışmanın amacına uygun olan bu bilgileri burada sunmak istiyoruz.

Tek değişkenli, iki örnekleme oluşturulan klasik regresyon doğrusu kestirici ve tepkimenin seçimine bağlıdır, fakat iki regresyondaki regresyon parametrelerinin (eğer sıfır ise) test edilmesi aynı t –istatistiklerini ve p –değerlerini verir. Birçok kestirici değişken var olduğunda t –testi bir F –testine genelleştirilir. Değişkenlerin her iki kümesi çok değişkenli olduğunda F –istatistiğinin birtakım genelleştirmesi, örneğin; Wilks’in lamdası vardır. İki regresyon modeli (x üzerine y ve y üzerinde x) için genelleştirilmiş test istatistikleri kesin olarak yine aynı değeri alır ve ilgili sıfır dağılımları da aynıdır.

İki testin denkliği birçok istatistikçi için bilinmezdir, ancak eğer testler geometrik olarak anlaşılırsa, denklik açıktır. Bu, testlerin geometrisinin nasıl daha iyi sezgisel anlamayı temin edebildiğinin bir örneğidir. Geometrik formüllemeler bazı yazarlar tarafından ifade edilir [Dempster, 1969; Eaton, 1983; Wickens, 1995; Saville ve Wood, 1996] fakat maalesef standart istatistiksel ders kitapları bu konuya çok dikkat etmezler. Bu çalışma en yaygın istatistiksel testlerin basit ve birleştirici bir geometrik tanımını sunar. Özellikle, Wilks’in lamda istatistiğine hoş bir yaklaşımla gösterir.

$H_0 : \Sigma_{xy} = \mathbf{0}$ hipotezini test etme:

n sayıda bağımsız gözleme sahip iki \mathbf{X} $n \times p$ ve \mathbf{Y} $n \times q$ veri matrisine dayanarak p sayıda x –değişkeni ve q sayıda y –değişkeni arasında herhangi bir bağımlılığın

var olup olmadığının test edilmesiyle ilgileniriz. $\mathbf{g} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = [x_1, \dots, x_p \mid y_1, \dots, y_q]'$

parçalanmış vektörü değişkenlerin her iki kümesine bağlı gözlemleri gösterebilir. \mathbf{g} nin çoklu dağılıma sahip olduğu ve \mathbf{g} nin kovaryans matrisinin

$$\text{kov}(\mathbf{g}) = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \quad (3.120)$$

olarak parçalandığını kabul ettik. Değişkenlerin iki kümesi arasında bağımsızlığa sahip olmamız için gerek ve yeter şart $\Sigma_{xy} = \mathbf{0}$ olmasıdır (Morrison 1976, Bölüm 7.5). Bu nedenle ilgili hipotezimiz

$$H_1 : \Sigma_{xy} \neq \mathbf{0} \text{ hipotezine karşı } H_0 : \Sigma_{xy} = \mathbf{0} \quad (3.121)$$

Aşağıda, \mathbf{X} ve \mathbf{Y} nin bir fonksiyonu olan bir test istatistiğinin bulunması problemini basitleştirmek istiyoruz.

Değişmezlik özellikleriyle problemi indirgeme:

Analizi etkileyen sabit terimlerden (regresyon sabitlerinden) kaçınmak için her iki veri kümesi merkezileştirilir, yani, her bir x - ve y - değişkeni için ortalama değer çıkarılır:

$$\mathbf{X}_c = \mathbf{X} - \frac{1}{n} \mathbf{j}_{(n \times 1)} \mathbf{j}'_{(n \times 1)} \mathbf{X} \quad (3.122)$$

$$\mathbf{Y}_c = \mathbf{Y} - \frac{1}{n} \mathbf{j}_{(n \times 1)} \mathbf{j}'_{(n \times 1)} \mathbf{Y} \quad (3.123)$$

yazılır. Şimdi geometrik olarak tüm sütun vektörleri, aşağıda \mathcal{S} denilen $\mathbf{j}_{(n \times 1)}$ vektörüne ortogonal olan, $(n-1)$ -boyutlu altuzay üzerine izdüşürülür.

Analizin, her iki kümenin tekil olmayan lineer dönüşümleri altında değişmediğini varsayalım. Başka bir deyişle, \mathbf{X}_c (veya \mathbf{Y}_c) yi başka bir $n \times p$ - matris (veya $n \times q$ - matris) ile yer değiştirme, yeni matrisin sütunları aynı altuzayı (sütun uzayı) orijinal matrisin sütunları gibi gerdiği sürece sonuçları değiştirmeyecek. Bu nedenle, merkezleştirilmiş matrisler ile çalışmanın yerine \mathbf{X}_c ve \mathbf{Y}_c nin sütun uzayları için ortonormal tabanları oluşturan vektörlerden oluşan matrisler ile çalışabiliriz. Böyle

matrisleri elde etmek için en basit yol **QR**-ayrışımı olarak da bilinen Gram-Schmidt ortogonalleştirme yöntemini uygulamaktır (Strang, 1988).

$$\mathbf{X}_c = \mathbf{X}_0 \mathbf{L}_X \quad (3.124)$$

$$\mathbf{Y}_c = \mathbf{Y}_0 \mathbf{L}_Y \quad (3.125)$$

Burada, \mathbf{X}_0 ve \mathbf{Y}_0 ; **QR**-ayrışımındaki "**Q**"-matrisleridir, \mathbf{L}_X ve \mathbf{L}_Y uygun "**R**"-matrisleridir.

Değişmezlik özellikleri veri matrisleri arasındaki ilişkileri test etmenin problemini altuzaylar arasındaki ilişkileri test etmeye indirgedi. Bu altuzaylar ortonormal sütunların \mathbf{X}_0 ve \mathbf{Y}_0 matrisleri ile temsil edilirler. Bu nedenle iki altuzay arasındaki ilişkinin bir ölçüsü bir test istatistiği olarak kullanılabilir. Problem böyle bir ölçünün nasıl tanımlanacağıdır.

QR ayrışımı tam anlamıyla ortogonal tabanlar üreten bir ayrışımın bir örneğidir. Bir test istatistiği özel bir ayrışımına bağlı olmamalı. Başka bir deyişle, \mathbf{U}_X ($p \times p$) ve \mathbf{U}_Y ($q \times q$) keyfi ortogonal matrisler olmak üzere test istatistiği,

$$(\mathbf{X}_0, \mathbf{Y}_0) \rightarrow (\mathbf{X}_0 \mathbf{U}_X, \mathbf{Y}_0 \mathbf{U}_Y) \quad (3.126)$$

dönüşümü altında değişmemelidir (aynı kalmalıdır).

Değişmezliğin başka bir tipi normal dağılımdan görülür. (\mathbf{X}, \mathbf{Y}) nin eş anlı dağılımı $(\mathbf{X}, \mathbf{Y}) \rightarrow (\mathbf{R}\mathbf{X}, \mathbf{R}\mathbf{Y})$ dönüşümü altında değişmezdir. Burada **R** matrisi

$$\mathbf{R}\mathbf{j}_{(n \times 1)} = \mathbf{j}_{(n \times 1)} \quad (3.127)$$

eşitliğini sağlayan keyfi bir ortogonal matristir. \mathbf{X}_0 ve \mathbf{Y}_0 in eş anlı dağılımının

$$(\mathbf{X}_0, \mathbf{Y}_0) \rightarrow (\mathbf{R}\mathbf{X}_0, \mathbf{R}\mathbf{Y}_0) \quad (3.128)$$

dönüşümü altında değişmediği doğrudan doğruya görülür. **R** ye \mathcal{S} deki dönme yapan bir matris olarak başvurulur. Bu çalışmada, tüm ortogonal matrislere dönme matrisleri denir. Bir veri matrisiyle çarpılan bir dönme matrisinin sonucu döndürülmüş bir koordinat sisteminde veri sütun vektörlerinin bir gösterimi olarak

veya orijinal verinin döndürülmüş bir versiyonu olarak yorumlanabilir. Dönme matrislerinin ayrıntıları için “Ek 3.1” e bakınız.

Λ İstatistiği:

(3.126) ve (3.128) değişmezlik özelliklerini sağlayan bir test istatistiği aşağıdaki geometrik hacim düşüncesinden yola çıkarak inşa edilebilir. \mathbf{X}_0 in sütun vektörleri p –boyutlu bir küp oluşturur. Tüm ortogonal vektörler bir uzunluğa sahip olduklarından, küpün hacmi birdir. $j \leq i$ olmak üzere, genel bir \mathbf{M} $i \times j$ matrisi için, sütun vektörleri j –boyutlu bir paralel yüzlü oluşturur (Strang, 1988). Bu paralelyüzlünün hacmi

$$\text{Hacim}^2 = |\mathbf{M}'\mathbf{M}| \quad (3.129)$$

bağıntısından hesaplanabilir. Eğer tüm sütun vektörleri bir uzunluğa sahip olmaya kısıtlanırsa, bu takdirde, maksimal hacim birdir. Tüm sütun vektörleri karşılıklı olarak (ikişer ikişer) ortogonal olduğunda maksimum elde edilir ve sütun vektörleri arasındaki bağımlılık artarken hacim azalır. \mathbf{X}_0 ve \mathbf{Y}_0 arasındaki ilişkiyi ölçmek için bir düşünce \mathbf{Y}_0 in q tane sütun vektörü ile birlikte \mathbf{X}_0 in p tane sütun vektörü tarafından oluşturulan $(p+q < n)$, $(p+q)$ –boyutlu paralelyüzlünün hacmini hesaplamaktır. \mathbf{X}_0 in ve \mathbf{Y}_0 in her ikisi de ortonormal sütunlara sahip olduğundan, hacim \mathbf{X}_0 in sütunlarının \mathbf{Y}_0 in sütunlarına nasıl bağlı olduğunun bir fonksiyonudur. Ayrıca, biri $(p+q)$ –boyutlu paralelyüzlüye; \mathbf{X}_0 ile temsil edilen altuzay ve \mathbf{Y}_0 ile temsil edilen altuzay ile gerilmiş gibi bakabilir. Böylece, hacim iki altuzay arasındaki ilişkiyi ölçmek için bir yöntemdir. Hacmin karesini Λ ile gösteririz.

$$\Lambda = \left| \begin{bmatrix} \mathbf{X}_0 & \mathbf{Y}_0 \end{bmatrix}' \begin{bmatrix} \mathbf{X}_0 & \mathbf{Y}_0 \end{bmatrix} \right| \quad (3.130)$$

$$= \left| \mathbf{I}_p - \mathbf{X}_0' \mathbf{Y}_0 \mathbf{Y}_0' \mathbf{X}_0 \right| \quad (3.131)$$

$$= \left| \mathbf{I}_q - \mathbf{Y}_0' \mathbf{X}_0 \mathbf{X}_0' \mathbf{Y}_0 \right| \quad (3.132)$$

dır. Burada, (3.131) ve (3.132), $\mathbf{X}_0' \mathbf{X}_0 = \mathbf{I}_p$, $\mathbf{Y}_0' \mathbf{Y}_0 = \mathbf{I}_q$ olması gerçeği ile birlikte standart matris cebiri ile elde edilir.

Altuzaylardan birinin bir boyutlu olduğu özel durumlarda, Λ istatistiği farklı bir şekilde yorumlanabilir. $q=1$ olduğunda \mathbf{Y}_0 vektörü ve X – altuzayı arasındaki θ açısını hesaplayabiliriz. θ ve Λ arasındaki ilişkinin

$$\Lambda = \sin^2 \theta \quad (3.133)$$

olarak yazılabildiği çıkar.

Yukarıda ifade edilen Λ istatistiği tam olarak olabilirlik oran istatistiğinin basit bir dönüşümü olan Wilks'in Lamdası'dır. Test istatistiği

$$\Lambda = \frac{|\mathbf{S}_{yy} - \mathbf{S}_{yx} \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy}|}{|\mathbf{S}_{yy}|} = \frac{|\mathbf{I} - \mathbf{S}_{yy}^{-1} \mathbf{S}_{yx} \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy}|}{|\mathbf{S}_{yy}|} \quad (3.134)$$

olarak yazılır (Mardia, 1979). Burada

$$\begin{bmatrix} \mathbf{S}_{xx} & \mathbf{S}_{xy} \\ \mathbf{S}_{yx} & \mathbf{S}_{yy} \end{bmatrix} = \frac{1}{n} \begin{bmatrix} \mathbf{X}'_c \mathbf{X}_c & \mathbf{X}'_c \mathbf{Y}_c \\ \mathbf{Y}'_c \mathbf{X}_c & \mathbf{Y}'_c \mathbf{Y}_c \end{bmatrix} \quad (3.135)$$

dir. (3.124) - (3.125) bağıntılarından, (3.134) ve (3.130) – (3.132) ifadelerinin denk olduklarını gerçeklemek kolaydır.

