T. C.

ORDU ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

TAM RANKLI OLMAYAN LİNEER MODELLERDE TAHMİN EDİLEBİLİRLİK, LİNEER KISITLAMALAR (EŞİTLİK/EŞİTSİZLİK) ALTINDA PARAMETRE TAHMİNLERİ, HİPOTEZ TESTLERİ VE BUNLARA AİT BAZI GEOMETRİK YAKLAŞIMLAR

HAKAN GEZGİN

DOKTORA TEZİ

MATEMATİK ANABİLİM DALI

ORDU 2021

TEZ BİLDİRİMİ

Tez yazım kurallarına uygun olarak hazırlanan ve kullanılan intihal tespit programının sonuçlarına göre; 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

TAM RANKLI OLMAYAN LİNEER MODELLERDE TAHMİN EDİLEBİLİRLİK, LİNEER KISITLAMALAR (EŞİTLİK/EŞİTSİZLİK) ALTINDA PARAMETRE TAHMİNLERİ, HİPOTEZ TESTLERİ VE BUNLARA AİT BAZI GEOMETRİK YAKLAŞIMLAR

HAKAN GEZGİN

ORDU ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

MATEMATİK ANABİLİM DALI

DOKTORA TEZİ, 216 SAYFA

(TEZ DANIŞMANI: PROF. DR. SELAHATTİN MADEN)

Bu tez beş bölümden oluşmaktadır. Birinci bölümde teze giriş verilmiştir. Giriş bölümünde matris teorisi ve bunun lineer modellerde kullanılması üzerine bazı bilgilere yer verilmiştir. İkinci bölümde lineer regresyon modelleri, hipotez testleri ve lineer modellerde tahmin edilebilirlikle ilgili temel bilgilerden söz edilmiştir. Üçüncü bölümde kısıtlanmış ve kısıtlanmamış lineer modeller ve tahminleri arasındaki ilişkiler açıklanmış bunlara geometrik yaklaşımlar yapılmıştır. Bu bölümde öncelikle tam rank durumu ele alınıp daha sonra eksik ranklı modeller üzerinde durulmuştur. Dördüncü bölümde sonuç ve öneriler verilmiştir. Beşinci bölümde kaynaklar listelenmiştir.

Anahtar Kelimeler: Matris, Lineer Model, Kısıtlamalı Model, Rank, Parametre Tahmini, En Küçük Kareler Tahmini, Maksimum Olabilirlik Tahmini, Genelleştirilmiş İnvers.

ABSTRACT

ESTIMABILITY IN THE LESS THAN FULL RANK LINEAR MODELS, PARAMETER ESTIMATIONS AND HYPOTHESIS TESTS SUBJECT TO LINEAR RESTRICTIONS (EQUALITY/INEQUALITY) AND SOME GEOMETRIC APPROACHES CONCERNING WITH THESE

HAKAN GEZGİN

ORDU UNIVERSITY INSTITUTE OF NATURAL AND APPLIED SCIENCES

MATHEMATICS

PHD THESIS, 216 PAGES

(SUPERVISOR: PROF. DR. SELAHATTİN MADEN)

This thesis consists of five chapters. In the first chapter, an introduction to the thesis is given. In the introduction, some information on matrix theory and its use in linear models is given. In the second chapter, it is given linear regression models, hypothesis tests and some basic informations about estimability in linear models. In the third chapter, the relationships between the restricted and unrestricted linear models and parameter estimations in these models are explained and geometric approaches have been given about them. In this section, firstly the models with full rank is discussed and then models with less than full rank are emphasized. In the fourth chapter, it is given some results and suggestions. Finally, it is listed some references in fifth chapter.

Keywords: Matrix, Linear Model, Restricted Linear Model, Rank, Parameter Estimation, Ordinary Least Squares Estimations, Maximum Likelihood Estimation, Generalized Invers.

TEŞEKKÜR

Tez konumun belirlenmesi ve tüm çalışmalarım boyunca her zaman engin bilgi ve deneyimleriyle yolumu aydınlatan ve 11.01.2021 tarihindeki vefatına kadar danışmanlığımı yürüten hocam merhum Sayın Prof. Dr. Cemil YAPAR' ı şükran, rahmet ve minnetle anarken, hocamın vefatı sonrasında danışmanlığımı üstlenerek hiçbir desteğini esirgemeyen hocam Sayın Prof. Dr. Selahattin MADEN' e en içten duygularımla teşekkür eder, saygılarımı sunarım.

Hem bu zorlu ve uzun süreçte hem de hayatım boyunca yanımda olan ve hiçbir desteğini esirgemeyerek ideallerimi gerçekleştirmemde maddi ve manevi katkı sağlayan değerli aileme yürekten teşekkürü bir borç bilirim.

Ayrıca Lisansüstü eğitimim sırasında kendilerinden ders aldığım ve engin tecrübelerinden yararlandığım Ordu Üniversitesi Fen Edebiyat Fakültesi Matematik Bölümündeki tüm değerli hocalarıma teşekkür ederim.

.

İÇİNDEKİLER

ŞEKİL LİSTESİ

Sayfa

ÇİZELGE LİSTESİ

Sayfa

SİMGELER ve KISALTMALAR LİSTESİ

$\mathbb{N}$	: Doğal sayılar kümesi sayılar kümesi
$\mathbb{R}$	: Reel sayılar kümesi
$\mathbb{K}$	: K cismi
$\mathbb{C}$	: Kompleks sayılar kümesi
$\mathbb{K}_{\boldsymbol{n}}^{\boldsymbol{m}}$	: $\mathbb{K}$ cismi üzerinde tanımlı $m \times n$ tipindeki tüm matrislerin kümesi
$\mathbb{C}_{\boldsymbol{n}}^{\boldsymbol{m}}$ **veya** $\mathbb{C}_{\boldsymbol{m,n}}$	: $\mathbb{C}$ üzerinde tanımlı $m \times n$ tipindeki tüm matrislerin kümesi
$\boldsymbol{I_n}$	: $n \times n$ tipindeki birim matris
$\boldsymbol{A}^{\mathrm{T}}$ **veya** $\boldsymbol{A'}$	: A matrisinin transpozu
$\overline{\boldsymbol{A}}$	: A matrisinin eşlenik matrisi (eş matrisi)
$\boldsymbol{A}^*$	: A matrisinin eşlenik transpoz matrisi (Hermitian matrisi)
$\lvert\boldsymbol{A}\rvert$	: A matrisinin determinantı
Ek$(\boldsymbol{A})$	: A matrisinin ek matrisi
$\boldsymbol{A_{ij}}$	: A matrisinin bir a_{ij} elemanının kofaktörü
$\boldsymbol{A^{-1}}$	: A matrisinin inversi
$\boldsymbol{r}(\boldsymbol{A})$	: A matrisinin rankı
$\boldsymbol{\mathcal{N}}(\boldsymbol{A})$	: A matrisinin null (sıfır) uzayı
$\boldsymbol{\Re}(\boldsymbol{A})$	: A matrisinin ranj (sütun) uzayı
$\boldsymbol{A}^-$ **veya** $\boldsymbol{A}^{(\mathbf{1})}$	: A matrisinin genelleştirilmiş inversi (iç inversi)
$\boldsymbol{A}^{(\mathbf{2})}$	: A matrisinin dış inversi
$\boldsymbol{A_0}$ **veya** $\boldsymbol{A}^{(\mathbf{1,2})}$	: A matrisinin yansımalı genelleştirilmiş inversi
$\boldsymbol{A}^{\dagger}$ **veya** $\boldsymbol{A}^{+}$	: A matrisinin Moore–Penrose tipi genelleştirilmiş inversi
BLUE	: En iyi lineer yansız tahmin edici
OLSE	: Alışılmış en küçük kareler tahmin edicisi
$\widehat{\boldsymbol{\beta}}$	: β parametresinin alışılmış en küçük kareler tahmin edicisi
Kov$(\boldsymbol{A},\boldsymbol{B})$	: A ve B değişkenleri arasındaki kovaryans
$\boldsymbol{E}(\boldsymbol{X})$	: X değişkeninin beklenen değeri
Köşeg$\boldsymbol{A}$	: A matrisinin köşegen elemanları
MLE	:En çok olabilirlik tahmini
OLSE	:Alışılmış en küçük kareler tahmin edicisi
SS	: Kareler toplamı
iz$(\boldsymbol{A})$	: A matrisinin izi

1. GİRİŞ

Günümüzde matris teorisi, teorik matematik, istatistik, sosyoloji, kimya, fizik eğitimi ve elektrik mühendisliği gibi çeşitli teknik alanlar içinde gerekli matematiksel temel bilginin ayrılmaz bir parçası haline gelmiştir. Matris hesabı, 19. yüzyıl ortalarından beri bilinmektedir. Bir İngiliz matematikçisi olan Cavley, 1858 yılında zamanında çok meşhur olan '*Memorie on the Theory of Matrices*' isimli çalışmasında matris cebirinin temel esaslarını ortaya koymuştur. Daha sonraları Fransız Laguerre ve Alman Frobenius matrislerle ilgili yeni kavram ve teoremler üzerinde çalışmışlardır.

Bir singüler matrisin inversi fikri ilk defa 1920 yılında Moore tarafından ortaya atılmıştır. Ancak, daha sonra 1955 yılına kadar bu konuda herhangi bir sistematik çalışmaya rastlanamamaktadır. 1955 yılında, önceki çalışmalardan habersiz olarak, Penrose (1955) biraz farklı bir yoldan Moore tarafından verilen invers kavramını tekrar tanımlamıştır. Penrose ile aynı zamanlarda yaşayan bilim adamlarından birisi olan Rao (1965) tarafından geliştirilen Pseuda invers, Moore ve Penrose tarafından ortaya konulan kısıtların tümünü sağlamamaktadır. Bu nedenle bu invers, Moore–Penrose inversten farklıdır. Rao, daha sonraki çalışmalarında, lineer denklemlerle ilgili problemlerinin çözümünde yeterli olacak ve Moore ve Penrose' un vermiş olduğu tanımdan daha zayıf bir tanım ortaya koymuştur. Böyle bir invers, bir genelleştirilmiş invers (g–invers) olarak adlandırılmış ve bunun uygulamaları Rao (1965) nun birçok çalışmasında yer almıştır. Genelleştirilmiş inverslerle ilgili sistematik gelişmeler ve onların çeşitli uygulamaları *Generalized Inverse of Matrices and Its Applications* (Wiley, 1971) adlı kitapta verilmiştir.

Y gözlemlerin $nx1$ mertebeli vektörü (rasgele vektör), $X: nxp$ ($n< p$) mertebeli bilinen sayıların matrisi, β $px1$ mertebeli bilinmeyen parametrelerin vektörü, ε ise $nx1$, rasgele değişkenlerin gözlenebilir olmayan bir vektörü öyle ki $E(\varepsilon)=0$, $Cov(\varepsilon)=\sum\sum(\text{Y-}\widehat{\text{Y}})^2$, olmak üzere bunlar arasında,

$$Y = X\beta + \varepsilon \qquad (1.1)$$

biçiminde varsayılan bağıntıya bir lineer model veya lineer regresyon modeli denir.