Çizelge 3.6. $\lambda_1, \dots, \lambda_s$ özdeğerleri için tanımlanan başka matrisler

Tanımlanan matris	i – yinci özdeğer
$\mathbf{H}(\mathbf{H} + \mathbf{E})^{-1}$ veya $\mathbf{S}_{yy}^{-1} \mathbf{S}_{yx} \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy}$ veya $\mathbf{Y}'_0 \mathbf{X}_0 \mathbf{X}'_0 \mathbf{Y}_0$	λ_i
$\mathbf{E}(\mathbf{H} + \mathbf{E})^{-1}$ veya $\mathbf{I}_q - \mathbf{Y}'_0 \mathbf{X}_0 \mathbf{X}'_0 \mathbf{Y}_0$	$1 - \lambda_{q+1-i}$
$\mathbf{H}\mathbf{E}^{-1}$ veya $\mathbf{Y}'_0 \mathbf{X}_0 \mathbf{X}'_0 \mathbf{Y}_0 (\mathbf{I}_q - \mathbf{Y}'_0 \mathbf{X}_0 \mathbf{X}'_0 \mathbf{Y}_0)^{-1}$	$\lambda_i / (1 - \lambda_i)$

Çizelge 3.7. Dört değişmez test istatistiği ve karşılık gelen blok tasarım ölçütü

Test istatistiği	İfade	Tasarım ölçütleri
Wilks'in Λ sı	$\prod_{i=1}^s (1 - \lambda_i)$	D -optimallik
Roy'un en büyük kökü	λ_1	E -optimallik
Hotelling-Lawley iz istatistiği	$\sum_{i=1}^s (\lambda_i / (1 - \lambda_i))$	A -optimallik
Pillay-Bartlett iz istatistiği	$\sum_{i=1}^s \lambda_i$	M -optimallik

Diğer Değişmez İstatistikler:

İki değişmez özelliğin, yani, (3.126) ve (3.128) un herhangi bir test istatistiğinin $\mathbf{Y}'_0\mathbf{X}_0\mathbf{X}'_0\mathbf{Y}_0$ in $s = \min(p, q)$ sayıda sıfırdan farklı özdeğerinin bir fonksiyonu olarak yazılabildiğini ifade ettiğini gösterebilir. Bu özdeğerleri $\lambda_1, \dots, \lambda_s$ ile gösteririz ve Çizelge 3.6 bazı değişik tanımlanan matrisleri listeler. λ_i ler kanonik korelasyon katsayılarının kareleri olarak da bilinir (Gittins, 1985). En büyük böyle katsayı, yani $\sqrt{\lambda_1}$ merkezleştirilen iki veri kümesinden lineer kombinasyonların bir çifti arasındaki maksimal korelasyon katsayısıdır. Kanonik açılar yönünden, $\arccos(\sqrt{\lambda_1})$, iki gerçek altuzaydan vektörlerin bir çifti arasındaki en küçük açıdır. İkinci korelasyon katsayısı vektörleri, birinci çiftten olanlara ortogonal olmaya sınırlayarak elde edilebilir.

(3.130) Λ istatistiği

$$\Lambda = \prod_{i=1}^s (1 - \lambda_i) \quad (3.136)$$

olarak yazılabilir. (Olson, 1976) en popüler dört test istatistiğini gözden geçirir. Bunlar Çizelge 3.7 de listelenir. Bir konu dışı söz olarak, bu tablo dört test istatistiğine karşılık gelen dört blok tasarım ölçütünü de listeler [John, 1982; Williams, 1995]. İki kategorik tasarım değişkenini kabul ederiz. t sayıda seviyeye sahip bir işlem değişkeni ve b sayıda seviyeye sahip bir bloklama değişkeni. Sadece işlem seviyelerine sahip bir model sabit vektöre ortogonal olan $(t-1)$ sütun vektörüne sahip bir regresyon biçiminde yazılabilir. Benzer şekilde, bir blok etki modeli $(b-1)$ sayıda vektörle yazılabilir. Bir optimal tasarımda $(t-1)$ sayıda işlem vektörü tarafından gerilen altuzay ve $(b-1)$ sayıda blok vektörü tarafından gerilen altuzay mümkün olduğunca ortogonaldirler. İyi tasarım için bir ölçüt olarak, altuzaylar arasındaki ilişkinin bir ölçüsüne gereksinim duyulur. Blok tasarım ölçütleri ve test istatistiklerinin aynı olduklarına dikkat etmek ilginçtir.

Blok tasarım ölçütlerine daha yakından bakmak için t sayıda işlemin r kez tekrarlandığı ve b sayıda bloğun her birinin $k = tr/b$ boyuta sahip olduğu bir

tasarımı göz önüne alırız. Tasarım sütunları kukla değişkenler (0 veya 1) olan işlem ve bloklar için sırasıyla \mathbf{C} ($tr \times t$) ve \mathbf{D} ($tr \times b$) ile gösterilebilir.

$$\mathbf{N} = \{n_{ij}\} = \mathbf{C}'\mathbf{D} \quad (3.137)$$

matrisi sıklık (incidence) matrisidir. Burada n_{ij} i -yinci işlemin j -yinci blokta vuku bulma keresinin sayısını gösterir. Farklı blok tasarım ölçütleri

$$\mathbf{A} = \mathbf{I}_t - (1/rk)\mathbf{N}\mathbf{N}' \quad (3.138)$$

matrisinin sıfırdan farklı özdeğerleri olan, kanonik etki (yeterlik) faktörlerinin fonksiyonlarıdır. \mathbf{A} yı,

$$\mathbf{A} = \mathbf{I}_t - \left(\frac{1}{\sqrt{r}}\mathbf{C}\right)' \left(\frac{1}{\sqrt{k}}\mathbf{D}\right) \left(\frac{1}{\sqrt{r}}\mathbf{D}\right)' \left(\frac{1}{\sqrt{r}}\mathbf{C}\right) \quad (3.139)$$

olarak yeniden yazabiliriz. \mathbf{A} için bu ifade, $\lambda_1, \dots, \lambda_s$ özdeğerleri için değişik tanımlanan matrisler olan (3.131) – (3.132) deki matrislere (Çizelge 3.6) çok benzerdir. [John, 1982; Williams, 1995] de ifade edildiği gibi, farklı blok tasarım ölçütlerinin test istatistiklerinin (Çizelge 3.7) denk oldukları doğrudan doğruya

görülür. \mathbf{X}_0 ve \mathbf{Y}_0 a benzer şekilde, $\left(\frac{1}{\sqrt{r}}\mathbf{C}\right)$ ve $\left(\frac{1}{\sqrt{r}}\mathbf{D}\right)$ nin ortonormal sütunlara sahip olduklarına dikkat ediniz. (3.139) ve (3.131) – (3.132) deki matrisler arasındaki yegane fark $\left(\frac{1}{\sqrt{r}}\mathbf{C}\right)$ ve $\left(\frac{1}{\sqrt{r}}\mathbf{D}\right)$ nin sütunlarının sabit vektöre ortogonal olmadığıdır. Gerçekten, sabit vektör bu matrislerin her ikisinin de sütun uzayında içerilir. Sonuç \mathbf{A} nın ilaveten sıfır olan bir özdeğere sahip olmasıdır.

Test istatistiklerinin sıfır dağılımı:

Değişmez test istatistiklerinin iki ortogonal matrisin, yani, \mathbf{X}_0 ve \mathbf{Y}_0 ın bir fonksiyonu olarak yazılabildiğini hatırlayınız. (3.128) değişmezliği x ve y nin eş anlı çoklu normal dağılımından çıkar. Sıfır hipotezi altında x ve y bağımsız olarak dağılır. Bu nedenle (3.128) değişmezliği x için ve y için ayrı ayrı yazılabilir. Başka bir deyişle, \mathbf{X}_0 ve \mathbf{Y}_0

$$(\mathbf{X}_0, \mathbf{Y}_0) \rightarrow (\mathbf{R}_1 \mathbf{X}_0, \mathbf{R}_2 \mathbf{Y}_0) \quad (3.140)$$

dönüşümü altında değişmezdir. Burada \mathbf{R}_1 ve \mathbf{R}_2 ; \mathcal{S} deki dönmeleri yapan iki keyfi dönme matrisidir. Değişmez (invariant) dağılımların genel teorisinden (Dempster, 1969), test istatistiklerinin sıfır dağılımının bu değişmezlik özelliği ile tek bir şekilde karakterize edildiği görülür. Genel bir test istatistiğinin, yani, $G(\mathbf{X}_0, \mathbf{Y}_0)$ ın sıfır dağılımına

$$G(\mathbf{R}_1 \mathbf{X}_0, \mathbf{R}_2 \mathbf{Y}_0) \quad (3.141)$$

ın dağılımı olarak bakılabilir. Burada \mathbf{X}_0 ve \mathbf{Y}_0 sabittir ve \mathbf{R}_1 ve \mathbf{R}_2 , \mathcal{S} deki iki bağımsız rasgele dönmeyi gösterir (bkz. Ek 3.1). Bu dağılım sezgisel bir geometrik yoruma sahiptir. Test istatistiği iki altuzay (\mathbf{X}_0 ve \mathbf{Y}_0 ile gösterilen) arasındaki ilişkinin bir ölçüsüdür ve sıfır dağılımı iki rasgele yönlendirilen altuzay ($\mathbf{R}_1 \mathbf{X}_0$ ve $\mathbf{R}_2 \mathbf{Y}_0$) arasındaki ilişkiden izlenir.

Bununla beraber, tüm değişmez test istatistikleri $\mathbf{X}_0' \mathbf{Y}_0$ vasıtasıyla \mathbf{X}_0 ve \mathbf{Y}_0 a bağlıdır. Başka bir deyişle tüm test istatistikleri bir $\tilde{G}(\mathbf{X}_0' \mathbf{Y}_0)$ fonksiyonu olarak yazılabilir. Bu nedenle, sıfır dağılımını

$$\tilde{G}(\mathbf{X}_0' \mathbf{R}_1' \mathbf{R}_2 \mathbf{Y}_0) \quad (3.142)$$

dağılımı olarak gösterebiliriz. Rasgele dönme matrislerinin özelliklerinden (bkz. Ek 3.1), $\mathbf{R}_1' \mathbf{R}_2$ çarpım matrisinin de \mathcal{S} de bir rasgele dönme yapan bir matris olduğu görülür. Ayrıca, çarpım matrisinin rasgele olması için iki matristen birinin, yani, \mathbf{R}_1 ve \mathbf{R}_2 den birinin rasgele olması yeter. (Fisher, 1939) tarafından belirtildiği gibi; x in normal dağılmış olduğu varsayımının düşürülebildiği görülür. Sıfır dağılımı kümelerden biri normal dağıldığı sürece aynı olacak. Geometrik olarak, bu; iki altuzaydan biri rasgele olarak yönlendirildiği sürece, iki altuzay arasındaki ilişkinin rasgele olduğunu söyler.

Wilks'in Λ – istatistiğinin sıfır dağılımı Wilks'in lamdası olarak bilinir ve üç parametreye sahiptir. Bu parametreler \mathcal{S} “çalışma uzayının” boyutuna ve iki

altuzayın boyutlarına karşılık gelir. Bununla beraber, bu parametreler için notasyon standartlaştırılmış değildir [Mardia ve ark., 1979; Kshirsagar, 1972]

Althipotezleri test etme:

Şu ana kadar ifade edilen test, değişkenlerin iki kümesi arasındaki geniş kapsamlı ilişkiyi ilgilendirir. Fakat, birtakım durumda biri çoğu kez özel x – değişkenlerine ilişkin althipotezleri test etmeyle ilgilendirir. Böyle testler geniş kapsamlı testlere çok benzeyen bir biçimde ifade edilebilir. İlk olarak, son matris hipotez için değişkenlerden ibaret olacak şekilde, \mathbf{X} matrisi iki \mathbf{X}_1 ($n \times p_1$) ve \mathbf{X}_2 ($n \times p_2$) kısma ayrılır. Şimdi, \mathcal{S} üzerine bir izdüşüm olan merkezileştirme adımı sabit vektörle birlikte \mathbf{X}_1 in sütunları tarafından gerilen uzayın ortogonal tümleyeni olan, \mathcal{S}^+ üzerine izdüşümle yer değiştirilir. Hedeflenen veriyi

$$\mathbf{X}_c^+ = \mathbf{X}_2 - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{X}_2 \quad (3.143)$$

$$\mathbf{Y}_c^+ = \mathbf{Y}_2 - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{Y} \quad (3.144)$$

olarak ifade edebiliriz. Burada

$$\mathbf{Z} = [\mathbf{j}_{n \times 1} \quad \mathbf{X}_1] \quad (3.145)$$

dir. Test istatistiğinin sıfır dağılımını formüllemek için \mathcal{S} deki dönmeler \mathcal{S}^+ daki dönmelerle yer değiştirir. Şimdi, “ X – altuzayı ” p yerine p_2 boyutuna sahiptir ve içinde çalışıyor olduğumuz uzay, yani, \mathcal{S}^+ ; $(n-1)$ yerine $(n-1-p_1)$ boyutuna sahiptir. x – değişkenlerinin bir lineer kombinasyonu ile ilgili hipotezlerin testleri, kural olarak x – değişkenlerini yeniden parametrelendirerek benzer bir biçimde inşa edilebilir. $\mathbf{j}_{n \times 1}$ vektörünün \mathbf{Z} matrisinin bir parçası olması gerekmediğine de dikkat ediniz.

Diğer modeller altında test etme:

Test istatistiklerinin sıfır dağılımında ifade edildiği gibi x in normal dağılımlı olduğu varsayımı düşürülebilir. Aşağıdaki çok değişkenli çoklu regresyon ve çok değişkenli varyans analizi kısımları x sabit olmak üzere normal olmayan x in özel durumundan bahseder.

Çok değişkenli çoklu regresyon:

\mathbf{B}'_0 sabit terimlerin (interceptlerin) vektörü ve \mathbf{B} diğer regresyon parametrelerinin matrisi olmak üzere,

$$\mathbf{Y} = \mathbf{j}_{(n \times 1)} \mathbf{B}_0 + \mathbf{X}\mathbf{B} + \mathbf{F} \quad (3.146)$$

ile tanımlanan alışlagelen çok değişkenli çoklu regresyon modelini kabul ediniz. \mathbf{F} matrisi satırları sıfır ortalamalı ve Σ ortak kovaryans matrisli ilişkisiz çoklu normal dağılıma sahip olan rasgele hatalardan oluşur.

$$H_1 : \mathbf{B} \neq \mathbf{0} \text{ hipotezine karşı } H_0 : \mathbf{B} = \mathbf{0} \quad (3.147)$$

hipotezi için, Wilks'in Λ sı, bu durumda

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{H} + \mathbf{E}|} \quad (3.148)$$

olarak yazılır. (Mardia, 1979). Burada, \mathbf{E} ve \mathbf{H} matrisleri H_0 ve H_1 altında Σ nın en çok olabilirlik tahminlerinde aşağıdaki biçimde tanımlanabilir:

$$\mathbf{E} + \mathbf{H} = n\hat{\Sigma}_{H_0} = \mathbf{L}'_Y \mathbf{L}_Y \quad (3.149)$$

$$\mathbf{E} = n\hat{\Sigma}_{H_1} = \mathbf{L}'_Y (\mathbf{I}_q - \mathbf{Y}'_0 \mathbf{X}_0 \mathbf{X}'_0 \mathbf{Y}_0) \mathbf{L}_Y \quad (3.150)$$

(3.148) ifadesi sadece daha evvel bahsedilen (3.130) test istatistiğini yazmak için başka bir yoldur. \mathbf{Y} çoklu normal olduğundan test istatistiği Wilks'in lamda dağılımına uyar. (3.148) ifadesinin, Wilks'in lamda dağılımının Wishart dağılımlarına dayanan ortak dağılımına karşılık geldiğine de dikkat ediniz (Mardia, 1979).

Bir y –değişkenine sahip olduğumuzda denk F istatistiğini kullanmak yaygındır:

$$F = \frac{n-1-p}{p} \cdot \frac{1-\Lambda}{\Lambda} \quad (3.151)$$

θ için bağıntı (3.3.14) ten çıkar:

$$F = \frac{n-1-p}{p} \cdot \frac{1}{\tan^2 \theta} \quad (3.152)$$

Bu durumda (Box ve Draper, 1987) geometrinin mükemmel örneklerini sunarlar. Onların şekillerinde anlam düzeyi bir küre yüzeyi alanının bir oranıdır.

Bu regresyon modeli altında althipotezlerin testleri alt hipotezleri test etme kısmında ifade edildiği gibi formüllenebilir.

Çok değişkenli varyans analizi:

Tüm MANOVA (çok değişkenli varyans analizi) kural olarak çok değişkenli regresyon modelleridir. p , model için serbestlik derecesine eşit olmak üzere, model daima bir regresyon biçiminde yazılabilir. \mathbf{X}_0 in sütun vektörleri bu durumda ortogonal kontrastlar için katsayı vektörleridir. Bir-yönlü MANOVA da g sayıda grup karşılaştırıldığında, bir olabilirlik $(g-1)$ sayıda ilk gruba karşılık gelen kukla değişkenleri (0 veya 1) X –değişkenleri olarak kullanmaktır. Bir-yönlü MANOVA durumunda, \mathbf{E} hata kareler toplamı ve çapraz çarpımların matrisidir ve \mathbf{H} işlem kareler toplamı ve çapraz çarpımların matrisidir (Johnson ve Wichern, 1992). Bu iki matris çoğu kez “ \mathbf{W} ” ve “ \mathbf{B} ” ile gösterilir.

$g = 2$ olduğunda, denk Hotelling T^2 istatistiğini kullanmak yaygındır:

$$T^2 = (n-2) \frac{1-\Lambda}{\Lambda} \quad (3.153)$$

Şimdi, \mathbf{X}_0 tek bir vektördür. Bu nedenle bu vektör ve y –uzayı arasındaki θ açısına başvurabiliriz. T^2 ve θ arasındaki ilişki:

$$T^2 = \frac{n-2}{\tan^2 \theta} \quad (3.154)$$

Bu bakiş açısından, Hotelling T^2 ve F dağılımının aynı (bir orantılılık katsayısı hariç) oldukları açıktır. Her iki durumda bir altuzay ve bir vektör arasında herhangi bir ilişki olup olmadığını test ederiz. Fark, bir F –testinde \mathbf{Y}_0 vektördür, fakat T^2 – testi için , burada \mathbf{X}_0 vektördür.

$H_0 : \mu = \mu_0$ hipotezi:

Şimdi, değişkenlerden sadece biri göz önüne alınır ve çoklu normallik kabul edilir:

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (3.155)$$

dir.

$$H_1 : \mu \neq \mu_0 \text{ hipotezine karşı } H_0 : \mu = \mu_0 \quad (3.156)$$

hipotezi için, olabilirlik oran testi, Hotelling T^2 test istatistiğine dayanır (Mardia, 1979).

$$T^2 = (n-1)(\bar{\mathbf{Y}} - \boldsymbol{\mu}_0)' \mathbf{S}^{-1} (\bar{\mathbf{Y}} - \boldsymbol{\mu}_0) \quad (3.157)$$

Burada,

$$\bar{\mathbf{Y}} = \frac{1}{n} \mathbf{j}'_{(n \times 1)} \mathbf{Y} \quad (3.158)$$

dir ve \mathbf{S} ; $\boldsymbol{\Sigma}$ nın en çok olabilirlik tahminidir. Bu T^2 istatistiği geometrik olarak Λ veya θ olarak yorumlanabilir:

$$T^2 = (n-1) \frac{1-\Lambda}{\Lambda} = \frac{n-1}{\tan^2 \theta} \quad (3.159)$$

Λ istatistiği ilk olarak tüm orijinal \mathbf{Y} gözlemlerinden varsayılan $\boldsymbol{\mu}_0$ ortalamasını çıkararak hesaplanabilir. Ortalamayı test ediyor olduğumuzdan, \mathbf{QR} ayrışımında önce merkezileştirme yapılmaz. $\frac{1}{\sqrt{n}} \mathbf{j}_{(n \times 1)}$ birim vektörü bu durumda \mathbf{X}_0 matrisi gibi çalışacak. Bu durumda θ ; bu vektör ve $\mathbf{Y} - \boldsymbol{\mu}_0$ ile gerilen uzay arasındaki açıdır. Küçük bir θ ; $\boldsymbol{\mu}_0$ nın gerçek bir ortalama olmadığını gösterir. Bu testin formüllemesi kısıtlanmış altuzay olmayan rasgele dönmelere dayanır (3.127). Şimdi çalışma uzayımızın $(n-1)$ -boyutlu \mathcal{S} altuzayının yerine n -boyutlu tüm uzay olduğunu söyleyebiliriz. Bu; (3.159) daki orantılılık sabitinin niçin (3.153) ve (3.154) den farklı olduğunu açıklar.

Son yorumlar:

Bu çalışmada, \mathbf{Y} nin (veya \mathbf{X} in) birçok değişkenli normal dağılımı izlediğini kabul ettik. Bununla beraber, sonuçlar $\mathbf{Y} \rightarrow \mathbf{R}\mathbf{Y}$ (burada \mathbf{R} ; \mathcal{S} deki bir dönme) gösterir dönüşümü altında değişmez olan \mathbf{Y} nin herhangi dağılımları için geçerlidir. Bu değişmezlik satırların bağımsızlığı ile birlikte normalliği ifade eder (James, 1954). Bu nedenle, teoremin pratiksel uygunluğu çok değişkenli normal dağılan veri için çoğu kez olacak. Bununla beraber, (Dawid, 1977) genelleştirilmiş dağılımların “doğal olarak ortaya çıktığına” değindi. Böyle genelleştirilmiş dağılımlar ve diğer ilgili dağılımlar için teori (Fang ve ark., 1990) tarafından kapsamlı olarak anlatılır.

EK 3.1 Rasgele dönme matrisleri

Bir dönme matrisi 1 determinantlı bir ortogonal matristir (Curtis, 1979). Bir ortogonal matrisin determinantı daima 1 veya -1 dir. Bununla beraber, bu çalışmada dönme matrislerinden bahsedildiğinde, tüm ortogonal matrisleri ifade edeceğiz.

Bir ortogonal $n \times n$ matrisin sütunları (ve aynı zamanda satırları) n – boyutlu uzay için bir taban oluşturur. Bir rasgele dönmede, birinci sütun bir rasgele birim vektördür. Birinci sütun verildiğinde, ikinci sütun birinciye ortogonal bir rasgele birim vektördür. Bir rasgele dönme matrisini oluşturmak için basit bir yöntem n^2 sayıda bağımsız standart normal sapmalarla bir $n \times n$ matrisi doldurmak ve sonra da **QR** ayrışımını yapmaktır. Bu durumda ortaya konan **Q** – matrisi bir rasgele dönme matrisidir. Transpoz (=ters) matrisin de bir rasgele dönme matrisi olduğuna dikkat ediniz. Bundan başka, başka bir dönme matrisiyle (sabit veya bağımsız rasgele) çarpılan bir $n \times n$ rasgele dönme matrisinin yine bir rasgele dönme matrisi olduğunu görmek kolaydır. Bir rasgele dönme matrisinin, bir rasgele “düzgün dağılan” ortogonal matris olarak bilindiğine de dikkat ediniz.

Bir $i < n$ boyutlu uzayda rasgele dönmeyi yapan bir matris, \mathbf{U}_1 sütunları dönme için altuzaya bir ortonormal taban oluşturan bir $n \times i$ matris olmak üzere,

$$\mathbf{R} = \mathbf{U}_1 \tilde{\mathbf{R}} \mathbf{U}_1' + \mathbf{U}_2 \mathbf{U}_2' \quad (3.160)$$

olarak inşa edilebilir. Benzer biçimde U_2 tümleyen için bir tabanı içerir. $\tilde{\mathbf{R}}$ $i \times i$ matrisi alışlagelen bir rasgele dönme matrisidir. Belirli (özel) bir altuzayda dönme yapan matrislerin bir sınıfı transpoz işlemi ve aynı sınıftan matrisler ile çarpma altında kapalıdır. Bu; “ \mathbf{R} – matrisleri” üzerindeki bu işlemlerin “ $\tilde{\mathbf{R}}$ – matrisleri” üzerindeki benzer işlemlere indirgenebilmesi gerçeğinden görülebilir. \mathcal{S} altuzayında, sabit vektöre ortogonal olan, dönmeyi yapan dönme matrislerinin sınıfı bu çalışmada önemli bir rol oynar. Bu sınıfın tüm permütasyon matrislerini kapsadığına da dikkat ediniz.



4. KOLAY KORELASYON BİLGİLERİ

Bu bölümde, Puntanen, S., Styan, G.P.H. Isolato, J. (2011) örneklem korelasyon katsayısının geometrik yorumu olan temel bir gerçeği ele aldı. Bu temel gerçek okuyucu için neredeyse hiç beklenmedik bir bilgi değildir. Fakat önemi nedeniyle onun başlı başına ele alınması değere şayandır.