Bu model pek çok özel durumlara sahiptir. Bu durumlar, ε rasgele vektörünün dağılımına ve kovaryans matrisine; X katsayı matrisinin yapısına ve rankına bağlıdır. Aksi belirtilmedikçe, $r(X)=p$ olduğunu kabul edeceğiz, yani modelimizdeki X matrisi tam sütun ranklı bir matris olacaktır. ε hata vektörünün dağılımı hakkında aşağıdaki üç durumu göz önüne alacağız:

1. Durum: $\varepsilon \sim N(0, \sigma^2 I)$

2. Durum: ε bilinmeyen bir dağılıma sahiptir ve $E(\varepsilon)=0$ $Cov(\varepsilon)= \sigma^2 I$ dır.

3. Durum: $\mathrm{Cov}(\varepsilon) = \sigma^2 V$, V bilinen pozitif definit bir matristir.

Birinci durumda her bir ε_i , 0 ortalamalı bilinmeyen σ^2 varyanslı normal dağılıma sahiptir ve $\varepsilon_i,\ i=1,2,\ldots,N$ ler bağımsızdır. İkinci durumda, her bir ε_i nin beklenen değeri sıfır ise, ε_i ler ilişkisiz ve ε_i ler bilinmeyen ortak σ^2 varyansına sahiptirler.

Birinci ve ikinci durumdaki varsayımlar altındaki modellere Gauss - Markov modeli denir. İkinci durumdaki modellere ise bazen en küçük kareler modelleri denir. Hata terimi normal dağılımlı olduğunda modellere hipotez modelleri de denir.

$Y = X\beta + \varepsilon$ lineer modelinde $X\beta$ çarpımına modelin deterministik kısmı, Y ve ε vektörlerine ise modelin stokastik kısmı denir. Y vektörü bağımlı değişken, tepki değişkeni, açıklanan değişken denen bir rastgele değişken ile ilgili gözlemlerin vektörüdür. X matrisine tasarım matrisi, açıklayıcı değişkenlerin matrisi, bağımlı değişkenlerin gözlem matrisi gibi isimler verilmektedir. ε vektörüne ise hata vektörü denmektedir. Gerçek dünyadaki olayların lineer model olarak modellenmesi sırasında Y, X, β ve ε çok değişik şekilde anlamlandırılmaktadır. Bazı modellerde Y üretim miktarı, bazılarında boy uzunluğu, bazılarında bir ekonomik değişken olabilir.

ε Doğrusal hareket eden, β_0 hızı ile hareketine başlayan ve ivmesi β_1 olan bir cismin zamana (t ye) bağlı olarak aldığı yol miktarı S, $S = \beta_0 + \beta_1 t$ formülü ile verilir. Böyle bir hareket yapan bir cismin hareket hızını ve ivmesini "belirlemek" istediğimizi ve daha sonra belli bir zamanda aldığı yol miktarını bilmek istediğimizi düşünelim. Bizim seçtiğimiz belli $t_i,\ i=1,2,\ldots,N$ zamanlarında yol uzunluklarını gözlemlemeye kalkıştığımızda ölçümlerdeki hatalardan dolayı $S_i,\ \ i=1,2,\ldots,N$

gözlemleri için $S_i = \beta_0 + \beta_1 t_i + \varepsilon_i,\ i = 1, 2, ..., N$ gibi bir model düşünmemiz uygun görünmektedir.

$$Y = \begin{bmatrix} S_1 \\ S_2 \\ \cdots \\ S_N \end{bmatrix}_{Nx1}, \ X = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \cdots & \cdots \\ 1 & t_N \end{bmatrix}_{Nx2}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_N \end{bmatrix}_{Nx1}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}_{2x1}$$

gösterimleri altında yukarıda söylenenler,

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

lineer modeli olarak ifade edilmektedir. Bu modelde Y gözlem vektöründeki gözlemleri veren açıklayıcı (bağımlı) değişkeni Y harfi, X matrisinin ikinci sütunundaki gözlemler (bunların hatasız olarak gözlendiğini kabul ettik) ile ilgili bağımsız değişkeni X harfi ve hatayı da ε harfi ile gösterirsek aralarındaki bağıntı,

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{1.2}$$

biçimindedir.

İkinci bir örnek olarak, belli bir tür elmadaki meyve suyu miktarını elmanın ağırlığına bağlı olarak incelemeyi düşünelim. Gerçekte bir elmadaki meyve suyu miktarı sadece elmanın ağırlığına bağlı değildir, ama ağırlık ile meyve suyu arasında bir fonksiyonel bağıntının (bilinmeyen parametrelere göre lineer bir ifade olabilir) varlığını kabul edip gözlemlerin bunu doğrulayıp doğrulamadığını, gözlemlerden çıkıp bir bağıntının bulunmasını ve bunların neticesinde ağırlığa bağlı olarak meyve suyu miktarını "belirlemeyi" (tahmin etmeyi) düşünebiliriz. Bu örnekteki açıklayıcı değişken olan elmanın ağırlığı ile açıklanan (bağımlı) değişken olan elmadaki meyve suyu miktarı birer rasgele değişkendir. Ağırlığı X, meyve suyu miktarını Y ile gösterirsek X ile Y değişkeninin bir ortak dağılımı söz konusu olacaktır.

$E(Y \text{ I } X = x) = g(x)$ ifadesine Y nin X üzerindeki regresyon denklemi dendiğini ve X ile Y değişkeninin ortak dağılımı normal olduğunda bunun, $E(Y \text{ I } X = x) = \beta_0 + \beta_1 x$ biçiminde olduğunu hatırlatalım. Bu takdirde (X, Y) iki boyutlu rasgele değişkeninin dağılımından N birimlik örneklem, (X_i, Y_i) $i = 1, 2, ..., N$ olmak üzere,

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i\ ,\ i = 1, 2, ..., N \qquad \varepsilon_i \sim N(0, \sigma^2 I)$$

veya

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_N \end{bmatrix}_{Nx1}, \quad X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \dots & \dots \\ 1 & X_N \end{bmatrix}_{Nx2}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_N \end{bmatrix}_{Nx1}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}_{2x1}$$

gösterimi altında,

$$Y = X\beta + \varepsilon\,, \ \varepsilon \sim N(0, \sigma^2 I),$$

modeline bir basit lineer regresyon modeli denir.

Lineer regresyon modelleri de Lineer Modeller çerçevesinde düşünülebilir. X ile Y rasgele değişkenleri arasında bir ortak dağılım düşünmeden sadece Y bağımlı değişken ile ilgili gözlemlere dayalı olarak,

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i\,,\ i = 1, 2, \dots, N \tag{1.3}$$

biçiminde bir ifade söz konusu olduğunda modele basit lineer model denir.

Elmanın ağırlığı X değişkeni ile elmadaki meyve suyu miktarı Y değişkeninin ortak dağılımı normal olmayabilir. Amacımız X değişkeninin gözlenen değerine bağlı olarak Y değişkeninin gözlenen değerini ön görmek olduğunda,

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i\,,\ i = 1, 2, ..., N$$

biçiminde bir lineer model söz konusudur. Bu durumda e hata terimi, birinci örnekteki yol uzunluğunun ölçülmesi sırasındaki hataya benzer bir hatayı içermekle birlikte, X değişkeninin belli bir değeri için Y değişkenindeki rastgeleliği ve ayrıca model belirlemesindeki hatayı da içermektedir.

Bir lineer modelde açıklayıcı değişken sayısı birden çok olduğunda bu modele çoklu lineer model (multiple linear model) denir. Bir modelde bağımlı değişken birden çok olduğunda modele çok değişkenli model (multivariate model) denir.

Sıcaklık ile basıncın, sertlik üzerindeki etkisinin fonksiyon biçiminde bir bağıntı ile ifade edilip edilemeyeceği, bu bağıntının biçiminin ne olacağı veya sıcaklık ile basınç değişkenlerinin sertliği ne derece etkileyip etkilemediği gibi sorunlar ilk olarak metalürji biliminin sorunlarıdır. Metalürji biliminin kanunlarına göre sıcaklık ile basıncın sertlik üzerindeki etkisi tam olarak belirlenmiş olabilir, bağıntı biçimsel olarak belirlenmiş ancak içinde bilinmeyen katsayılar vardır veya aralarında bir bağıntı var ama ne olduğu belirlenmemiş olabilir. İlk durumda istatistikçinin yapacağı fazla bir şey kalmamıştır. Belki belirlenmiş olan modelin geçerliliğinin sınanmasında yardımcı olabilir. İkinci ve üçüncü durumlarda istatistikçiye önemli görevler düşmektedir. Amaç belirlendikten sonra (Örneğin bu amaç hangi sıcaklık ve basınçta malzemenin sertliği maksimum olmaktadır olabilir) gözlemlerin alınacağı en iyi deney tasarımının ve ardından istatistiksel sonuç çıkarımı yapılması istatistik biliminin sorunudur.

İkinci bir örnek olarak belli bir m ısır türünün verimini incelemeyi düşünelim. Verim, toprak ve hava ile ilgili birçok tabiat şartı yanında sulama, gübreleme, toprağı işleme gibi bazı etkenlere bağlıdır. Modelleme sırasında, çok karma şık olan gerçek dünyadaki ilişkilerden bazılarını ihmal ederek, verim (Y) için toplam yağış miktarı (X_1), sıcaklık ortalaması (bitkinin yetişmesi boyunca her gün bir defa ölçüleri sıcaklıkların ortalaması (X_2), gübre miktarı (X_1), bir metrekaredeki bitki sayısına (X_4) bağlı olarak,

$$y = \beta_0 + X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + X_4\beta_4 + \varepsilon \tag{1.4}$$

gibi bir modelin geçerli olduğunu varsayalım. Gerek modelin geçerliliğinin sınanması, gerekse geçerli olacak bir modelde aç ıklayıcı değişkenlerin etkilerinin, yani parametrelerin tahmin edilmesi amacıyla yapılacak araştırmada veri toplama safhası uygulamada pek kolay olmayacaktır. Modeldeki yağış miktarı ve sıcaklık ortalaması ile ilgili açıklayıcı değişkenler birer rasgele değişkendir, gübre miktarı ile ilgili açıklayıcı değişken bir deterministik değişken olarak görülebilir. Açıklayıcı değişkenlerin birer rasgele değişken olup olmamasına bakmaksızın, bundan sonra açıklayıcı değişkenler ile ilgili X matrisini, gözlem değerlerinin bir matrisi, yani sabitlerin bir matrisi olarak düşüneceğiz.