\mathbf{U} $n \times 2$ veri matrisi

$$\mathbf{U} = (\mathbf{x} : \mathbf{y}) = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix} = \begin{pmatrix} \mathbf{u}'_{(1)} \\ \mathbf{u}'_{(2)} \\ \vdots \\ \mathbf{u}'_{(n)} \end{pmatrix} \quad (4.1)$$

olarak parçalanmış olsun. Burada $\mathbf{u}_{(i)} = \begin{pmatrix} x_i \\ y_i \end{pmatrix} \in \mathbb{R}^2$ gözlem uzayındaki i -yinci durumu veya i -yinci gözlemi gösterir, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ vektörleri değişken uzayındaki iki değişkeni gösterir. $\bar{\mathbf{u}} = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} \in \mathbb{R}^2$, x - ve y -değişkenlerinin ortalama vektörünü ve

\mathbf{S} örneklem kovaryans matrisini gösterebilir:

$$\bar{\mathbf{u}} = \frac{1}{n} (\mathbf{u}_{(1)} + \mathbf{u}_{(2)} + \cdots + \mathbf{u}_{(n)}) = \frac{1}{n} \mathbf{U}' \mathbf{j}_n = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} \quad (4.2a)$$

$$\mathbf{S} = \frac{1}{n-1} \mathbf{U}' \mathbf{C} \mathbf{U} = \frac{1}{n-1} \mathbf{T} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{u}_{(i)} - \bar{\mathbf{u}}) (\mathbf{u}_{(i)} - \bar{\mathbf{u}})' \quad (4.2b)$$

dür. Devamında, \mathbf{j}_n vektöründen çoğu kez alt indisi kaldırırız, içerikten onun boyutu açık olmalı. Burada \mathbf{C} ; $\mathbf{C} = \mathbf{I}_n - \mathbf{P}_j$ biçimindeki $n \times n$ merkezileştirme matrisidir ve

$$\mathbf{P}_j = \mathbf{j}(\mathbf{j}'\mathbf{j})^{-1} \mathbf{j}' = \frac{1}{n} \mathbf{j}\mathbf{j}' \text{ dür.}$$

Teorem 4.1 (Geometrik olarak korelasyon katsayısı): Uygun (çarpılabilir) matrisler için aşağıdaki ifade gerçekleşir. (Styan ve ark., 2006).

- (a) $\bar{\mathbf{y}} = \bar{y}\mathbf{j}$ vektörü, \mathbf{y} değişken vektörünün $\mathcal{C}(\mathbf{j})$ sütun uzayı üzerine ortogonal izdüşümüdür, yani,

$$\bar{\bar{\mathbf{y}}} = \bar{\mathbf{y}}\mathbf{j} = \mathbf{P}_j\mathbf{y} \quad (4.3)$$

dir.

- (b) $\tilde{\mathbf{y}}$ merkezleştirilmiş vektörü \mathbf{y} nin $\mathcal{C}(\mathbf{j})^\perp$ sütun uzayı üzerine ortogonal izdüşümüdür, yani,

$$\tilde{\mathbf{y}} = \mathbf{y} - \mathbf{P}_j\mathbf{y} = \mathbf{C}\mathbf{y} = (\mathbf{I} - \mathbf{P}_j)\mathbf{y} = \mathbf{P}_{j^\perp}\mathbf{y} = \mathbf{P}_C\mathbf{y} \quad (4.4)$$

dir. Burada $\mathbf{P}_j = \mathbf{j}(\mathbf{j}'\mathbf{j})^{-1}\mathbf{j}' = \frac{1}{n}\mathbf{j}\mathbf{j}'$, $\mathcal{C}(\mathbf{j})$ üzerine ortogonal izdüşüm matrisi ve $(\mathbf{I} - \mathbf{P}_j)$, $\mathcal{C}(\mathbf{j})^\perp$ üzerine yani \mathbf{j} nin sütun uzayının tümleyeni üzerine ortogonal izdüşüm matrisidir.

- (c) x ve y değişkenlerinin varyansları sıfır olmasın, yani, $\mathbf{x} \notin \mathcal{C}(\mathbf{j})$ ve $\mathbf{y} \notin \mathcal{C}(\mathbf{j})$ olsun. Bu takdirde r_{xy} örneklem korelasyon katsayısı merkezleştirilmiş değişken vektörleri arasındaki açının kosinüsüdür, yani,

$$\begin{aligned} r_{xy} &= \text{kor}_s(x, y) = \text{kor}_d(\mathbf{x}, \mathbf{y}) \\ &= \cos(\mathbf{C}\mathbf{x}, \mathbf{C}\mathbf{y}) = \cos(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \\ &= \frac{\mathbf{x}'\mathbf{C}\mathbf{y}}{\sqrt{\mathbf{x}'\mathbf{C}\mathbf{x} \cdot \mathbf{y}'\mathbf{C}\mathbf{y}}} \end{aligned} \quad (4.5)$$

dir.

- (d) \mathbf{y} nin merkezleştirilmiş olması için gerek ve yeter şart $\mathbf{y} \in \mathcal{C}(\mathbf{j})^\perp = \mathcal{C}(\mathbf{C}) = \mathcal{N}(\mathbf{j}')$ olmasıdır.

İspat. İspatların tümü \mathbf{P}_j ve \mathbf{C} izdüşürücülerinin tanımlarından doğrudan doğruya ortaya çıkar. Aynı zamanda bkz. Şekil 4.1 (s.108). Somutluk için, y nin y_1, y_2, \dots, y_n gözlenen değerlerini göz önüne alabilir ve

$$f(\alpha) = \sum_{i=1}^n (y_i - \alpha)^2 \quad (4.6)$$

yi minimumlaştıracak şekilde $\alpha \in \mathbb{R}$ yi bulmaya çalışırız. Çözüm,

$$\begin{aligned} \min_{\alpha} f(\alpha) &= \min_{\alpha} \sum_{i=1}^n (y_i - \alpha)^2 = \min_{\alpha} \sum_{i=1}^n \|\mathbf{y} - \alpha \mathbf{j}\|^2 = \min_{\mathbf{u} \in \mathcal{C}(\mathbf{j})} \|\mathbf{y} - \mathbf{u}\|^2 \\ &= \|\mathbf{y} - \mathbf{P}_{\mathbf{j}} \mathbf{y}\|^2 = \|\mathbf{y} - \bar{y} \mathbf{j}\|^2 = \sum_{i=1}^n (y_i - \bar{y})^2 \end{aligned} \quad (4.7)$$

den çıkar. Böylece ispat tamamlanır.

4.1 Ortogonalite ve İlişkizlik

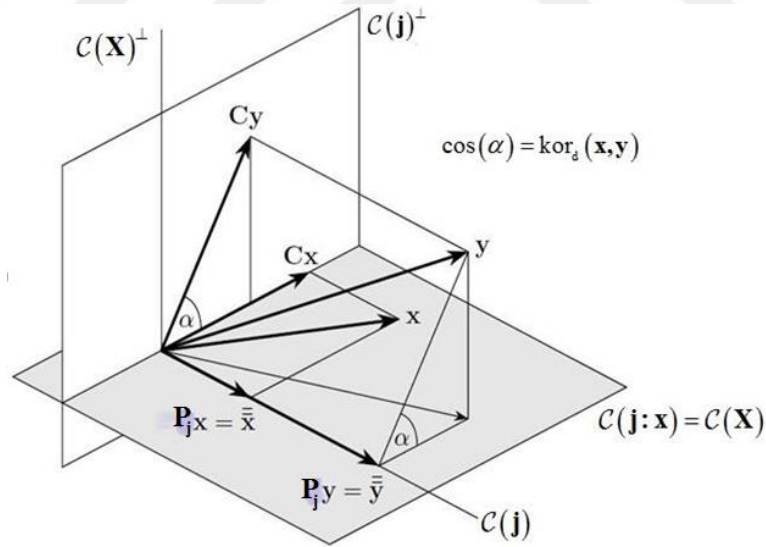
Asıl gerekli olan ortogonalite ve ilişkizliğin ister istemez aynı kavramlar olmadıklarını hatırlatmaktadır. Aşağıdaki yardımcı teorem bazı ilgili özellikleri özetler.

Yardımcı Teorem 4.1: \mathbf{x} ve \mathbf{y} sıfır olmayan n -boyutlu değişken vektörler olsun.

Bu takdirde

- (a) $\cos(\mathbf{x}, \mathbf{y}) = \text{yüksek}$, fakat $\text{kor}_d(\mathbf{x}, \mathbf{y}) = 0$
- (b) $\cos(\mathbf{x}, \mathbf{y}) = 0$, fakat $\text{kor}_d(\mathbf{x}, \mathbf{y}) = 1$

olması olasıdır. (Styan ve ark., 2006).



Şekil 4.1. Geometrik olarak \mathbf{x} ve \mathbf{y} arasındaki korelasyon

Bundan başka, $\mathbf{x} \in \mathcal{C}(\mathbf{j})$ ve $\mathbf{y} \notin \mathcal{C}(\mathbf{j})$ olsun. Bu takdirde

$$\text{kor}_d(\mathbf{x}, \mathbf{y}) = 0 \Leftrightarrow \mathbf{y} \in \mathcal{C}(\mathbf{Cx})^\perp = \mathcal{C}(\mathbf{j}:\mathbf{x})^\perp \oplus \mathcal{C}(\mathbf{j}) \quad (4.8)$$

olmasıdır. Burada \oplus ilgili vektörlerin direkt toplamını gösterir.

İspat. (a) durumu için, (Belsley, 1991) korelasyonun sıfır fakat, kosinüsün 1 e çok yakın olduğu bir açıklayıcı örneğe sahiptir. Bu amaç için,

$$(\mathbf{x} : \mathbf{y}) = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 0 & -2 \end{pmatrix} \quad (4.9)$$

veri matrisini göz önüne alınız. Burada, $\mathbf{x}, \mathbf{y} \in \mathcal{C}(\mathbf{C}) = \mathcal{N}(\mathbf{j}')$ ve $\mathbf{x}'\mathbf{y} = 0$, yani, \mathbf{x} ve \mathbf{y} merkezleştirilmişlerdir ve ilişkisizdirler: $\text{kor}_d(\mathbf{x}, \mathbf{y}) = 0$ dır. α sıfırdan farklı bir reel sayı olmak üzere,

$$\mathbf{t} = \mathbf{x} + \alpha \mathbf{j}, \mathbf{u} = \mathbf{y} + \alpha \mathbf{j} \quad (4.10)$$

olacak şekilde \mathbf{t} ve \mathbf{u} değişkenlerini tanımlayınız. Bu takdirde, $\text{kor}_d(\mathbf{t}, \mathbf{u}) = 0$, fakat

$$\cos(\mathbf{t}, \mathbf{u}) = \frac{3\alpha^2}{\sqrt{(\mathbf{x}'\mathbf{x} + 3\alpha^2)(\mathbf{y}'\mathbf{y} + 3\alpha^2)}} = \frac{3}{\sqrt{\left(\frac{\mathbf{x}'\mathbf{x}}{\alpha^2} + 3\right)\left(\frac{\mathbf{y}'\mathbf{y}}{\alpha^2} + 3\right)}} \quad (4.11)$$

dür ki bu α yı çok büyük seçerek keyfi olarak 1 e yakın yapabilir. Aynı zamanda, $\alpha \neq 0$ olmak üzere,

$$\mathbf{v} = \alpha \mathbf{x} + \mathbf{j}, \mathbf{w} = \alpha \mathbf{y} + \mathbf{j} \quad (4.12)$$

değişkenlerini de düşünebilirdik, ki bu durumda $\text{kor}_d(\mathbf{v}, \mathbf{w}) = 0$ dır, fakat

$$\cos(\mathbf{v}, \mathbf{w}) = \frac{3}{\sqrt{(\mathbf{x}'\mathbf{x}\alpha^2 + 3)(\mathbf{y}'\mathbf{y}\alpha^2 + 3)}} \quad (4.13)$$

dür. Şimdi α , 0 a giderken $\cos(\mathbf{v}, \mathbf{w})$ 1 e yaklaşır. (b) için aşağıdaki veri

$$(\mathbf{x} : \mathbf{y}) = \begin{pmatrix} 0 & -1 \\ 0 & -1 \\ 1 & 0 \end{pmatrix} \quad (4.14)$$

örneğini ortaya koyar. Bu durumda (4.8) bağlantısını düşününüz.

$$\mathbf{x}'\mathbf{C}\mathbf{y} = 0 \Leftrightarrow \mathbf{y} \in \mathcal{N}(\mathbf{x}'\mathbf{C}) = \mathcal{C}(\mathbf{C}\mathbf{x})^\perp \quad (4.15)$$

dir. Buradan sonra,

$$\mathcal{C}(\mathbf{C}\mathbf{x}) = \mathcal{C}(\mathbf{j}:\mathbf{x}) \cap \mathcal{C}(\mathbf{j})^\perp = \mathcal{C}(\mathbf{j}:\mathbf{x}) \cap \mathcal{C}(\mathbf{C}) \quad (4.16)$$

olduğunu gösteririz. Şimdi,

$$\mathbf{C}\mathbf{x} = \mathbf{x} - \bar{x}\mathbf{j} = (\mathbf{j}:\mathbf{x}) \begin{pmatrix} -\bar{x} \\ 1 \end{pmatrix} \quad (4.17a)$$

$$(\mathbf{j}:\mathbf{C}\mathbf{x}) = (\mathbf{j}:\mathbf{x}) \begin{pmatrix} 1 & -\bar{x} \\ 0 & 1 \end{pmatrix} = (\mathbf{j}:\mathbf{x}) \mathbf{A} \quad (4.17b)$$

dır. Burada $\mathbf{A} = \begin{pmatrix} 1 & -\bar{x} \\ 0 & 1 \end{pmatrix}$ matrisi tekil değildir ve bundan dolayı (daha sonra da kullanılacak çok yararlı bir gerçek)

$$\mathcal{C}(\mathbf{j}:\mathbf{x}) = \mathcal{C}(\mathbf{j}:\mathbf{C}\mathbf{x}) \quad (4.18)$$

elde ederiz. Şimdi, $\mathbf{C}\mathbf{x} \in \mathcal{C}(\mathbf{j}:\mathbf{C}\mathbf{x}) \cap \mathcal{C}(\mathbf{C})$ olduğu açıktır. Ters kapsamı göstermek için,

$$\mathbf{u} \in \mathcal{C}(\mathbf{j}:\mathbf{C}\mathbf{x}) \cap \mathcal{C}(\mathbf{C}) \quad (4.19)$$

ifadesini düşünelim, ki bu ifade, herhangi $a, b \in \mathbb{R}$ ve $\mathbf{c} \in \mathbb{R}^n$ için

$$\mathbf{u} = a\mathbf{j} + b\mathbf{C}\mathbf{x} = \mathbf{C}\mathbf{c} \quad (4.20)$$

olduğunu ifade eder. (4.20) eşitliğini \mathbf{C} ile önden çarpma $\mathbf{u} = b\mathbf{C}\mathbf{x} \in \mathcal{C}(\mathbf{C}\mathbf{x})$ olduğunu gösterir, böylece (4.16) ispatlanır.

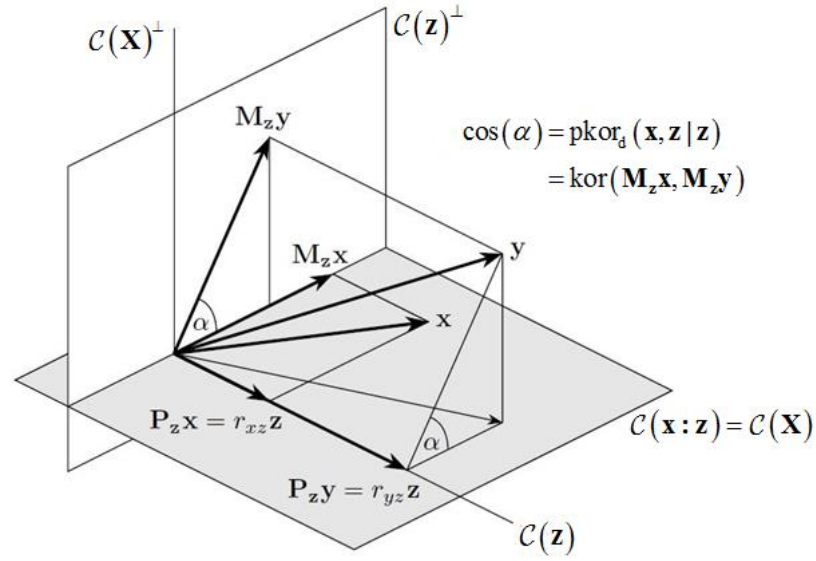
$$\mathcal{C}(\mathbf{C}\mathbf{x})^\perp = \mathcal{C}(\mathbf{j}:\mathbf{x})^\perp \oplus \mathcal{C}(\mathbf{j}) \quad (4.21)$$

iddiası (4.16) nın her bir yanının ortogonal tümleyenlerini alarak görülür. Böylece ispat tamamlanır. Bu arada (4.18) bağıntısının

$$\begin{aligned} \mathbf{P}_{(\mathbf{j}:\mathbf{x})} &= \mathbf{P}_{(\mathbf{j}:\mathbf{C}\mathbf{x})} = (\mathbf{j}:\mathbf{C}\mathbf{x}) \begin{pmatrix} n^{-1} & 0 \\ 0 & (\mathbf{x}'\mathbf{C}\mathbf{x})^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{j} \\ \mathbf{C}\mathbf{x} \end{pmatrix} \\ &= \mathbf{P}_{\mathbf{j}} + \mathbf{C}\mathbf{x}\mathbf{x}'\mathbf{C}/\mathbf{x}'\mathbf{C}\mathbf{x} = \mathbf{P}_{\mathbf{j}} + \mathbf{P}_{\mathbf{C}\mathbf{x}} \end{aligned} \quad (4.22a)$$

$$\mathbf{I} - \mathbf{P}_{(\mathbf{j}:\mathbf{x})} = \mathbf{C} - \mathbf{P}_{\mathbf{C}\mathbf{x}} \quad (4.22b)$$

ifade ettiğine dikkat edebiliriz. (4.22) ayrışımı aynı zamanda “matrix tricks for linear statistical models” Teorem 8 (s.155) den de bir kez daha görülür.



Şekil 4.2. Geometrik olarak $r_{xy.z}$ kısmi korelasyonu

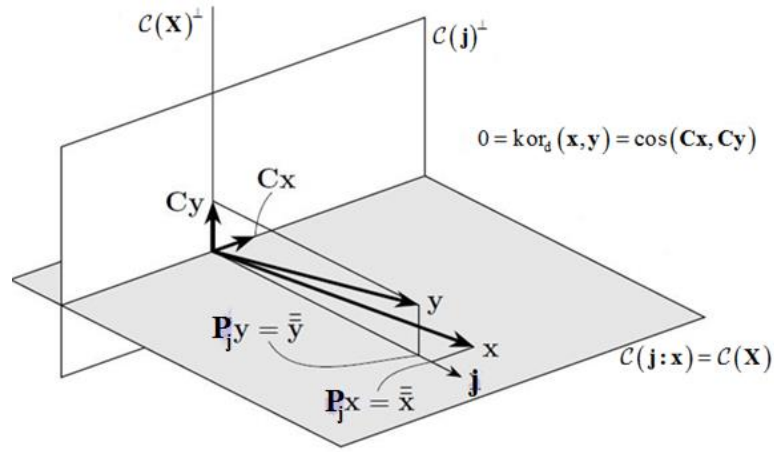
4.2 j Vektörünün Rolü Üzerine

Bu kısımda, lineer regresyondaki sabit terimin bazı özellikleri, yani, \mathbf{j} vektörünün rolü üzerinde duracağız bkz. Isotalo, Puntanen & Styan (2006). Bu nedenle, çok basit bir

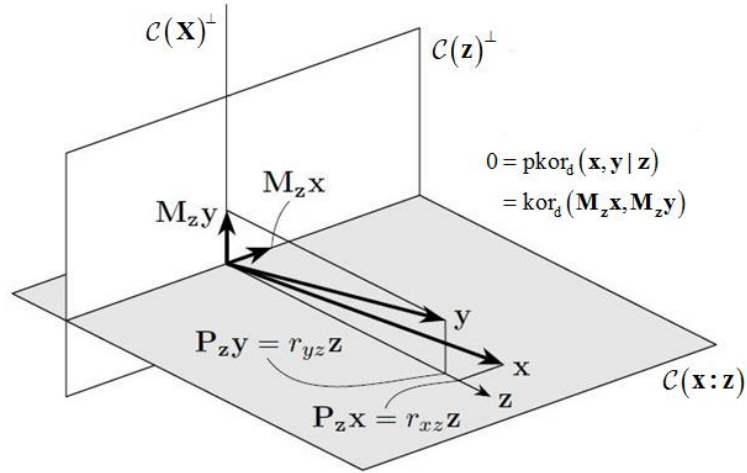
$$\mathcal{M}_0 = \{\mathbf{y}, \mathbf{j}\beta_0, \sigma^2 \mathbf{I}\} \quad (4.23)$$

lineer modelini göz önüne alınız. Bu durumda açık olarak,

- $\mathbf{P}_{\mathbf{j}} = \text{OLSE}(\mathbf{j}\beta_0) = \mathbf{j}(\mathbf{j}'\mathbf{j})^{-1} \mathbf{j}'\mathbf{y} = \mathbf{y}\mathbf{j}$ uygun değişkenlerin vektörüdür. Burada OLSE; alışılmış en küçük kareler tahminini gösterir.
- $\mathbf{C}\mathbf{y} = \mathbf{y} - \mathbf{P}_{\mathbf{j}}\mathbf{y} = \tilde{\mathbf{y}}$ hataların (tahmin edilen hataların) vektörüdür.
- \mathcal{M}_0 altında hata kareler toplamı



Şekil 4.3. x ve y arasında kosinüs yüksek fakat korelasyon yok
 x ve y vektörleri çok yakın.



Şekil 4.4. $r_{y:z}$ kısmi korelasyonu 0 a eşit

Eğer y ; x ve z ile açıklırsa, bu takdirde x in regresyon katsayısının 0 olduğunu kabul ediniz.

$$\text{HKT}(\mathcal{M}_0) = \text{HKT}_0 = \mathbf{y}'(\mathbf{I} - \mathbf{P}_j)\mathbf{y} = \mathbf{y}'\mathbf{C}\mathbf{y} = \sum_{i=1}^n (y_i - \bar{y})^2 = \text{GKT} \quad (4.24)$$

dır. Burada olan şey, birlerin sütun vektörünün, yani, \mathbf{j} nin \mathbf{y} den etkisini gideririz ve kalan şey hata (tahmin edilen hata) dır, ki bu durumda o; merkezileştirilmiş \mathbf{y} , yani, $\tilde{\mathbf{y}}$ dır. Şimdi

$$\mathcal{M}_{x,y} = \{\mathbf{y}, \mathbf{x}\beta, \sigma^2 \mathbf{I}\} \quad (4.25)$$

modelini göz önüne alınız. Bu durumda,

- $\hat{\mathbf{y}} = \text{OLSE}(\mathbf{x}\beta)$; uygun değerlerin vektörüdür:

$$\hat{\mathbf{y}} = \mathbf{P}_x \mathbf{y} = \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1} \mathbf{x}'\mathbf{y} = \hat{\beta}\mathbf{x} \quad (4.26)$$

dir. Burada, $\hat{\beta} = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}}$ dir.

- $(\mathbf{I} - \mathbf{P}_x)\mathbf{y}$ hataların (tahmin edilen hataların) bir vektörüdür.
- $\mathcal{M}_{x,y}$ altında hata kareler toplamı

$$\text{HKT}(\mathcal{M}_{x,y}) = \mathbf{y}'(\mathbf{I} - \mathbf{P}_x)\mathbf{y} = \|(\mathbf{I} - \mathbf{P}_x)\mathbf{y}\|^2 \quad (4.27)$$

dir. Bu modelin “mükemmelliği (iyiliği)” için çok doğal bir ölçü

$$R_{xy}^2 = \frac{\|\mathbf{P}_x \mathbf{y}\|^2}{\|\mathbf{y}\|^2} = \frac{\|\hat{\mathbf{y}}\|^2}{\|\mathbf{y}\|^2} = \frac{\mathbf{y}'\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1} \mathbf{x}'\mathbf{y}}{\mathbf{y}'\mathbf{y}} = \cos^2(\mathbf{x}, \mathbf{y}) \quad (4.28)$$

olmalıdır. Yukarıdaki iyilik ölçüsünün

$$R_{xy}^2 = 1 - \sin^2(\mathbf{x}, \mathbf{y}) = 1 - \frac{\|(\mathbf{I} - \mathbf{P}_x)\mathbf{y}\|^2}{\|\mathbf{y}\|^2} = 1 - \frac{\text{HKT}(\mathcal{M}_{x,y})}{\|\mathbf{y}\|^2} \quad (4.29)$$

olarak ifade edilebildiğine de dikkat ediniz. Burada tanımlanan R_{xy}^2 niceliğine (sayısına) $\mathcal{M}_{x,y}$ altında determinasyon (belirleme) katsayısı denir. Esas özellik, $\mathcal{M}_{x,y}$ modelinde \mathbf{j} vektörünün olmadığına ve bu nedenle, R_{xy}^2 nin $\mathcal{M}_{12} = \{\mathbf{y}, \beta_0 \mathbf{j} + \beta_1 \mathbf{x}, \sigma^2 \mathbf{I}\}$ modelinde belirleme katsayısı olan r_{xy}^2 ye eşit olması gerekmediğine dikkat etmektir.