2. GENEL BİLGİLER

2.1 Çoklu Lineer Regresyon Modeli

Açıklanan değişken birden fazla açıklayıcı veya bağımsız değişkene bağlı olduğunda, çoklu lineer regresyon modeli olarak adlandırılan regresyon problemi göz önüne alınacaktır. Bu model basit lineer regresyonu iki şekilde genelleştirir. Bu model $E(y)$ ortalama fonksiyonunun birden fazla açıklayıcı değişkene bağlanmasına ve düz doğrulardan üç ya da daha fazlasına sahip olmasına izin verir, bununla birlikte, o keyfi biçimler için izin vermez.

y; $\beta_1, \beta_2, \ldots, \beta_k$ parametreleri vasıtasıyla k sayıda bağımsız (veya açıklayıcı) $X_1, X_2, \ldots, X_k$ değişkenine bağlı olan bağımlı (veya açıklanan) değişkeni göstersin ve bu durumu

$$y = X_1\beta_1 + X_2\beta_2 + \cdots + X_k\beta_k + \varepsilon \qquad (2.1)$$

şeklinde yazalım. Böyle bir modele bir çoklu lineer regresyon modeli denir. $\beta_1, \beta_2, \ldots, \beta_k$ parametreleri sırasıyla $X_1, X_2, \ldots, X_k$ ile ilgili regresyon katsayılarıdır ve ε gözlenen ve uydurulan lineer ilişki arasındaki farkı yansıtan rasgele hata bileşenidir. Böyle bir fark için çeşitli nedenler var olabilir, örneğin, modelde içerilmeyen değişkenlerin ortak etkisi, modelde hesaba katılamayan rasgele faktörler gibi.

Burada $j-$yinci β_j regresyon katsayısının $j-$yinci X_j bağımsız değişkeninde birim başına değişme için y deki beklenen değişmeyi gösterdiğine dikkat edelim. $E(\varepsilon) = 0$ kabul ederek, $\beta_j = \frac{\partial E(y)}{\partial X_j}$ yazılabilir.

Bir modele parametrelere göre lineer olduğunda lineer denecektir. Böyle bir durumda $\frac{\partial y}{\partial \beta_j}$ (veya eşdeğer olarak $\frac{\partial E(y)}{\partial X_j}$), başka β lara bağlı olmamalıdır. Örneğin,

i) $y = \beta_0 + \beta_1 X$ bir lineer modeldir. Çünkü bu model β_0 ve β_1 e göre lineerdir.

ii) $y = \beta_0 X^{\beta_1}$ modeli

$$\log y = \log \beta_0 + \beta_1 \log X$$

veya

$$y^* = \beta_0^* + \beta_1 X^*$$

olarak da yazılabilir yani, bu model β_0^* ve β_1 parametrelerine göre lineerdir, ancak $y^* = log\,y\,,\; X^* = log\,X$ değişkenleri lineer olmadığından dolayı bu model bir lineer model değildir.

iii) $y = \beta_0 + \beta_1 X + \beta_2 X^2$; β_0, β_1 ve β_2 parametrelerine göre lineerdir, fakat X değişkenine göre lineer değildir. Bu nedenle, o bir lineer modeldir.

iv) $y = \beta_0 + \frac{\beta_1}{X-\beta_2}$, parametrelere göre ve her iki değişkene göre lineer değildir. Bu nedenle, bu model lineer olmayan modeldir.

v) $y = \beta_0 + \beta_1 X^{\beta_2}$, parametrelere göre ve her iki değişkene göre lineer değildir. Bu nedenle, o bir lineer olmayan modeldir.

vi) $y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$ modeli $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$ olarak yazılabilen bir kübik polinominal modeldir ki bu $\beta_0, \beta_1, \beta_2, \beta_3$ parametrelerine göre lineerdir ve $X_1 = X, X_2 = X^2, X_3 = X^3$ değişkenlerine göre lineerdir. Bu nedenle o bir lineer modeldir.

Örnek 2.1. Bir kişinin gelir ve eğitimi ilişkilidir. Ortalama olarak, eğitimin daha yüksek düzeyinin daha yüksek gelir sağlayacağı beklenir. Bu nedenle, basit bir regresyon modeli

$$gelir = \beta_0 + \beta_1 eğitim + \varepsilon$$

olarak ifade edilebilir. β_1 in eğitimdeki birim başına değişmeye göre gelirdeki değişmeyi ve bir eğitimsiz kişinin bile bir gelir sahip olabildiği beklenildiğinden, β_0, eğitim sıfır olduğunda geliri yansıtır. Dahası bu model eğitime bakılmaksızın, birçok kişinin yaşlandıklarında gençliklerinden daha yüksek gelire sahip oluğunu ihmal eder. Bu nedenle, β_1 eğitimin marjinal etkisini abartacaktır. Eğer yaş ve eğitim pozitif olarak ilişkilendirilirse, bu takdirde regresyon modeli eğitimdeki bir artış ile gelirdeki tüm gözlenen artışı ilgilendirecektir. Bu nedenle daha iyi bir model,

$$gelir = \beta_0 + \beta_1 eğitim + \beta_2 yaş + \varepsilon$$

olur. Çoğu kez, gelirin erken yaşlardan ziyade daha geç kazanımda artmaya meylettiği gözlenir. Böyle bir olabilirliği yerleştirmek için, modeli

$$gelir = \beta_0 + \beta_1 eğitim + \beta_2 yaş + \beta_3 yaş^2 + \varepsilon$$

modeline genişletebiliriz. Bunun sebebi gerçek yaşam durumundaki regresyon modellemesi için nasıl ilerlediğimizdir. Bu nedenle araştırmacının kaç tane bağımlı ve bağımsız değişkeni niçin ve nasıl seçme kararını almadan önce deneysel şartları ve olayı göz önüne alması gerekir.

2.1.1 Model Kurgusu

Bir deneme n kez yapılsın ve veri aşağıdaki gibi elde edilsin.

Gözlem Numarası	y tepkimesi	$X_1 \quad X_2 \quad \cdots \quad X_k$ açıklayıcı değişkenleri
1	y_1	$\boldsymbol{x}_{11} \quad \boldsymbol{x}_{12} \quad \cdots \quad \boldsymbol{x}_{1k}$
2	y_2	$\boldsymbol{x}_{21} \quad \boldsymbol{x}_{22} \quad \cdots \quad \boldsymbol{x}_{2k}$
$\vdots$	$\vdots$	$\vdots \quad \vdots \quad \ddots \quad \vdots$
n	y_n	$\boldsymbol{x}_{n1} \quad \boldsymbol{x}_{n2} \quad \cdots \quad \boldsymbol{x}_{nk}$

Modelin

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

olduğunu kabul ederek, gözlemlerin $n-$lilerin aynı modele işleyeceği de kabul edilir. Bu nedenle onlar

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + \varepsilon_1$$
$$y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_k x_{2k} + \varepsilon_2$$
$$\vdots$$
$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_k x_{nk} + \varepsilon_n$$

bağıntısını sağlar. Bu n tane denklem

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

olarak yazılabilir. Genel olarak, k sayıda bağımsız değişkene sahip model

$$y = X\beta + \varepsilon$$

olarak ifade edilebilir. Burada $\mathrm{y} = (\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_\mathrm{n})'$ inceleme değişkeni üzerinde n sayıda gözlemin bir $n \times 1$ vektörüdür.

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

k sayıda açıklayıcı değişkenin her birisi üzerinde n sayıda gözlemin bir $n \times k$ matrisidir. $\beta = (\beta_1, \beta_2, \dots, \beta_k)'$ regresyon katsayılarının bir $k\times 1$ vektörü olup $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots, \boldsymbol{\varepsilon}_n)'$ ise rasgele hata bileşenlerinin veya hata teriminin bir $n\times 1$ vektörüdür. Eğer regresyon sabiti mevcut ise, X in birinci sütunu $(1,1,\dots,1)'$ olur. İstatistiksel çıkarımları ortaya koymak için $y = X\beta + \varepsilon$ modelinde bazı varsayımlara ihtiyaç duyulur. Aşağıdaki varsayımlar yapılır:

(i) $E(\boldsymbol{\varepsilon}) = 0.$

(ii) $E(\varepsilon\varepsilon') = \sigma^2 I_n.$

(iii) $r(X) = k$

(iv) X stokastik (rasgele) olmayan bir matristir.

(v) $\varepsilon \sim N(0, \sigma^2 I_n).$

Bu varsayımlar regresyon katsayılarının tahmin edicisinin istatistiksel özelliklerini incelemek için kullanılır. Aşağıdaki varsayım özellikle tahmin edicilerin büyük örneklem özelliklerini incelemek için gereklidir.

(vi) $\lim\limits_{n\to\infty} \left(\frac{X'X}{n}\right) = \Delta$

mevcuttur ve stokastik olmayan ve tekil olmayan bir matristir (sonlu elemanlı). Açıklayıcı değişkenler bazı durumlarda stokastik de olabilir. Burada aksi ifade edilmedikçe X in stokastik olmadığını kabul edilecektir. İfade edilen varsayım altında regresyon katsayı vektörü hakkındaki tahmin ve hipotezi testi problemleri göz önüne alınacaktır.

2.1.2 Parametrelerin Tahmini

Regresyon katsayı vektörünün tahmini için bir genel yöntem, uygun şekilde seçilen bir M fonksiyonu için,

$$\sum_{i=1}^{n} M(\varepsilon_i) = \sum_{i=1}^{n} M(y_i - x_{i1}\beta_1 - x_{i2}\beta_2 - \cdots - x_{ik}\beta_k) \tag{2.2}$$

ifadesini minimumlaştırmaktır. M fonksiyonunun seçiminin bazı örnekleri $M(\boldsymbol{x}) = |\boldsymbol{x}|$, $M(\boldsymbol{x}) = \boldsymbol{x}^2$ veya daha genel olarak $M(\boldsymbol{x}) = |\boldsymbol{x}|^p$ şeklindedir. Bu durumda parametrelerin tahmini için $M(\boldsymbol{x}) = \boldsymbol{x}^2$ ile ilgili olan en küçük kareler yöntemini ve en çok olabilirlik yöntemini göz önüne alalım.

2.1.3 Alışılmış En Küçük Kareler Yöntemi (OLS)

B tüm β vektörlerinin kümesi olsun. Eğer özel bir ek bilgi verilmez ise, B kümesi $k-$boyutlu reel öklid uzayındadır. Amaç ε_i sapmalarının kareleri toplamını, yani, verilen y ve X için,

$$S(\beta)=\sum_{i=1}^{n}\varepsilon_i^2=\varepsilon'\varepsilon=(y-X\beta)'(y-X\beta) \qquad (2.3)$$

ifadesini minimum yapan B den bir $b'=(b_1,b_2,\dots,b_k)$ vektörü bulmaktır. Bu durumda $S(\beta)$ bir reel değerli, konveks ve diferansiyellenebilir bir fonksiyon iken, bir minimum daima mevcut olacaktır. Bunun için

$$S(\beta)=y'y+\beta'X'X\beta-2\beta'X'y$$

yazılabilir. $S(\beta)$ nın β ya göre diferansiyeli alınırsa

$$\frac{\partial S(\beta)}{\partial\beta}=2X'X\beta-2X'y$$

$$\frac{\partial^2 S(\beta)}{\partial\beta^2}=2X'X$$

yazılabilir. Bu durumda normal denklem

$$\frac{\partial S(\beta)}{\partial\beta}=0 \;\Rightarrow X'Xb=X'y \qquad (2.4)$$

dır. Burada aşağıdaki sonuç kullanılmıştır.