Şimdi eğer, “ \mathbf{y} ve \mathbf{x} den \mathbf{j} nin etkisini giderirsek” (tabiri caizse), yani, onları merkezileştirirsek, bu takdirde şerk koyulan kovaryans matrisini bile bile terk etmek üzere,

$$\mathcal{M}_{x,y,1} = \{\tilde{\mathbf{y}}, \tilde{\mathbf{x}}\beta, \#\} \quad (4.30)$$

modelini göz önüne alabiliriz. $\mathcal{M}_{x,y-1}$ modelinin, $\mathcal{M}_{x,y}$ modelini \mathbf{C} matrisi ile önden çarpmak suretiyle elde edilebildiğine dikkat ediniz. Şimdi, $\mathcal{M}_{x,y-1}$ modelinin iyiliği için doğal bir ölçü,

$$R_{xy-1}^2 = \frac{\|\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{y}}\|^2}{\|\tilde{\mathbf{y}}\|^2} = \frac{\tilde{\mathbf{y}}'\tilde{\mathbf{x}}(\tilde{\mathbf{x}}'\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}'\tilde{\mathbf{y}}}{\tilde{\mathbf{y}}'\tilde{\mathbf{y}}} = \cos^2(\tilde{\mathbf{x}},\tilde{\mathbf{y}}) \quad (4.31)$$

olacaktır. $R_{xy-1}^2 = \text{kor}_d^2(\mathbf{x},\mathbf{y}) = r_{xy}^2$ olduğu ve bu nedenle r_{xy}^2 nin; \mathbf{j} nin etkisi hem \mathbf{y} ve hem de \mathbf{x} den giderildikten sonra, \mathbf{y} ; \mathbf{x} üzerinde regresyonlandığında, iyiliğin bir ölçüsü olarak yorumlanabildiğini gösterdiğimiz açıktır. Sonuçta yukarıda, lineer regresyonda herkesçe bilinen bazı kavramlara ulaştık. Düşüncelerimiz için sadece doğal ve sezgisel yolları kullandık. Bir sonra duruma daha biçimsel bakalım.

Aşağıdaki dört modeli göz önüne alınız:

$$\mathcal{M}_{12} = \{\mathbf{y}, \beta_0\mathbf{j} + \beta_1\mathbf{x}, \sigma^2\mathbf{I}\}, \quad \mathcal{M}_{12,1} = \{\mathbf{C}\mathbf{y}, \beta_1\mathbf{C}\mathbf{x}, \sigma^2\mathbf{C}\}, \quad (4.32a)$$

$$\mathcal{M}_C = \{\mathbf{y}, \beta_0\mathbf{j} + \beta_1\mathbf{C}\mathbf{x}, \sigma^2\mathbf{I}\}, \quad \mathcal{M}_r = \{\mathbf{y}, \beta_1\mathbf{C}\mathbf{x}, \sigma^2\mathbf{I}\} \quad (4.32b)$$

\mathcal{M}_{12} modeli tam modeldir ve diğer tüm modeller onun çeşitli versiyonlarıdır. $\mathcal{M}_{12,1}$ in yukarıda göz önüne alınan $\mathcal{M}_{x,y-1}$ e karşılık geldiği açıktır. Bu versiyonların tümünde sabit terime bağlı bir şey yaptık:birşeyi merkezileştirdik. Yukarıdaki modeller uygulamada (ve istatistik derslerinde regresyonu öğretmede) sık sık görülür. (4.32) deki modellere bir göz atarak, bir yardımcı teorem olarak ifade edebileceğimiz ilginç bir özelliği hemen görürüz.

Yardımcı Teorem 4.2: (4.32) de tanımlanan modelleri göz önüne alınız ve $\mathbf{X} = (\mathbf{j} : \mathbf{x})$ tam sütun ranka sahip olsun. Bu takdirde $\hat{\beta}_1$ her modelde aynıdır, yani,

$$\hat{\beta}_1(\mathcal{M}_{12}) = \hat{\beta}_1(\mathcal{M}_{12,1}) = \hat{\beta}_1(\mathcal{M}_r) = \hat{\beta}_1(\mathcal{M}_C) \quad (4.33)$$

dir. Bundan başka, (\mathcal{M}_{12}) , $(\mathcal{M}_{12,1})$ ve (\mathcal{M}_C) modelleri altında hatalar (tahmin edilen hatalar) aynıdır. (Styan ve ark., 2006).

İspat. (4.33) deki ilk iki eşitlik açıktır. (\mathcal{M}_C) modeli (\mathcal{M}_{12}) nin yeniden parametrelendirilmiştir. Bu;

$$\mathbf{A} = \begin{pmatrix} 1 & -\bar{x} \\ 0 & 1 \end{pmatrix} \quad (4.34)$$

olmak üzere, $(\mathbf{j} : \mathbf{C}\mathbf{x}) = (\mathbf{j} : \mathbf{x})\mathbf{A}$ dan görülür. $\hat{\beta}_0(\mathcal{M}_C) = \bar{y}$ ve $\hat{\beta}_1(\mathcal{M}_C) = \hat{\beta}_1(\mathcal{M}_{12})$ olduğunu doğrulamak kolaydır. (4.18) e göre,

$$\mathcal{C}(\mathbf{j} : \mathbf{C}\mathbf{x}) = \mathcal{C}(\mathbf{j} : \mathbf{x}), \quad \mathbf{P}_{(\mathbf{j}:\mathbf{C}\mathbf{x})} = \mathbf{P}_{(\mathbf{j}:\mathbf{x})} \quad (4.35)$$

yazarız ve bu nedenle (\mathcal{M}_C) ve (\mathcal{M}_{12}) altında hata tahminleri vektörleri aynıdır, yani,

$$\text{hata}(\mathcal{M}_C) = \text{hata}(\mathcal{M}_{12}) = \mathbf{y} - \mathbf{H}\mathbf{y} = \mathbf{M}\mathbf{y} \quad (4.36)$$

dir. Burada, $\mathbf{H} = \mathbf{P}_{(\mathbf{d},\mathbf{x})}$ ve $\mathbf{M} = \mathbf{I} - \mathbf{H}$ dir.

$(\mathcal{M}_{12,1})$ altında hata tahminleri vektörü

$$\begin{aligned} \text{hata}(\mathcal{M}_{12,1}) &= \mathbf{C}\mathbf{y} - \mathbf{P}_{\mathbf{C}\mathbf{x}}\mathbf{C}\mathbf{y} \\ &= \mathbf{y} - \left[\mathbf{P}_{\mathbf{j}}\mathbf{y} + \mathbf{C}\mathbf{x}(\mathbf{x}'\mathbf{C}\mathbf{x})^{-1}\mathbf{x}'\mathbf{C}\mathbf{y} \right] \\ &= \mathbf{y} - \left[\bar{y}\mathbf{j} + (\mathbf{x} - \bar{x}\mathbf{j})\hat{\beta}_1 \right] \\ &= \mathbf{y} - \left[(\bar{y} - \bar{x}\hat{\beta}_1)\mathbf{j} + \hat{\beta}_1\mathbf{x} \right] \\ &= \mathbf{y} - (\hat{\beta}_0\mathbf{j} + \hat{\beta}_1\mathbf{x}) \\ &= \mathbf{y} - (\mathbf{j} : \mathbf{x}) \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \\ &= \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \text{hata}(\mathcal{M}_{12}) \end{aligned} \quad (4.37)$$

olur. Burada

$$\hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1 \quad \text{ve} \quad \hat{\beta}_1 = \mathbf{x}'\mathbf{C}\mathbf{y}/\mathbf{x}'\mathbf{C}\mathbf{x} \quad (4.38)$$

dir. \mathbf{H} ortogonal izdüşürücünün bir ayrışımını kullanarak, (4.37) için değişik bir ispat aşağıdaki gibi verilebilir:

$(\mathcal{M}_{12,1})$ indirgenmiş modeli altında OLSE lerin hatası (tahmin edilen hataları) (\mathcal{M}_{12}) tam modeli altında OLSE lerin hatasına eşit olduğunu, $\text{hata}(\mathcal{M}_{12,1}) = \text{hata}(\mathcal{M}_{12})$, yani, $\text{hata}(\mathbf{M}_1\mathbf{y}; \mathbf{M}_1\mathbf{x}_2) = \text{hata}(\mathbf{y}; \mathbf{x})$ olduğunu görmek

ilginçtir. Model matrisi sadece $(\mathbf{j}; \mathbf{x})$ olduğunda, bu; yukarıdaki yardımcı teoremden ifade edildi. Genel durumda, $(\mathcal{M}_{12,1})$ altında hata vektörü

$$\begin{aligned}\mathbf{H} &= \mathbf{P}_{(\mathbf{x}_1; \mathbf{x}_2)} = \mathbf{P}_{\mathbf{x}_1} + \mathbf{P}_{\mathbf{M}_1; \mathbf{x}_2} \\ \mathbf{H}\mathbf{y} &= \mathbf{x}_1\hat{\beta}_1 + \mathbf{x}_2\hat{\beta}_2 = \mathbf{P}_{\mathbf{x}_1}\mathbf{y} + \mathbf{P}_{\mathbf{M}_1; \mathbf{x}_2}\mathbf{y}\end{aligned}$$

$$\text{hata}(\mathcal{M}_{12,1}) = \mathbf{M}_1\mathbf{y} - \mathbf{P}_{\mathbf{M}_1; \mathbf{x}_2}\mathbf{M}_1\mathbf{y} = \mathbf{y} - (\mathbf{P}_{\mathbf{x}_1} + \mathbf{P}_{\mathbf{M}_1; \mathbf{x}_2})\mathbf{y} = \mathbf{y} - \mathbf{P}_{(\mathbf{x}_1; \mathbf{x}_2)}\mathbf{y} = \mathbf{M}\mathbf{y} = \text{hata}(\mathcal{M}_{12})$$

olur. $\mathbf{P}_{\mathbf{M}_1; \mathbf{x}_2}\mathbf{M}_1\mathbf{y} = \mathbf{P}_{\mathbf{M}_1; \mathbf{x}_2}\mathbf{y}$ olduğuna dikkat ediniz. bkz. “matrix tricks for linear statistical models” (8.72, s. 164). Böylece ispat tamamlanır.

(4.33) deki birinci eşitlik Frisch-Wang-Lovell Teoremi (bu teoremin özel bir durumu) olarak bilinir.

Farklı yollarla R^2 için sonuca ulaşabiliriz. \mathbf{j} nin önemli bir role sahip olduğu çok doğal bir yaklaşım yegane açıklayıcı değişkenin bir sabit olduğu basit temel modeli, yani, $\mathcal{M}_0 = \{\mathbf{y}, \mathbf{j}\beta_0, \sigma^2\mathbf{I}\}$ modelini göz önüne almaktır. Tahmin edilen hata vektörü merkezleştirilmiş \mathbf{y} , yani, $\mathbf{C}\mathbf{y}$ iken \mathcal{M}_0 altında tahmin edilmiş hataların kareleri toplamı $\text{OLSE}(\mathbf{j}\beta_0) = \mathbf{P}_j\mathbf{y} = \bar{y}\mathbf{j}$ elde ederiz ve bu nedenle

$$\text{HKT}_0 = \mathbf{y}'(\mathbf{I} - \mathbf{P}_j)\mathbf{y} = \mathbf{y}'\mathbf{C}\mathbf{y} = \|\mathbf{C}\mathbf{y}\|^2 = t_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = \text{GKT} \quad (4.39)$$

dir. Hata kareler toplamı vasıtasıyla $\mathcal{M}_{12} = \{\mathbf{y}, \mathbf{j}\beta_0 + \mathbf{x}\beta_1, \sigma^2\mathbf{I}\}$ tam modeli \mathcal{M}_0 basit temel modeliyle karşılaştırmak isteyebiliriz, aynı zamanda x -değişkeni bir açıklayıcı değişken olarak kullanıldığında hata kareler toplamında elde edilen kazanç ne kadardır. \mathcal{M}_{12} altında hata kareler toplamı

$$\begin{aligned}\text{HKT} &= \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} = \|(\mathbf{I} - \mathbf{H})\mathbf{y}\|^2 \\ &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \left[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right]^2\end{aligned} \quad (4.40)$$

dir. \mathcal{M}_0 dan \mathcal{M}_{12} ye gidildiğinde HKT deki değişim

$$\text{HKT}_0 - \text{HKT} = \mathbf{y}'(\mathbf{I} - \mathbf{P}_j)\mathbf{y} - \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{y}'(\mathbf{H} - \mathbf{P}_j)\mathbf{y} = \text{RKT} \quad (4.41)$$

dir. RKT ye “regresyon kareler toplamı” denir. Bu durumda $(\mathbf{H} - \mathbf{P}_j)$ matrisi simetrik ve idempotenttir. bkz. “matrix tricks for linear statistical models yardımcı teorem 7.1” (s.152) ve (8.119) (s.171).

$$\mathbf{H} - \mathbf{P}_j = \mathbf{P}_{\mathcal{C}_X} = \mathbf{P}_{\bar{\mathbf{x}}} = \mathbf{P}_{\mathcal{C}(\mathbf{X}) \cap \mathcal{C}(\mathbf{j})^\perp} \quad (4.42)$$

dir. Bu nedenle,

$$\text{RKT} = \|(\mathbf{H} - \mathbf{P}_j)\mathbf{y}\|^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2 \quad (4.43)$$

dir. \mathcal{M}_0 in yerine \mathcal{M}_{12} kullanıldığında RKT nin değeri HKT deki azalmayı bildirir, fakat o; HKT deki bağıl (göreceli) azalmayı incelemek için, kesinlikle daha aydınlatıcıdır, yani,

$$\frac{\text{HKT}_0 - \text{HKT}}{\text{HKT}_0} = \frac{\text{GKT} - \text{HKT}}{\text{GKT}} = \frac{\text{RKT}}{\text{GKT}} = 1 - \frac{\text{HKT}}{\text{GKT}} = R^2 \quad (4.44)$$

oranını hesaplamak için nedenlere sahibiz. Yukarıda tanımlanan R^2 nin temel bir özelliği onun \mathbf{y} ve $\mathbf{H}\mathbf{y}$ arasındaki korelasyon katsayısının karesine eşit olmasıdır. Bu sonuca daha sonra döneceğiz, ancak o; \mathbf{j} nin önemli rolünü hoş bir şekilde açıkladığından burada onu ifade ettik.

Yardımcı Teorem 4.3: $\mathbf{j} \in \mathcal{C}(\mathbf{X})$ olmak üzere $\mathcal{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\}$ modelini göz önüne alınız ve R^2 ; (4.44) te tanımlandığı gibi olsun. Bu takdirde

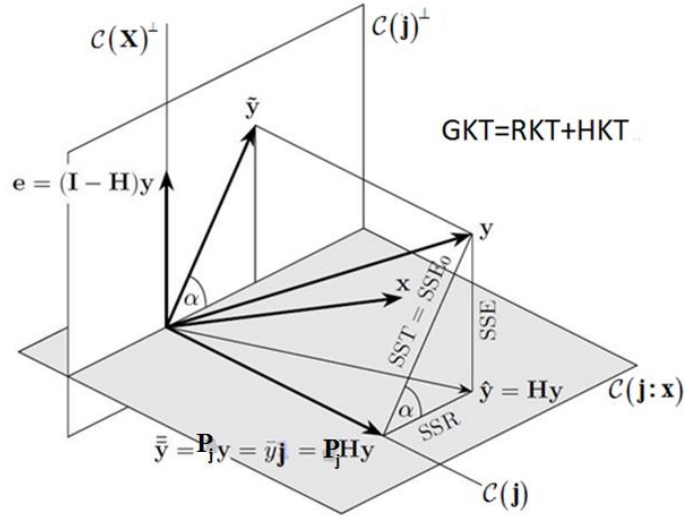
$$R = \text{kor}_d(\mathbf{y}, \mathbf{H}\mathbf{y}) = \cos(\mathbf{C}\mathbf{y}, \mathbf{C}\mathbf{H}\mathbf{y}) \quad (4.45)$$

dir. (Styan ve ark., 2006).

İspat. İspat için, bkz. “matrix tricks for linear statistical models” Yardımcı teorem 8.6 (s.172).

$\mathbf{X} = (\mathbf{j} : \mathbf{x}_1 : \dots : \mathbf{x}_k)$ $n \times (k+1)$ model matrisine göz atıldığında ilk \mathbf{j} sütunu diğer herhangi bir sütun gibi görünür fakat büyük bir fark vardır: \mathbf{X} model matrisinde gösterilen diğer tüm değişkenler sıfırdan farklı bir örnek varyansına sahiptir.

Şüphesiz bu, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ arasında \mathbf{j} nin katlarının olmamasını isteyebilmemizdir. $GKT = RKT + HKT$ ayrışımının arkasındaki geometri Şekil 5 te resimlendirilir.



Şekil 4.5. $GKT = RKT + HKT$ nin açıklaması

4.3 Geometrik Olarak Kısmi Korelasyon

Burada örnek kısmi korelasyonları çok kısaca ele alacağız ve okuyucuya onların sadece özel hata tahminleri arasındaki korelasyonlar olduğunu hatırlatacağız. Aşağıdaki iki lineer modeli göz önüne alınız:

$$\mathcal{M}_x = \{\mathbf{x}, (\mathbf{j}; \mathbf{z})\boldsymbol{\beta}, \sigma^2 \mathbf{I}\}, \quad \mathcal{M}_y = \{\mathbf{y}, (\mathbf{j}; \mathbf{z})\boldsymbol{\beta}, \sigma^2 \mathbf{I}\} \quad (4.46)$$

Bu modellerdeki hatalar (tahmin edilmiş hatalar)

$$\mathbf{u} = \mathbf{e}_x \mathbf{z} = \text{hata}(\mathbf{x}; \mathbf{Z}) = (\mathbf{I} - \mathbf{P}_Z) \mathbf{x} \quad (4.47a)$$

$$\mathbf{v} = \mathbf{e}_y \mathbf{z} = \text{hata}(\mathbf{y}; \mathbf{Z}) = (\mathbf{I} - \mathbf{P}_Z) \mathbf{y} \quad (4.47b)$$

dir. Burada $\mathbf{Z} = (\mathbf{j}; \mathbf{z})$ dir. (4.46) daki modelleri \mathbf{C} merkezileştirme matrisiyle önden çarparak, (4.47) dekiyle aynı hatalara sahip olan modelleri elde ederiz. Notasyonu basitleştirmek için, \mathbf{x} , \mathbf{y} ve \mathbf{z} nin hepsini merkezileştirdiğimizi ve karşılık gelen modelleri

$$\mathcal{M}_{cx} = \{\mathbf{x}, \mathbf{z}\boldsymbol{\beta}, \#\}, \quad \mathcal{M}_{cy} = \{\mathbf{y}, \mathbf{z}\boldsymbol{\beta}, \#\} \quad (4.48)$$

olarak gösterdiğimizi kabul edelim. Ayrıca, şimdi bu iki hata arasındaki korelasyon (ilişki) ile ilgilendiğimizden, (4.48) deki \mathbf{x} , \mathbf{y} ve \mathbf{z} vektörlerinin sadece merkezileştirilmediklerini, aynı zamanda birim uzunluğa sahip olduklarını da kabul edebiliriz; yani,

$$\mathbf{j}'_n(\mathbf{x} : \mathbf{y} : \mathbf{z}) = \mathbf{0}, \quad \mathbf{x}'\mathbf{x} = \mathbf{y}'\mathbf{y} = \mathbf{z}'\mathbf{z} = 1 \quad (4.49)$$

olduğunu da kabul edebiliriz. Şimdi

$$\mathbf{x}'\mathbf{y} = r_{xy}, \quad \mathbf{x}'\mathbf{z} = r_{xz}, \quad \mathbf{y}'\mathbf{z} = r_{yz} \quad (4.50a)$$

$$\mathbf{u} = \mathbf{e}_{\mathbf{x}:\mathbf{z}} = \text{hata}(\mathbf{x}; \mathbf{z}) = (\mathbf{I} - \mathbf{P}_{\mathbf{z}})\mathbf{x} = (\mathbf{I} - \mathbf{z}\mathbf{z}')\mathbf{x} = \mathbf{x} - r_{xz}\mathbf{z} \quad (4.50b)$$

$$(4.50c)$$

$$\mathbf{v} = \mathbf{e}_{\mathbf{y}:\mathbf{z}} = \text{hata}(\mathbf{y}; \mathbf{z}) = (\mathbf{I} - \mathbf{P}_{\mathbf{z}})\mathbf{y} = (\mathbf{I} - \mathbf{z}\mathbf{z}')\mathbf{y} = \mathbf{y} - r_{yz}\mathbf{z}$$

ve

$$\mathbf{u}'\mathbf{v} = \mathbf{x}'(\mathbf{I} - \mathbf{z}\mathbf{z}')\mathbf{y} = r_{xy} - r_{xz}r_{yz} \quad (4.51a)$$

$$\mathbf{u}'\mathbf{u} = \mathbf{x}'(\mathbf{I} - \mathbf{z}\mathbf{z}')\mathbf{x} = 1 - r_{xz}^2 \quad (4.51b)$$

$$\mathbf{v}'\mathbf{v} = \mathbf{y}'(\mathbf{I} - \mathbf{z}\mathbf{z}')\mathbf{y} = 1 - r_{yz}^2 \quad (4.51c)$$

dir ve bu nedenle, kısmi korelasyon katsayısı için meşhur

$$r_{xy \cdot z} = \text{pkor}_d(\mathbf{x}, \mathbf{y} | \mathbf{z}) = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} \quad (4.52)$$

formülünü elde ederiz. Eğer $r_{xz}^2 \neq 1$ ve $r_{yz}^2 \neq 1$ ise, açık olarak, $r_{xy \cdot z}$ iyi tanımlıdır.

Matris terimleri cinsinden

$$\begin{aligned} (\mathbf{u} : \mathbf{v})'(\mathbf{u} : \mathbf{v}) &= (\mathbf{x} : \mathbf{y})'(\mathbf{I} - \mathbf{P}_{\mathbf{z}})(\mathbf{x} : \mathbf{y}) \\ &= (\mathbf{x} : \mathbf{y})'(\mathbf{x} : \mathbf{y}) - (\mathbf{x} : \mathbf{y})'\mathbf{z}\mathbf{z}'(\mathbf{x} : \mathbf{y}) \\ &= \mathbf{R}_{11} - \mathbf{r}_{12}\mathbf{r}'_{12} : \mathbf{R}_{11 \cdot 2} \end{aligned} \quad (4.53)$$

elde ederiz. Burada,

$$\text{kor}_d(\mathbf{x} : \mathbf{y} : \mathbf{z}) = \mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{r}_{12} \\ \mathbf{r}'_{12} & 1 \end{pmatrix} \quad (4.54)$$

dür. Buradan, kısmi korelasyonların matrisi

$$\text{pkor}_d[(\mathbf{x}:\mathbf{y})|\mathbf{z}] = [\text{köşeg}(\mathbf{R}_{11.2})]^{-1/2} \mathbf{R}_{11.2} [\text{köşeg}(\mathbf{R}_{11.2})]^{-1/2} \quad (4.55)$$

olarak ifade edilebilir. $\mathbf{Z} = (\mathbf{j}:\mathbf{z})$ olmak üzere, orijinal değişkenler cinsinden,

$$r_{xy:z} = \cos[\text{hata}(\mathbf{x};\mathbf{Z}), \text{hata}(\mathbf{y};\mathbf{Z})] = \cos[(\mathbf{I} - \mathbf{P}_z)\mathbf{x}, (\mathbf{I} - \mathbf{P}_z)\mathbf{y}] \quad (4.56)$$

elde ederiz.

$$\mathbf{U} = (\mathbf{x}:\mathbf{y}:\mathbf{z}) = (\mathbf{U}_1:\mathbf{z}) \quad (4.57)$$

göstererek,

$$\mathbf{E}_{\mathbf{U}_1:\mathbf{z}} = (\mathbf{I} - \mathbf{P}_{(\mathbf{j}:\mathbf{z})})\mathbf{U}_1 = (\mathbf{e}_{x:z}:\mathbf{e}_{y:z}) \quad (4.58)$$

hata tahmini vektörlerinin matrisini elde ederiz. \mathbf{C} merkezileştirme matrisi olmak üzere, $\mathbf{I} - \mathbf{P}_{(\mathbf{j}:\mathbf{z})} = \mathbf{C} - \mathbf{P}_{\mathbf{Cz}}$ olduğundan, bkz (4.22), $\mathbf{E}_{\mathbf{U}_1:\mathbf{z}} = (\mathbf{C} - \mathbf{P}_{\mathbf{Cz}})\mathbf{U}_1$ yazabiliriz.