Sonuç 2.1. Eğer $f(z)=z'Az$ bir kuadratik form, z bir $m\times 1$ tipinde vektör ve A herhangi bir $m\times m$ tipinde simetrik matris ise, bu takdirde

$$\frac{\partial}{\partial z}f(z)=2Az$$

olacaktır. Bu durumda $r(X)=k$ olduğu kabul edildiğinden, bu durumda $X'X$ pozitif tanımlıdır ve normal denklemin yegane çözümü

$$b=(X'X)^{-1}X'y$$

dir ki bu β nın alışılmış en küçük kareler (OLS) tahmin edici olarak adlandırılır.

X in tam ranklı olmadığı durumda, $(X'X)^{-}$, $X'X$ in genelleştirilmiş inversi ve ω keyfi bir vektör olmak üzere;

$$b=(X'X)^{-}X'y+[I-(X'X)^{-}X'X]\omega$$

olacaktır. $X'X$ in $(X'X)^{-}$ genelleştirilmiş inversi aşağıdaki bağıntıları sağlar:

$$X'X(X'X)^{-}X'X=X'X$$

$$X\left(X'X\right)^{-}X'X = X$$

$$X'X\left(X'X\right)^{-}X' = X'$$

dir.

Teorem 2.1.

(i) $\hat{y} = Xb$; y nin deneysel tahmin edicisi olsun. Bu takdirde $\hat{y}$; $X'Xb = X'y$ nin tüm *b* çözümleri için aynı değere sahiptir.

(ii) $S(\beta)$; $X'Xb = X'y$ nin herhangi bir çözümü için minimuma ulaşır.

İspat. (i) b paramatresi

$$b = \left(X'X\right)^{-}X'y + \left[I - \left(X'X\right)^{-}X'X\right]\omega$$

daki herhangi bir eleman olsun. Bu takdirde $X\left(X'X\right)^{-}X'X = X$ olduğundan,

$$Xb = X\left(X'X\right)^{-}X'y + X\left[I - \left(X'X\right)^{-}X'X\right]\omega = X\left(X'X\right)^{-}X'y$$

olur ki bu ω ya bağlı değildir. Bu ise, $\hat{y}$ nın $X'Xb = X'y$ nin her *b* çözümü için aynı değere sahip olduğunu belirtir.

(ii) Herhangi β için,

$$\begin{aligned}
S(\beta) &= \left[y - Xb + X(b-\beta)\right]'\left[y - Xb + X(b-\beta)\right] \\
&= (y-Xb)'(y-Xb) + (b-\beta)'X'X(b-\beta) + 2(b-\beta)'X'(y-Xb) \\
&= (y-Xb)'(y-Xb) + (b-\beta)'X'X(b-\beta) \qquad \left(X'Xb = X'y \text{ kullanarak}\right) \\
&\geq (y-Xb)'(y-Xb) = S(b) \\
&= y'y - 2y'Xb + b'X'Xb \\
&= y'y - b'X'Xb \\
&= y'y - \hat{y}'\hat{y}
\end{aligned}$$

olduğuna dikkat edelim.

Eğer $y = X\beta + \varepsilon$ modeli için, $\hat{\beta}$ tahmini β nın herhangi bir tahmin edicisi ise, bu takdirde uydurulan değerler $\hat{y} = X\hat{\beta}$ olarak tanımlanır.

$\hat{\beta} = b$ olması durumunda,

$$\hat{y} = Xb = X\left(X'X\right)^{-1}X'y = Hy \tag{2.5}$$

dir. Burada $H = X\left(X'X\right)^{-1}X'$ matrisi Şapka matrisi olarak adlandırılır. Bu şekilde tanımlanan H matrisi simetriktir, idempotentdir (yani, $HH = H$ dir) ve

$$iz\left(H\right) = iz\left[X\left(X'X\right)^{-1}X'\right] = iz\left[X'X\left(X'X\right)^{-1}\right] = iz\left(I_k\right) = k \tag{2.6}$$

dır.

İnceleme (tepkime) değişkeninin gözlenen ve uydurulmuş değerleri arasındaki fark tahmin edilen hata olarak adlandırılır. Bu durumda; $\bar{H} = I - H$ olmak üzere,

$$e = y - \hat{y} = \bar{H}y$$

olarak gösterilebilir. Burada kolayca görülebilir ki $\bar{H}$ bir simetrik ve idempotent matris olup,

$$iz(\bar{H}) = iz(I_n) - iz(H) = (n-k)$$

eşitliği sağlanır.

2.1.4 OLSE'nin Özellikleri

(i) Tahmin Hatası: b nin tahmin hatası

$$b - \beta = \left(X'X\right)^{-1}X'y - \beta = \left(X'X\right)^{-1}X'\varepsilon \tag{2.7}$$

şeklindedir.

(ii) Yan (Yanlılık - Bias): X in rasgele olmadığı kabul edildiğinden ve $E\left(\varepsilon\right) = 0$ olduğundan,

$$E\left(b - \beta\right) = \left(X'X\right)^{-1}X'E\left(\varepsilon\right) = 0 \tag{2.8}$$

olacaktır. OLSE (alışılmış en küçük kareler tahmin edicisi) β nın bir yansız tahmin edicisidir.

(iii) b nin kovaryans matrisi

$$Var(b) = \sigma^2 (X'X)^{-1} \tag{2.9}$$

olacaktır.

(iv) *b* nin varyansı, *b* nin kovaryans matrisinin izidir, yani

$$Var(b) = iz[V(b)] = \sum_{i=1}^{k} E(b_i - \beta_i)^2 = \sum_{i=1}^{k} Var(b_i) \tag{2.10}$$

dir.

2.1.5 σ^2 nin Tahmini

σ^2, $S(\beta)$ da görünmediğinden dolayı en küçük kareler kriteri σ^2 yi tahmin etmek için kullanılamaz. $E(\varepsilon_i^2) = \sigma^2$ olduğundan σ^2 yi tahmin etmek için e_i hata tahminleri kullanılarak aşağıdaki yol izlenebilir:

$$e = y - \hat{y} = \bar{H}y$$

olup tahmin edilen hata kareler toplamı,

$$SS_{hata} = \sum_{i=1}^{n} e_i^2 = e'e = y'\bar{H}y = (X\beta + \varepsilon)' \bar{H} (X\beta + \varepsilon) = \varepsilon'\bar{H}\varepsilon \tag{2.11}$$

olarak yazılabilir. Öte yandan $\varepsilon \sim N(0, \sigma^2 I)$ olduğundan, $y \sim N(X\beta, \sigma^2 I)$ dır. Bunun sonucu olarak $y'\bar{H}y \sim \chi^2(n-k)$ elde edilir. Bu nedenle,

$$E[y'\bar{H}y] = (n-k)\sigma^2$$

veya

$$E\left[\frac{y'\bar{H}y}{n-k}\right] = \sigma^2$$

veya

$$E[MS_{hata}] = \sigma^2$$

olduğu görülür. Burada $MS_{hata} = \frac{SS_{hata}}{n-k}$ hata tahminine bağlı ortalama kareler toplamıdır. Bu nedenle, σ^2 nin yansız tahmin edicisi $\hat{\sigma}^2 = MS_{hata} = s^2$ (diyelim) olur. Buradan $\hat{y}$ nın varyansının

$$Var(\hat{y}) = Var(Xb) = XVar(b)X' = \sigma^2 X(X'X)^{-1}X' = \sigma^2 H \tag{2.12}$$

olduğu kolaylıkla görülebilir.

Teorem 2.2. (Gauss – Markov Teoremi): Alışılmış en küçük kareler tahmin edicisi (OLSE) β nın en iyi lineer yansız tahmin edicisi (BLUE) dir.

İspat. β nın OLSE si y nin bir lineer fonksiyonu olan

$$b = (X'X)^{-1}X'y$$

dir. a nın elemanları keyfi sabitler olmak üzere, $l'\beta$ lineer parametrik fonksiyonunun $b^* = a'y$ keyfi lineer tahmin edicisini göz önüne alalım. Bu takdirde b^* için

$$E(b^*) = E(a'y) = a'X\beta$$

yazılabilir ve bu nedenle,

$$E(b^*) = a'X\beta = l'\beta \Rightarrow a'X = l'$$

olduğunda, b^* $l'\beta$ nın yansız tahmin edicisidir. Sadece lineer ve yansız olan tahmin edicileri göz önüne almak istediğimizden dolayı kendimizi $a'X = l'$ için olan tahmin edicilere sınırlayacağız. Ayrıca

$$Var(a'y) = a'Var(y)a = \sigma^2 a'a$$

ve

$$Var(l'b) = l'Var(b)l = \sigma^2 a'X(X'X)^{-1}X'a$$

olacaktır. Öte yandan

$$Var(a'y) - Var(l'b) = \sigma^2\left[a'a - a'X(X'X)^{-1}X'a\right] = \sigma^2 a'(I-H)a$$

eşitliği göz önüne alınırsa $(I-H)$ bir pozitif – yarı tanımlı matris olduğundan

$$Var(a'y) - Var(l'b) \geq 0$$

dır. Bu, eğer b^* herhangi bir lineer yansız tahmin edici ise, bu takdirde onun varyansının b nin varyansından daha küçük olmaması gerektiğini ortaya koyar. Sonuç olarak "en iyi b" nin lineer yansız tahmin ediciler içinde etkin olduğu gerçeğini işaret etmek üzere, b en iyi lineer yansız tahmin edicidir.

2.2 En Çok Olabilirlik Tahmini

$y = X\beta + \varepsilon$ modelinde, hataların σ^2 sabit varyansı ile normal ve bağımsız olarak dağıldıkları veya $\varepsilon \sim N\left(0, \sigma^2 I\right)$ oldukları kabul edilir. Bu durumda hatalar için normal yoğunluk fonksiyonu

$$f(\varepsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}\varepsilon_i^2\right] \quad i = 1, 2, \ldots n \tag{2.13}$$

olacaktır. Olabilirlik fonksiyonu $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ nin

$$\begin{aligned} L\left(\beta, \sigma^2\right) &= \prod_{i=1}^{n} f(\varepsilon_i) \\ &= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\varepsilon_i^2\right] \\ &= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\varepsilon'\varepsilon\right] \\ &= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right] \end{aligned} \tag{2.14}$$

olarak verilen ortak yoğunluğudur. Öte yandan logaritma dönüşümü monoton bir dönüşüm olduğundan $L\left(\beta, \sigma^2\right)$ fonksiyonu yerine $\ln L\left(\beta, \sigma^2\right)$ fonksiyonu da maksimumlaştırılabilir. Bu durumda

$$\ln L\left(\beta, \sigma^2\right) = -\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta) \tag{2.15}$$

olacaktır. β ve σ^2 nin en çok olabilirlik tahmin edicileri (MLE) $\ln L\left(\beta, \sigma^2\right)$ nin β ve σ^2 ye göre birinci mertebeden türevlerini sıfıra eşitleyerek aşağıdaki gibi elde edilir:

$$\frac{\partial \ln L\left(\beta,\sigma^2\right)}{\partial \beta} = \frac{1}{2\sigma^2} 2X'\left(y - X\beta\right) = 0 \tag{2.16a}$$

$$\frac{\partial \ln L\left(\beta,\sigma^2\right)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2}\left(y - X\beta\right)'\left(y - X\beta\right) \tag{2.16b}$$

Bu durumda olabilirlik denklemleri

$$X'X\beta = X'y \tag{2.17}$$

ve

$$\sigma^2 = \frac{1}{n}\left(y - X\beta\right)'\left(y - X\beta\right) \tag{2.18}$$

ile verilir. Bu nedenle $r\left(X\right) = k$ olduğundan, β ve σ^2 nin yegane en çok olabilirlik tahmin edicileri

$$\tilde{\beta} = \left(X'X\right)^{-1} X'y \tag{2.19}$$

ve

$$\tilde{\sigma}^2 = \frac{1}{n}\left(y - X\tilde{\beta}\right)'\left(y - X\tilde{\beta}\right) \tag{2.20}$$

olarak elde edilir. Ayrıca, bu değerlerin olabilirlik fonksiyonu maksimumlaştırdığını doğrulamak için

$$\frac{\partial^2 \ln L\left(\beta,\sigma^2\right)}{\partial \beta^2} = -\frac{1}{\sigma^2} X'X$$

$$\frac{\partial^2 \ln L\left(\beta,\sigma^2\right)}{\partial^2 \left(\sigma^2\right)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\left(y - X\beta\right)'\left(y - X\beta\right)$$

$$\frac{\partial^2 \ln L\left(\beta,\sigma^2\right)}{\partial \beta \partial \sigma^2} = -\frac{1}{\sigma^4} X'\left(y - X\beta\right)$$

ikinci mertebeden türevlerine bakmak gerekir. Bu nedenle $\ln L\left(\beta,\sigma^2\right)$ nin β ve σ^2 ye göre ikinci mertebeden türevlerinin Hessian matrisi

$$\begin{pmatrix} \dfrac{\partial^2 \ln L(\beta,\sigma^2)}{\partial\beta^2} & \dfrac{\partial^2 \ln L(\beta,\sigma^2)}{\partial\beta\partial\sigma^2} \\ \dfrac{\partial^2 \ln L(\beta,\sigma^2)}{\partial\sigma^2\partial\beta} & \dfrac{\partial^2 \ln L(\beta,\sigma^2)}{\partial^2(\sigma^2)^2} \end{pmatrix} \tag{2.21}$$

şeklindedir ki bu matris $\beta = \tilde{\beta}$ ve $\sigma^2 = \tilde{\sigma}^2$ olduğunda negatif tanımlıdır. Bu da olabilirlik fonksiyonunun bu değerlerde maksimumlaştırıldığını garantiler. Dolayısıyla MLE leri OLSE lerle karşılaştırarak aşağıdaki ifadeler yazılabilir.

(i) β nın OLSE ve MLE si aynıdır. Bu nedenle β nın MLE si de β nın yansız bir tahmin edicisidir.

(ii) σ^2 nin OLSE si, $\tilde{\sigma}^2 = \dfrac{n-k}{n} s^2$ olarak σ^2 nin MLE siyle ilgili olan s^2 dir. Bu nedenle, σ^2 nin MLE si σ^2 nın bir yanlı tahmin edicisidir.

2.2.1 Tahmin Edicilerin Tutarlılığı

(i) *b* nin Tutarlılığı: $\lim_{n\to\infty}\left(\frac{X'X}{n}\right) = \Delta$ nın stokastik olmayan ve tekil olmayan bir matris (sonlu elemanlara sahip) olarak mevcut olduğu varsayımı altında,

$$\lim_{n\to\infty} V(b) = \sigma^2 \lim_{n\to\infty} \frac{1}{n}\left(\frac{X'X}{n}\right)^{-1} = \sigma^2 \lim_{n\to\infty} \frac{1}{n}\Delta^{-1} = 0 \tag{2.22}$$

elde edilir. Bu ise, OLSE nin kuadratik anlamda β ya yakınsadığını ifade eder. Bu nedenle OLSE β nın bir tutarlı tahmin edicisidir. Bu durum en çok olabilirlik tahmin edicileri için de doğrudur.

Aynı sonuç olasılığa göre yakınsaklık kavramını kullanarak da gösterilebilir. Eğer herhangi bir $\delta > 0$ için

$$\lim_{n\to\infty} P\left[\left|\hat{\theta}_n - \theta\right| \geq \delta\right] = 0$$

ise $\hat{\theta}_n$ tahmin edicisi olasılığa göre θ ya yakınsar denir ve bu durum $p\lim\left(\hat{\theta}_n\right) = \theta$ olarak gösterilir. Buna göre OLSE nin Tutarlılığı;

$$p\lim\left(\frac{X'\varepsilon}{n}\right)=0$$

olacak şekilde,

$$p\lim\left(\frac{X'X}{n}\right)=\Delta_*$$

nın mevcut; tekil olmayan ve stokastik olmayan matris olduğu daha zayıf varsayımı altında elde edilebilir. Bu durumda

$$b-\beta=\left(X'X\right)^{-1}X'\varepsilon=\left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{n}$$

olduğundan,

$$p\lim\left(b-\beta\right)=p\lim\left(\frac{X'X}{n}\right)^{-1}p\lim\left(\frac{X'\varepsilon}{n}\right)=\Delta_*^{-1}.0=0 \tag{2.23}$$

yazılabilir. Bu nedenle, b tahmini β nın bir tutarlı tahmin edicisi olacaktır. Aynı durum MLE için de doğrudur.

(ii) s^2 nin Tutarlılığı: Şimdi

$$\begin{aligned}
s^2&=\frac{1}{n-k}e'e\\
&=\frac{1}{n-k}\varepsilon'\bar{H}\varepsilon\\
&=\frac{1}{n}\left(1-\frac{k}{n}\right)^{-1}\left[\varepsilon'\varepsilon-\varepsilon'X\left(X'X\right)^{-1}X'\varepsilon\right]\\
&=\left(1-\frac{k}{n}\right)^{-1}\left[\frac{\varepsilon'\varepsilon}{n}-\frac{\varepsilon'X}{n}\left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{n}\right]
\end{aligned}$$

iken, σ^2 nın bir tahmini olarak s^2 nin tutarlılığına bakılabilir. Bunun için $\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2$ den oluşan $\varepsilon'\varepsilon/n$ nin σ^2 ortalamalı, bağımsız ve aynı tür dağılmış rasgele değişkenlerin bir dizisi olduğuna dikkat edelim. Bu durumda Büyük Sayılar Kanununa göre

$$p\lim\left(\frac{\varepsilon'\varepsilon}{n}\right)=\sigma^2$$

$$p\lim\left[\frac{\varepsilon' X}{n}\left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{n}\right]=\left(p\lim\frac{\varepsilon' X}{n}\right)\left[p\lim\left(\frac{X'X}{n}\right)^{-1}\right]\left(p\lim\frac{X'\varepsilon}{n}\right)$$

$$=0.\Delta_*^{-1}.0$$

$$=0$$

$$\Rightarrow p\lim\left(s^2\right)=\left(1-0\right)^{-1}\left[\sigma^2-0\right]$$

$$=\sigma^2$$

yazılabilir. Bu nedenle, s^2 istatistiği σ^2 nin bir tutarlı tahmin edicisi olacaktır. Aynı durum MLE için de doğrudur.

2.2.2 Cramer – Rao Alt Sınırı

$\theta=\left(\beta,\sigma^2\right)'$ olsun. β ve σ^2 nin bilinmediğini kabul edelim. Eğer $E\left(\hat{\theta}\right)=\theta$ ise, bu takdirde $\hat{\theta}$ için Cramer – Rao alt sınırı

$$I\left(\theta\right)=-E\left[\frac{\partial^2 \ln L\left(\theta\right)}{\partial\theta\partial\theta'}\right]$$

$$=\begin{bmatrix} -E\left[\frac{\partial \ln L\left(\beta,\sigma^2\right)}{\partial\beta^2}\right] & -E\left[\frac{\partial \ln L\left(\beta,\sigma^2\right)}{\partial\beta\partial\sigma^2}\right] \\ -E\left[\frac{\partial \ln L\left(\beta,\sigma^2\right)}{\partial\sigma^2\partial\beta}\right] & -E\left[\frac{\partial \ln L\left(\beta,\sigma^2\right)}{\partial^2\left(\sigma^2\right)^2}\right] \end{bmatrix}$$

$$=\begin{bmatrix} -E\left[-\frac{X'X}{\sigma^2}\right] & -E\left[\frac{X'\left(y-X\beta\right)}{\sigma^4}\right] \\ -E\left[\frac{\left(y-X\beta\right)' X}{\sigma^4}\right] & -E\left[\frac{n}{2\sigma^4}-\frac{\left(y-X\beta\right)'\left(y-X\beta\right)}{\sigma^6}\right] \end{bmatrix}$$

$$=\begin{bmatrix} \frac{X'X}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$

matrisinin tersine eşit ya da büyüktür. Bu takdirde

$$\left[I(\theta)\right]^{-1} = \begin{bmatrix} \sigma^2\left(X'X\right)^{-1} & 0 \\ 0 & \dfrac{2\sigma^4}{n} \end{bmatrix} \tag{2.24}$$

matrisi β ve σ^2 nin Cramer–Rao alt sınır matrisidir. Öte yandan β ve σ^2 nin OLSE lerinin kovaryans matrisi

$$\Sigma_{OLS} = \begin{bmatrix} \sigma^2\left(X'X\right)^{-1} & 0 \\ 0 & \dfrac{2\sigma^4}{n-k} \end{bmatrix} \tag{2.25}$$

dır ki bu Cramer–Rao nun sahip olduğu sınıra sadece s^2 için değil b nin kovaryansı için de ulaşılır.

2.2.3 Standartlaştırılmış Regresyon Katsayıları

Genel olarak, $\hat{\beta}_j$ nin büyüklüğü $j-$yinci X_j açıklayıcı değişkeninin ölçümünün birimlerini yansıttığından, regresyon katsayılarını karşılaştırmak oldukça zordur. Örneğin, $\hat{y} = 5 + X_1 + 1000X_2$ uydurulmuş modelinde, y litreler cinsinden, X_1 litreler cinsinden ve X_2 mililitreler cinsinden ölçülür. Her ne kadar $\hat{\beta}_2 > \hat{\beta}_1$ ise de her iki açıklayıcı değişkenin etkisi aynıdır. Diğer değişken sabit tutulduğunda X_1 ve X_2 nin herhangi birindeki bir litrelik değişim $\hat{y}$ da bir litrelik değişime sebep olur.