Ayrıca,

$$\mathbf{T}_{11.2} := \mathbf{E}'_{\mathbf{U}_1:\mathbf{z}} \mathbf{E}_{\mathbf{U}_1:\mathbf{z}} = \mathbf{U}'_1 (\mathbf{C} - \mathbf{P}_{\mathbf{Cz}}) \mathbf{U}_1 = \mathbf{U}'_1 \mathbf{C} \mathbf{U}_1 - \mathbf{U}'_1 \mathbf{P}_{\mathbf{Cz}} \mathbf{U}_1 \quad (4.59)$$

göstererek, kısmi korelasyonların matrisini

$$\text{pkor}_d[\mathbf{U}_1|\mathbf{z}] = [\text{köşeg}(\mathbf{T}_{11.2})]^{-1/2} \mathbf{T}_{11.2} [\text{köşeg}(\mathbf{T}_{11.2})]^{-1/2} \quad (4.60)$$

olarak ifade edebiliriz. Bir parçalanmış matrisin tersi için aşağıdaki yardımcı teoremi yazarız:

Yardımcı Teorem 4.4: \mathbf{A} negatif tanımlı olmayan bir simetrik matris olsun. Bu takdirde, $\mathbf{L} = (\mathbf{L}_1:\mathbf{L}_2)$ olmak üzere, $\mathbf{A} = \mathbf{L}'\mathbf{L}$ ve bu nedenle

$$\mathbf{A} = \mathbf{L}'\mathbf{L} = \begin{pmatrix} \mathbf{L}'_1\mathbf{L}_1 & \mathbf{L}'_1\mathbf{L}_2 \\ \mathbf{L}'_2\mathbf{L}_1 & \mathbf{L}'_2\mathbf{L}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \quad (4.61)$$

olacak şekilde bir \mathbf{L} matrisi mevcuttur. Eğer aşağıdaki ifadelerden herhangi biri gerçekleşirse, bu takdirde her üçü de gerçekleşir.

$$(a) \text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}_{11}) + \text{rank}(\mathbf{A}_{22}), \text{ yani, } \mathcal{C}(\mathbf{L}_1) \cap \mathcal{C}(\mathbf{L}_2) = \{\mathbf{0}\},$$

$$\mathbf{A}_{22.1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} = \mathbf{L}'_2(\mathbf{I} - \mathbf{P}_{L_1})\mathbf{L}_2 \quad (4.62)$$

$$\mathbf{A}_{11.2} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} = \mathbf{L}'_1(\mathbf{I} - \mathbf{P}_{L_2})\mathbf{L}_1$$

olmak üzere;

(b)

$$\mathbf{A}^+ = \begin{pmatrix} \mathbf{A}_{11.2}^+ & -\mathbf{A}_{11.2}^+\mathbf{A}_{12}\mathbf{A}_{22}^+ \\ -\mathbf{A}_{22}^+\mathbf{A}_{21}\mathbf{A}_{11.2}^+ & \mathbf{A}_{22}^+ + \mathbf{A}_{22}^+\mathbf{A}_{21}\mathbf{A}_{11.2}^+\mathbf{A}_{12}\mathbf{A}_{22}^+ \end{pmatrix} \quad (4.63)$$

(c)

$$\mathbf{A}^+ = \begin{pmatrix} \mathbf{A}_{11.2}^+ & -\mathbf{A}_{11.2}^+\mathbf{A}_{12}\mathbf{A}_{22}^+ \\ -\mathbf{A}_{22}^+\mathbf{A}_{21}\mathbf{A}_{11.2}^+ & \mathbf{A}_{22}^+ + \mathbf{A}_{22}^+\mathbf{A}_{21}\mathbf{A}_{11.2}^+\mathbf{A}_{12}\mathbf{A}_{22}^+ \end{pmatrix} \quad (4.64)$$

Burada “ \mathbf{A}^+ ”; \mathbf{A} matrisinin Moure-Penrose genelleştirilmiş tersini gösterir. Ayrıca bir parçalanmış matrisin (normal) tersi için aşağıda verilen formül geçerlidir.

$\mathbf{E} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ ve $\mathbf{F} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ olmak üzere,

$$\begin{aligned} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} &= \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{E}^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{E}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{F}^{-1} & -\mathbf{F}^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{F}^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\mathbf{F}^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix} \end{aligned} \quad (4.65)$$

elde edilir.

Bu yardımcı teoreme göre $\mathbf{T}_{11.2}^{-1}$ in,

$$\mathbf{T} = \mathbf{U}'\mathbf{C}\mathbf{U} = \begin{pmatrix} \mathbf{U}'_1\mathbf{C}\mathbf{U}_1 & \mathbf{U}'_1\mathbf{C}\mathbf{z} \\ \mathbf{z}'\mathbf{C}\mathbf{U}_1 & \mathbf{z}'\mathbf{C}\mathbf{z} \end{pmatrix} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{t}_{12} \\ \mathbf{t}'_{12} & t_{zz} \end{pmatrix} \quad (4.66)$$

matrisinin tersinin sol üst bloğu olduğuna dikkat ediniz.

5. SONUÇ VE ÖNERİLER

Dört bölümden oluşan bu tezde lineer modeller genel olarak tanıtıldı. Konunun anlaşılmasındaki bazı güçlükler, geometrik yorum ve gösterimlerle ortadan kaldırılmaya çalışıldı. Özellikle lineer modellerin geometrik yorumuna yeterince yer verildi. Bunun dışında regresyon analizinde kullanılan singüler değer ayrışımı yine geometrik yorumlar ve şekillerle açıklanmaya çalışıldı. Bu tezde açıklanmaya çalışılan konular ve sunulan geometrik yorumlar istatistikte birçok alanda karşılaşılan zorlukları yeneceği kanaatindeyim.



KAYNAKLAR

- Belsley, D. A., Kuh, E., Welsch, R. E. 1980. Identifying Influential Data and Sources of Collinearity. John Wiley, New Jersey.
- Belsley, D. A. 1991. Conditioning Diagnostics: Collinearity and Weak Data in Regression. Wiley, New York, 20, 93, 100, 162, 178, 412.
- Box, G. E. P., Draper, N. R. 1987. Empirical Model-Building and Response Surfaces. John Wiley & Sons, New York.
- Bradu, D., Gabriel, K. R. 1978. The Biplot as a Diagnostic Tool for Models of Two-Way Tables. *Technometrics*, 20: 47-68.
- Chambers, J. M. 1977. Computational Methods for Data Analysis. John Wiley, New York.
- Chatterjee, S., Price, B. 1977. Regression Analysis by Example. John Wiley, New York.
- Curtis, M. L. 1979. Matrix Groups, Springer-Verlag, New York.
- Dawid, A. P. 1977. Spherical Matrix Distributions and a Multivariate Model. *Journal of Royal Statistical Society Series B*, 39: 254-261.
- Dempster, A. P. 1969. Elements of Continuous Multivariate Analysis. Addison-Wesley, Reading Mass, 388.
- Draper, N. R., Smith, H. 1981. Applied Regression Analysis (2nd ed.). John Wiley, New York.
- Eaton, M. L. 1983. Multivariate Statistics: A Vector Space Approach. John Wiley & Sons, New York.
- Fang, K. T., Anderson T. W. 1990. Statistical Inference in Elliptically Contoured and Related Distributions. Allerton Press, New York.
- Fang, K. T., Zhang Y. T. 1990. Generalized Multivariate Analysis. Springer-Verlag, New York.
- Fisher, R. A. 1939. The Sampling Distribution of some Statistics Obtained from Non-Linear Equations. *Annals of Eugenics*, 9: 238-249.
- Frisch, R., Waugh, F. V. 1933. Partial time regressions as compared with individual trends. *Econometrica*, 1: (163); 331.
- Gittins, R. 1985. Cononical Analysis: A Review with Applications in Ecology. Springer-Verlag, New York.
- Hahn, G. J., Meeker, W. Q. Jr., Feder, P. I. 1976. The Evaluation and Comparison of Experimental Design for Fitting Regression Relationships. *Journal of Quality Technology*, 8: 140-157.

- Hocking, R. R., Speed, F. M., Lynn, M. J. 1976. A Class of Biased Estimators in Linear Regression. *Technometrics*, 18: 425-437.
- Hoerl, A. E., Kennard, R. W. 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12: 55-67.
- Isotalo, J., Puntanen, S., Styan, G. P. H. 2006. On the role of the constant term in linear regression. In *Festschrift for Tarmo Pukkila on his 60th Birthday* (E. P. Liski, J. Isotalo, J. Niemelä, S. Puntanen & G. P. H. Styan, eds.), Dept. of Mathematics, Statistics and Philosophy, University of Tampere, 243–259.
- James, A. T. 1954. Normal Multivariate Analysis and The Orthogonal Group. *The Annals of Mathematical Statistics*, 25: 40-75.
- John, J. A., Williams, E. R. 1982. Conjectures for Optimal Block Designs. *Journal of Royal Statistical Society, Series B*(44): 221-225.
- John, J. A., Williams, E. R. 1995. *Cyclic and Computer Generated Designs*, Sec, Ed. Chapman and Hall, New York.
- Johnson, R. A., Wichern, D. W. 1992. *Applied Multivariate Statistical Analysis* 3rd ed. Prentice-Hall, London.
- Kshirsagar, A. M. 1972. *Multivariate Analysis*. Marcel Dekker, New York.
- Langsrud Ø. 2004. The geometrical interpretation of statistical tests in multivariate linear regression. *Springer-Verlag*, 45:111-122.
- Mandel, J. 1971. A New Analysis of Variance Models for Non-Additive Data. *Technometrics*, 13: 1-18.
- Mandel, J. 1981. Fitting Curves and Surfaces with Monotonic and Non-Monotonic Four Parameter Equations. *Journal of Research of the National Bureau of Standards*, 86(1): 1-25.
- Mandel, J. 1982. Use of the Singular Value Decomposition in Regression Analysis. *The American Statistician*, 36(1):15-24.
- Mardiva, K. V., Kent, J. T, Bibby, J. M. 1979. *Multivariate Analysis*. Academic Press Limited, London.
- Marquardt, D. W. 1970. Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. *Technometrics*, 12: 591-612.
- Marquardt, D. W., Snee, R. D. 1975. Ridge Regression in Practise. *The American Statistician*, 29: 3-20.
- Miller, A., 2015. *The Geometry of Linear Models Lecture Notes 1*. Auckland, <https://www.stat.auckland.ac.nz/~lee/330/lects/762slides1.pdf>-(Erişim tarihi:27.04.2016).

- Miller, A., 2015. The Geometry of Linear Models Lecture Notes 2. Auckland, https://www.stat.auckland.ac.nz/~stats330/lects/762_slides2_2015.pdf-(Erişim tarihi:27.04.2016).
- Miller, A., 2015. The Geometry of Linear Models Lecture Notes 3. Auckland, https://www.stat.auckland.ac.nz/~stats330/lects/762_slides3_2015.pdf-(Erişim tarihi:27.04.2016).
- Miller, A., 2015. The Geometry of Linear Models Lecture Notes 4. Auckland, https://www.stat.auckland.ac.nz/~stats330/lects/762_slides4_2015.pdf-(Erişim tarihi:27.04.2016).
- Monahan, J. F. 2008. A Primer on Linear Models. Chapman & Hall / CRC, New York, 276.
- Morrison, D. F. 1976. Multivariate Statistical Methods, Sec ed. McGraw-Hill, New York.
- Olson, L. 1976. On Choosing a Test Statistics in Multivariate Analysis of Variance. Psychological Bulletin, 83: 579-586.
- Puntanen, S., Styan, G.P.H. Isolato, J. 2006. Matrix Tricks for Linear Statistical Models. Springer-Verlag, New York, 486.
- Rao, C. R. 1973. Linear Statistical Inference and Its Applications. John Wiley, New York.
- Saville, D. J., Wood, G. R. 1996. Statistical Methods: A Geometric Primer. Springer-Verlag, New York.
- Snee, R. D. 1973. Some Aspects of Nonorthogonal Data Analysis, Part 1 Developing Prediction Equations. Journal of Quality Technology, 5: 67-79.
- Stein, C. M. 1960. Multiple Regression Contributions to Probability and Statistics in Essays in Honor of Harold Hotelling ed. 1. Olkin. Stanford University Press, 424-443.
- Strang, G. 1988. Linear Algebra and its Applications 3rd ed. Harcourt Brace Javanovich, San Diego.
- Tebbs, J. M. 2010. Linear statistical models stat 714 lecture notes. The University of South Carolina. <http://people.stat.sc.edu/Tebbs/stat714/f10notes.pdf> - - (Erişim tarihi:27.04.2016).
- Webster, J. T., Gunst, R. F., Mason, R. L. 1974. Latent Root Regression Analysis. Technometrics, 16: 513-522.
- Weisberg, S. 1980. Applied Linear Regression. John Wiley, New York, 218.
- Wickens, T. D. 1995. The Geometry of Multivariate Statistics. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

ÖZGEÇMİŞ

Adı Soyadı : Hakan GEZGİN
Doğum Yeri : Ünye
Doğum Tarihi : 18.10.1990
Yabancı Dili : İngilizce
E-mail : hakangezgin@gmail.com
İletişim Bilgileri : Ordu Üniversitesi Fen Edebiyat Fakültesi

Öğrenim Durumu :

Derece	Bölüm/ Program	Üniversite	Yıl
Lisans	Matematik	Mustafa Kemal Üniversitesi	2014

İş Deneyimi:

Görev	Görev Yeri	Yıl
Öğretmen	Eğitim Bilimleri Ortaokulu Avcılar/İSTANBUL	2014