Bazen, boyutsuz regresyon katsayılarını üreten ölçeklendirilmiş açıklayıcı değişkenler ve açıklanan değişkenle çalışmak yararlıdır. Bu boyutsuz regresyon katsayıları standartlaştırılmış regresyon katsayıları olarak adlandırılır. Standartlaştırılmış regresyon katsayıları veren ölçekleme için iki popüler yaklaşım vardır. Bunlar aşağıdaki gibi verilebilir.

Birim Normal Ölçekleme: Her bir açıklayıcı değişken ve açıklanan değişken için birim normal ölçekleme kullanılabilir. Bu nedenle,

$$s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij} - \overline{x}_j\right)^2$$

ve

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$$

sırasıyla, $j-$yinci açıklayıcı değişkenin ve açıklanan değişkeninin örneklem varyansları olmak üzere,

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \ i = 1, 2, \ldots, n, \ j = 1, 2, \ldots, k$$
$$y_i^* = \frac{y_i - \bar{y}}{s_y} \tag{2.26}$$

yi tanımlayalım. Ölçeklendirilmiş açıklayıcı değişken ve ölçeklendirilmiş açıklanan değişkeninin her biri sıfır ortalamasına ve bir örneklem varyansına sahiptir, yani bu yeni değişkenleri kullanarak model

$$y_i^* = \gamma_1 z_{i1} + \gamma_2 z_{i2} + \ldots + \gamma_k z_{ik} + \varepsilon_i, \ i = 1, 2, \ldots, n \tag{2.27}$$

olarak yazılabilir. Böylece merkezileştirilmiş modelden regresyon sabiti terimi yok edilir. Bu durumda $\gamma = \left(\gamma_1, \gamma_2, \ldots, \gamma_k\right)'$ nin en küçük kareler tahmini

$$\hat{\gamma} = \left(Z'Z\right)^{-1} Z'y^* \tag{2.28}$$

olacaktır. Bu ölçeklendirme bir normal rasgele değişkeni standartlaştırmaya bir benzerliktir, yani gözlem eksi onun ortalaması onun standart sapması ile bölünür. Bu nedenle, bu ölçeklendirme bir normal ölçeklendirme olarak adlandırılır.

Birim Uzunluk Ölçeklendirmesi: Birim uzunluk ölçeklendirmesinde *j*–yinci X_j açıklayıcı değişkeni için düzeltilmiş kareler toplamı

$$S_{jj} = \sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2$$

ve genel kareler toplamı

$$S_T = SS_T = \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$$

olmak üzere,

$$\omega_{ij} = \frac{x_{ij} - \overline{x}_j}{S_{jj}^{1/2}}, \ i = 1, 2, \ldots, n; \ \ j = 1, 2, \ldots, k$$
$$y_i^0 = \frac{y_i - \overline{y}}{SS_T^{1/2}} \tag{2.29}$$

yi tanımlayalım. Bu ölçeklendirmede, her bir yeni W_j açıklayıcı değişkeni $\overline{\omega}_j = \frac{1}{n}\sum_{i=1}^{n} \omega_{ij} = 0$ ortalamasına ve $\sqrt{\sum_{i=1}^{n} (\omega_{ij} - \overline{\omega}_j)^2} = 1$ uzunluğuna sahiptir. Bu durumda bu yeni değişkenler cinsinden regresyon modeli

$$y_i^0 = \delta_1 \omega_{i1} + \delta_2 \omega_{i2} + \ldots + \delta_k \omega_{ik} + \varepsilon_i, \ i = 1, 2, \ldots, n \tag{2.30}$$

şeklindedir. Bunun sonucu olarak $\delta = \left(\delta_1, \delta_2, \ldots, \delta_k\right)'$ regresyon katsayısının en küçük kareler tahmini

$$\hat{\delta} = \left(W'W\right)^{-1} W'y^0 \tag{2.31}$$

olacaktır. Bu durumda, $W'W$ matrisi korelasyon matrisi formundadır, yani

$$r_{ij} = \frac{\sum_{u=1}^{n} \left(x_{ui} - \overline{x}_i\right)\left(x_{uj} - \overline{x}_j\right)}{\left(S_{ii} S_{jj}\right)^{1/2}} = \frac{S_{ij}}{\left(S_{ii} S_{jj}\right)^{1/2}}$$

şeklindedir. X_i ve X_j açıklayıcı değişkenleri arasındaki basit korelasyon katsayısı olmak üzere,

$$W'W = \begin{pmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1k} \\ r_{12} & 1 & r_{23} & \cdots & r_{2k} \\ r_{13} & r_{23} & 1 & \cdots & r_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{1k} & r_{2k} & r_{3k} & \cdots & 1 \end{pmatrix} \tag{2.32}$$

dir. Benzer şekilde, *j*–yinci X_j açıklayıcı değişkeni ve *y* açıklanan değişkeni arasındaki basit korelasyon katsayısı

$$r_{jy} = \frac{\sum_{u=1}^{n} \left(x_{uj} - \overline{x}_j\right)\left(y_u - \overline{y}\right)}{\left(S_{jj} SS_T\right)^{1/2}} = \frac{S_{iy}}{\left(S_{jj} SS_T\right)^{1/2}}$$

olmak üzere,

$$W'y^0 = (r_{1y}, r_{2y}, \ldots, r_{ky})' \qquad (2.33)$$

olacaktır. Öte yandan r_{ij} ve r_{jy} lere X_i ler rasgele değişken olmazsa bile korelasyon katsayısı olarak bakmak geleneğe uygundur.

Eğer birim normal ölçeklendirme kullanılırsa, bu takdirde

$$Z'Z = (n-1)W'W \qquad (2.34)$$

olacaktır. Bu nedenle birim normal ölçeklendirmedeki regresyon katsayısının tahmini ($\hat{\gamma}$) ve uzunluk ölçeklendirmesindeki regresyon katsayısının tahmini ($\hat{\delta}$) aynıdır. Bundan dolayı, hangi ölçeklendirmenin kullanıldığı önemli değildir, bu nedenle $\hat{\gamma} = \hat{\delta}$ dır. Böyle bir ölçeklendirmeden sonra elde edilen regresyon katsayılarına, yani, $\hat{\gamma}$ veya $\hat{\delta}$ ya standartlaştırılmış regresyon katsayıları denir. Orijinal ve standartlaştırılmış regresyon katsayıları arasındaki ilişki

$$b_j = \hat{\delta}_j \left(\frac{SS_T}{S_{jj}} \right)^{1/2}, \quad j = 1, 2, \ldots, k$$

ve

$$b_0 = \bar{y} - \sum_{j=1}^{k} b_j \bar{x}_j$$

şeklindedir. Burada b_0 regresyon sabitinin OLSE si ve b_j ler eğim parametrelerinin OLSE leridir.

2.2.4 Sapma Biçiminde Model

Çoklu lineer regresyon modeli sapma biçiminde de ifade edilebilir. Bunun için ilk önce veri örneklem ortalamasından sapmalara göre ifade edilir. Regresyon parametrelerinin tahmini iki adımda gerçekleştirilir. Birinci adımda eğim parametreleri tahmin edilirken ikinci adımda ise regresyon sabiti tahmin edilir. Sapma formunda çoklu lineer regresyon modeli aşağıdaki gibi ifade edilir:

$$A = I - \frac{1}{n} ll' \qquad (2.35)$$

olsun. Burada $l = (1,1,\dots,1)'$ bir $n\times 1$ vektörüdür. Bu nedenle

$$A=\begin{bmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix}-\frac{1}{n}\begin{bmatrix}1 & 1 & \cdots & 1\\ 1 & 1 & \cdots & 1\\ \vdots & \vdots & \ddots & \vdots\\ 1 & 1 & \cdots & 1\end{bmatrix} \tag{2.36}$$

alınabilir. Buradan

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i=\frac{1}{n}(1,1,\dots,1)\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix}=\frac{1}{n}l'y \tag{2.37}$$

ve

$$Ay=y-l\,\bar{y}=\left(y_1-\bar{y},y_2-\bar{y},\dots,y_n-\bar{y}\right)' \tag{2.38}$$

dir. Bu nedenle herhangi bir sütun vektörünün A ile soldan çarpımı gözlemleri sapma formunda gösteren bir vektörü üretir. Burada

$$Al=l-\frac{1}{n}ll'l=0 \tag{2.39}$$

olup A nın simetrik ve idempotent olduğuna dikkat edelim. Öte yandan

$$y=X\beta+\varepsilon$$

modelinde, β nın OLSE si

$$b=\left(X'X\right)^{-1}X'y \tag{2.40}$$

ve tahmin edilen hata vektörü ise

$$e=y-Xb \tag{2.41}$$

dır. Burada $Ae=e$ olduğuna dikkat edelim.

Eğer $n\times k$ tipindeki X matrisi, $X_1=(1,1,\dots,1)'$ şeklinde $n\times 1$ vektörü X_2^*; $k-1$ tane $X_2,X_3,\dots X_k$ açıklayıcı değişkenin gözlemlerinin $n\times(k-1)$ matrisi olmak üzere,

$$X = \begin{bmatrix} X_1 & X_2^* \end{bmatrix}$$

olarak parçalanır ve $b = \left(b_1, b_2^{*\prime}\right)$ OLSE si, β_1 sabit teriminin OLSE si, b_1 ve b_2, $\beta_2, \beta_3, \ldots, \beta_k$ ile ilgili OLSE lerin bir $(k-1)\times 1$ vektörü olmak üzere, uygun olarak parçalanırsa, bu takdirde

$$y = X_1 b_1 + X_2^* b_2^* + e \tag{2.42}$$

yazılabilir. Bu ifade A ile soldan çarpılırsa

$$Ay = AX_1 b_1 + AX_2^* b_2^* + Ae = AX_2^* b_2^* + e \tag{2.43}$$

elde edilir. Bu durumda bu son ifadeyi $X_2^{*}{}'$ ile soldan çarparak

$$X_2^{*\prime} Ay = X_2^{*\prime} AX_2^* b_2^* + X_2^{*\prime} e = X_2^{*\prime} AX_2^* b_2^* \tag{2.44}$$

bağıntısı elde edilir. Öte yandan A simetrik ve idempotent olduğundan,

$$\left(AX_2^*\right)'\left(Ay\right) = \left(AX_2^*\right)'\left(AX_2^*\right)b_2^* \tag{2.45}$$

yazılabilir. Bu bağıntı $y = X\beta + \varepsilon$ modelindeki $X'y = X'Xb$ normal denklemleri ile karşılaştırılabilir. Böyle bir karşılaştırma aşağıdaki sonuçları verir:

- b_2^* OLSE'nin bir alt vektörüdür.
- Ay sapma vektörü formundaki açıklanan değişken vektörüdür.
- AX_2^* sapma formundaki açıklayıcı değişken matrisidir.
- Bu sapmalara göre bir normal denklem olup çözümü eğim katsayıların OLSE sini

$$b_2^* = \left[\left(AX_2^*\right)'\left(AX_2^*\right)\right]^{-1}\left(AX_2^*\right)'\left(Ay\right) \tag{2.46}$$

olarak verir. Regresyon sabitinin tahmini ikinci adımda aşağıdaki gibi elde edilir. Bunun için $y = Xb + e$ ifadesini $\frac{1}{n}l'$ ile soldan çarparak

$$\frac{1}{n}l'y = \frac{1}{n}l'Xb + \frac{1}{n}l'e$$

$$\bar{y} = \begin{bmatrix} 1 & \bar{X}_2 & \bar{X}_3 & \cdots & \bar{X}_k \end{bmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix} + 0$$

$$\Rightarrow b_1 = \bar{y} - b_2\bar{X}_2 - b_3\bar{X}_3 - \ldots - b_k\bar{X}_k \tag{2.47}$$

olduğu görülür. Şimdi bu modele göre çeşitli kareler toplamlarını açıklayabiliriz. Genel kareler toplamı (TSS) nin ifadesi daha öncekiyle aynı olup

$$TSS = y'Ay$$

ile verilir. Öte yandan $Ay = AX_2^*b_2^* + e$ olup

$$\begin{aligned} y'Ay &= y'AX_2^*b_2^* + y'e \\ &= (Xb + e)'AX_2^*b_2^* + y'e \\ &= (X_1b_1 + X_2^*b_2^* + e)'AX_2^*b_2^* + (X_1b_1 + X_2^*b_2^* + e)'e \\ &= b_2^*{}'X_2^*{}'AX_2^*b_2^* + e'e \end{aligned} \tag{2.48}$$

olduğundan, $TSS = SS_{reg} + SS_{hata}$ dır. Burada regresyona bağlı kareler toplamı $SS_{reg} = b_2^*{}'X_2^*{}'AX_2^*b_2^*$ ve tahmin edilen hata kereler ortalaması ise $SS_{hata} = e'e$ olacaktır.

2.3 Hipotezin Test Edilmesi

Regresyon katsayılarını ilgilendiren hipotezin test edilmesiyle ilgili olarak birkaç önemli soru vardır. Örneğin,

1. Modelin genel yeterliliği (uygunluğu) nedir?
2. Hangi özel açıklayıcı değişken anlamlı olarak görünür?

Bu gibi soruları cevaplamak için, genel çerçevede bir hipotezin, yani genel lineer hipotezin testini ortaya koymak gerekir. Bu takdirde çeşitli hipotez testleri onun özel durumu olarak elde edilebilir. Bu nedenle, bir genel lineer hipotezin testini irdeleyebiliriz.

$H_0 : R\beta = r$ için Hipotezin Testi: Bu kısımda R bilinen elemanların bir $j \times k$ matrisi ve r bilinen elemanların $j \times 1$ vektörü olmak üzere, β daki parametrelerin parametre uzayının $R\beta = r$ olduğu bir alt uzayında ihtiva edilen bir genel lineer hipotez göz önüne alınacaktır. Genel olarak, $H_0 : R\beta = r$ sıfır hipotezi genel lineer hipotez olarak adlandırılır ve bu durumda $H_1 : R\beta \neq r$ alternatif (karşıt) hipotezdir. Burada $r(R) = j$, yani, tam ranklı (tam satır ranklı) olduğununu ve hipotezde lineer bağımlılık olmadığını kabul edelim. $H_0 : R\beta = r$ nin bazı özel durumları ve örneği aşağıdaki gibidir:

(i) $H_0 : \beta_i = 0$ seçelim. Burada 1, R de i–yinci konumda vuku bulur. Bu özel hipotez X_i nin lineer model üzerinde herhangi bir etkiye sahip olmadığını açıklar.

(ii) $H_0 : \beta_3 = \beta_4$ veya $H_0 : \beta_3 - \beta_4 = 0$, $j = 1, r = 0, R = [0,0,1,-1,0 \ldots,0]$ seçelim.

(iii) $H_0 : \beta_3 = \beta_4 = \beta_5$ veya $H_0 : \beta_3 - \beta_4 = 0, \beta_3 - \beta_5 = 0$, $j = 2, r = (0,0)'$,

$$R = \begin{bmatrix} 0 & 0 & 1 & -1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & \ldots & 0 \end{bmatrix}$$

seçelim.

(iv) $H_0 : \beta_3 + 5\beta_4 = 2$ ve $J = 1, r = 2,\ R = [0,0,1,5,0,\ldots,0]$ seçelim.

(v) $H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0$, $J = k-1$, $r = (0,0,\ldots 0)'$

$$R = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_{(k-1)\times k} = \left[0_{(k-1)\times 1} : I_{k-1} \right]$$

Bu özel hipotez uyumun iyiliğini açıklar. Aynı zamanda $X_2, X_3, \ldots, X_k$ nın y nin belirlenmesinde etkili olup olmadığını da test eder. Asıl ilgimiz, açıklayıcı değişkenlerin y deki değişimin onun ortalama değer etrafında olup olmadığını açıklamaya yardım edip etmediğini bilmektir. Bunun için $H_0 : R\beta = r$ için olabilirlik oran testini geliştirelim.

2.3.1 Olabilirlik Oran Testi

Ω tam parametrik uzay ve ω örneklem uzayı olmak üzere, olabilirlik oranı test istatistiği

$$\lambda = \frac{\max L\left(\beta, \sigma^2 \mid y, X\right)}{\max L\left(\beta, \sigma^2 \mid y, X, R\beta = r\right)} = \frac{\hat{L}(\Omega)}{\hat{L}(\omega)} \tag{2.49}$$

şeklindedir. Bu durumda eğer her iki olabilirlik maksimum yapılırsa, bu takdirde kısıtlanmamışın değeri kısıtlanmışın değerinden daha küçük olamayacaktır. Bu nedenle $\lambda \geq 1$ dir.

İlk olarak, $R = I_k$ ve $r = \beta_0$, yani $\beta = \beta_0$ olması basit durumu için, olabilirlik oran testini ele alalım. Bu durumda önce bu küçük model için daha iyi ve ayrıntılı incelemeyi verecek ve daha sonra bu durumu $R\beta = r$ için genelleştireceğiz.

$H_0 : \beta = \beta_0$ için olabilirlik oran testi: β $k \times 1$ vektörü ile ilgili sıfır hipotezi $H_0 : \beta = \beta_0$ olsun. Burada β_0 araştırmacı tarafından belirlenir. β_0 ın elemanları, sıfır dahil herhangi bir değeri alabilir. İlgili alternatif (karşıt) hipotez

$$H_1 : \beta \neq \beta_0$$

olacaktır. $y = X\beta + \varepsilon$ modelinde, $\varepsilon \sim N\left(0, \sigma^2 I\right)$ olduğundan, $y \sim N\left(X\beta, \sigma^2 I\right)$ dır. Bu nedenle, tam parametre uzayı ve örneklem uzayı sırasıyla

$$\Omega : \left\{\left(\beta, \sigma^2\right) : -\infty < \beta_i < \infty, \sigma^2 > 0,\ i = 1, 2, \ldots, k\right\}$$

ve

$$\omega : \left\{\left(\beta, \sigma^2\right) : \beta = \beta_0, \sigma^2 > 0\right\}$$

olur. Ω altında kısıtlanmış olabilirlik

$$L\left(\beta, \sigma^2 \mid y, X\right) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right] \tag{2.50}$$

dır. $\tilde{\beta}$ ve $\tilde{\sigma}^2$, β ve σ^2 nin olabilirlik fonksiyonunu maksimumlaştıran değerler olmak üzere,

$$\tilde{\beta} = \left(X'X \right)^{-1} X'y$$

ve

$$\tilde{\sigma}^2 = \frac{1}{n}\left(y - X\tilde{\beta} \right)'\left(y - X\tilde{\beta} \right)$$

olduğunda bu Ω üzerinde maksimumlaştırılır. Bu nedenle

$$\hat{L}(\Omega) = \max L\left(\beta, \sigma^2 \mid y, X \right)$$

$$= \frac{1}{\left[\frac{2\pi}{n}\left(y - X\tilde{\beta} \right)'\left(y - X\tilde{\beta} \right) \right]^{\frac{n}{2}}} \exp\left[-\frac{\left(y - X\tilde{\beta} \right)'\left(y - X\tilde{\beta} \right)}{\left(\frac{2\left(y - X\tilde{\beta} \right)'\left(y - X\tilde{\beta} \right)}{n} \right)} \right]$$

$$= \frac{n^{n/2} \exp\left(-\frac{n}{2} \right)}{(2\pi)^{n/2}\left[\left(y - X\tilde{\beta} \right)'\left(y - X\tilde{\beta} \right) \right]^{\frac{n}{2}}}$$

dir. ω altında kısıtlanmış olabilirlik ise

$$\hat{L}(\omega) = \max L\left(\beta, \sigma^2 \mid y, X, \beta = \beta_0 \right) = \frac{1}{\left(2\pi\sigma^2 \right)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\left(y - X\beta_0 \right)'\left(y - X\beta_0 \right) \right]$$

şeklindedir. Bu durumda β_0 bilinmediğinden, kısıtlanmış olabilirlik fonksiyonu bir

$$\tilde{\sigma}_\omega^2 = \frac{1}{n}\left(y - X\beta_0 \right)'\left(y - X\beta_0 \right)$$

optimum varyans tahmin edicisine sahiptir. Yani

$$\hat{L}(\omega) = \frac{n^{n/2} \exp\left(-\frac{n}{2} \right)}{(2\pi)^{n/2}\left[\left(y - X\beta_0 \right)'\left(y - X\beta_0 \right) \right]^{\frac{n}{2}}} \tag{2.51}$$

dir. Bu durumda olabilirlik oranı

$$\frac{\hat{L}(\Omega)}{\hat{L}(\omega)}=\frac{\left(\dfrac{n^{n/2}\exp(-n/2)}{(2\pi)^{n/2}\left[\left(y-X\tilde{\beta}\right)'\left(y-X\tilde{\beta}\right)\right]^{n/2}}\right)}{\left(\dfrac{n^{n/2}\exp(-n/2)}{(2\pi)^{n/2}\left[\left(y-X\tilde{\beta}_0\right)'\left(y-X\tilde{\beta}_0\right)\right]^{n/2}}\right)}$$

$$\begin{aligned}&=\left[\frac{(y-X\beta_0)'(y-X\beta_0)}{\left(y-X\tilde{\beta}\right)'\left(y-X\tilde{\beta}\right)}\right]^{n/2}\\&=\left(\frac{\tilde{\sigma}_\omega^2}{\tilde{\sigma}^2}\right)^{n/2}=(\lambda)^{n/2}\end{aligned}\tag{2.52}$$

olacaktır. Burada

$$\lambda=\frac{(y-X\beta_0)'(y-X\beta_0)}{\left(y-X\tilde{\beta}\right)'\left(y-X\tilde{\beta}\right)}\tag{2.53}$$

dır. Bu durumda λ nın ifadesindeki payı aşağıdaki gibi sadeleştirebiliriz:

$$\begin{aligned}&(y-X\beta_0)'(y-X\beta_0)=\left[\left(y-X\tilde{\beta}\right)+X\left(\tilde{\beta}-\beta_0\right)\right]'\left[\left(y-X\tilde{\beta}\right)+X\left(\tilde{\beta}-\beta_0\right)\right]\\&=\left(y-X\tilde{\beta}\right)'\left(y-X\tilde{\beta}\right)+2y'\left[I-X\left(X'X\right)^{-1}X'\right]X\left(\tilde{\beta}-\beta_0\right)+\left(\tilde{\beta}-\beta_0\right)'X'X\left(\tilde{\beta}-\beta_0\right)\\&=\left(y-X\tilde{\beta}\right)'\left(y-X\tilde{\beta}\right)+\left(\tilde{\beta}-\beta_0\right)'X'X\left(\tilde{\beta}-\beta_0\right)\end{aligned}$$

Bu nedenle,

$$\begin{aligned}\lambda&=\frac{\left(y-X\tilde{\beta}\right)'\left(y-X\tilde{\beta}\right)+\left(\tilde{\beta}-\beta_0\right)'X'X\left(\tilde{\beta}-\beta_0\right)}{\left(y-X\tilde{\beta}\right)'\left(y-X\tilde{\beta}\right)}\\&=1+\frac{\left(\tilde{\beta}-\beta_0\right)'X'X\left(\tilde{\beta}-\beta_0\right)}{\left(y-X\tilde{\beta}\right)'\left(y-X\tilde{\beta}\right)}\end{aligned}\tag{2.54}$$

veya buna denk olarak

$$\lambda-1=\lambda_0=\frac{\left(\tilde{\beta}-\beta_0\right)'X'X\left(\tilde{\beta}-\beta_0\right)}{\left(y-X\tilde{\beta}\right)'\left(y-X\tilde{\beta}\right)}\tag{2.55}$$

olacaktır. Burada $0 \le \lambda_0 < \infty$ dur. Bu durumda λ_0 ın dağılımını bulmak için λ_0 da içerilen kuadratik formların dağılımını aşağıdaki gibi yazabiliriz:

$$\left(y - X\tilde{\beta}\right)'\left(y - X\tilde{\beta}\right) = \tilde{e}'\tilde{e} = y'\left[I - X\left(X'X\right)^{-1}X'\right]y = \left(n-k\right)\hat{\sigma}^2 \tag{2.56}$$

dir.

Sonuç 2.2. Eğer $Z \sim N\left(0, \sigma^2 I_n\right)$ bir $n \times 1$ rasgele vektör ve A $n \times n$ tipinde p ranklı herhangi bir simetrik idempotent matris ise, bu takdirde $\frac{Z'AZ}{\sigma^2} \sim \chi^2(p)$ dir. Eğer B, q ranklı başka bir $n \times n$ tipinde simetrik idempotent matris ise, bu takdirde $\frac{Z'BZ}{\sigma^2} \sim \chi^2\left(q\right)$ dur. Eğer $AB = 0$ ise, bu takdirde $Z'AZ$, $Z'BZ$ den bağımsız olarak dağılır. Bu sonucu kullanarak

$$\frac{y'\bar{H}y}{\sigma^2} = \frac{\left(n-k\right)\hat{\sigma}^2}{\sigma^2} \sim \chi^2\left(n-k\right) \tag{2.57}$$

elde edilir. Ayrıca, eğer H_0 doğru ise, $\beta = \beta_0$ dır ve λ_0 daki pay elde edilir. Genel olarak, λ_0 daki payı yeniden yazarak, H, k ranklı bir idempotent matris olmak üzere,

$$\begin{aligned}\left(\tilde{\beta} - \beta\right)'X'X\left(\tilde{\beta} - \beta\right) &= \varepsilon'X\left(X'X\right)^{-1}X'X\left(X'X\right)^{-1}X'\varepsilon \\ &= \varepsilon'H\varepsilon\end{aligned} \tag{2.58}$$

olduğu görülür. Bu nedenle bu sonucu kullanarak,

$$\frac{\varepsilon'H\varepsilon}{\sigma^2} = \frac{\varepsilon'X'\left(X'X\right)^{-1}X'\varepsilon}{\sigma^2} \sim \chi^2\left(k\right) \tag{2.59}$$

sonucuna ulaşılır. Ayrıca, λ_0 ın payındaki $\varepsilon'\bar{H}\varepsilon$ ve paydasındaki $\varepsilon'H\varepsilon$ kuadratik form matrislerinin çarpımı

$$\begin{aligned}&\left[I - X\left(X'X\right)^{-1}X'\right]X\left(X'X\right)^{-1}X' \\ &= X\left(X'X\right)^{-1}X' - X\left(X'X\right)^{-1}X'X\left(X'X\right)^{-1}X' = 0\end{aligned}$$

dır ve bu nedenle λ_0 ın pay ve paydasındaki χ^2 rasgele değişkenleri bağımsızdır. χ^2 rasgele değişkenlerinin her birini onların kendi serbestlik derecesine bölerek, H_0 altında

$$\lambda_1 = \left(\frac{\dfrac{\dfrac{\left(\tilde{\beta}-\beta_0\right)' X' X \left(\tilde{\beta}-\beta_0\right)}{\sigma^2}}{k}}{\dfrac{\dfrac{(n-k)\hat{\sigma}^2}{\sigma^2}}{n-k}} \right)$$

$$= \frac{\left(\tilde{\beta}-\beta_0\right)' X' X \left(\tilde{\beta}-\beta_0\right)}{k\hat{\sigma}^2}$$

$$= \frac{(y-X\beta_0)'(y-X\beta_0)-\left(y-X\tilde{\beta}\right)'\left(y-X\tilde{\beta}\right)}{k\hat{\sigma}^2} \sim F(k,n-k) \qquad (2.60)$$

elde edilir. Burada $(y-X\beta_0)'(y-X\beta_0)$ nın kısıtlanmış hata kareler toplamı ve $\left(y-X\tilde{\beta}\right)'\left(y-X\tilde{\beta}\right)$ nın ise kısıtlanmamış hata kereler toplamı olduğunu belirtelim. Bu durumda λ_1 deki pay kısıtlanmış ve kısıtlanmamış hata kareler toplamı arasındaki fark olacaktır. Dolayısıyla $\lambda_1 \geq F_\alpha(k,n-k)$ olduğunda, kararımız $H_0: \beta = \beta_0$ hipotezini α anlam düzeyinde reddetmektir. Burada $F_\alpha(k,n-k)$, k ve $n-k$ serbestlik dereceli merkezi *F*–dağılımı üzerindeki üst kritik noktalardır.

$H_0: R\beta = r$ için olabilirlik oran testi: $H_0: \beta = \beta_0$ için olabilirlik oran testinin geliştirilmesinde kullanılan aynı mantık ve nedenler $H_0: R\beta = r$ için olabilirlik oran testini geliştirmek için aşağıdaki gibi genişletilebilir:

$$\Omega = \left\{\left(\beta, \sigma^2\right): -\infty < \beta_i < \infty,\ \sigma^2 > 0,\ i = 1,2,\ldots,k\right\}$$

ve

$$\omega = \left\{\left(\beta, \sigma^2\right): -\infty < \beta_i < \infty,\ R\beta = r, \sigma^2 > 0\right\}$$

dır. $\tilde{\beta} = (X'X)^{-1} X'y$ olsun. Bu takdirde

$$E\left(R\tilde{\beta}\right) = R\beta$$

ve

$$Var\left(R\tilde{\beta}\right)=E\left[R\left(\tilde{\beta}-\beta\right)\left(\tilde{\beta}-\beta\right)'R'\right]=RV\left(\tilde{\beta}\right)R'=\sigma^2R\left(X'X\right)^{-1}R'$$

dir. Bu durumda $\tilde{\beta}\sim N\left[\beta,\sigma^2\left(X'X\right)^{-1}\right]$ olduğundan,

$$R\tilde{\beta}\sim N\left[R\beta,\sigma^2R\left(X'X\right)^{-1}R'\right],$$

$$R\tilde{\beta}-r=R\tilde{\beta}-R\beta=R\left(\tilde{\beta}-\beta\right)\sim N\left[0,\sigma^2R\left(X'X\right)^{-1}R'\right]$$

olacaktır. Öte yandan $[R\left(X'X\right)^{-1}R']^{-1}=QQ'$ olacak şekilde bir Q matrisi vardır ve bu takdirde $\xi=QR\left(b-\beta\right)\sim N\left(0,\sigma^2I_n\right)$ dır. Bu nedenle, $H_0:R\beta-r=0$ altında,

$$\begin{aligned}\frac{\xi\xi'}{\sigma^2}&=\frac{\left(R\tilde{\beta}-r\right)'QQ'\left(R\tilde{\beta}-r\right)}{\sigma^2}\\&=\frac{\left(R\tilde{\beta}-r\right)'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\tilde{\beta}-r\right)}{\sigma^2}\\&=\frac{\left(\tilde{\beta}-\beta\right)'R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}R\left(\tilde{\beta}-\beta\right)}{\sigma^2}\\&=\frac{\varepsilon'X\left(X'X\right)^{-1}R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}R\left(X'X\right)^{-1}X'\varepsilon}{\sigma^2}\sim\chi^2\left(J\right)\end{aligned}$$

dir, burada

$$X\left(X'X\right)^{-1}R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}R\left(X'X\right)^{-1}X'$$

bir idempotent matristir. Onun serbestlik derecesi ile ilgili olan izi J dir. Aynı zamanda, H_0 hipotezinin doğru olup olmadığına bakmaksızın

$$\frac{\tilde{e}'\tilde{e}}{\sigma^2}=\frac{\left(y-X\tilde{\beta}\right)'\left(y-X\tilde{\beta}\right)}{\sigma^2}=\frac{y'\bar{H}y}{\sigma^2}=\frac{\left(n-k\right)\hat{\sigma}^2}{\sigma^2}\sim\chi^2\left(n-k\right)$$

dır. Bundan başka, $\tilde{e}'\tilde{e}$ ve $\left(\tilde{\beta}-\beta\right)'R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}R\left(\tilde{\beta}-\beta\right)$ kuadratik form matrislerinin çarpımı, her iki kuadratik formun bağımsız olduğunu belirten, sıfırdır. Bu nedenle, olabilirlik oranı test istatistiği,

$$\lambda_1 = \frac{\dfrac{\dfrac{\left(R\tilde{\beta}-r\right)'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\tilde{\beta}-r\right)}{\sigma^2}}{J}}{\dfrac{\dfrac{\left(n-k\right)\hat{\sigma}^2}{\sigma^2}}{n-k}}$$

$$= \frac{\left(R\tilde{\beta}-r\right)'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\tilde{\beta}-r\right)}{J\hat{\sigma}^2} \sim F\left(J, n-k\right)$$

dır. $F_\alpha\left(J, n-k\right)$, J ve $n-k$ serbestlik dereceli merkezi F dağılımı üzerinde üst kritik noktalar olmak üzere, $\lambda_1 \geq F_\alpha\left(J, n-k\right)$ olduğunda, kararımız H_0 hipotezini reddetmektir.

2.3.2 Regresyon Anlamlılığının Testi (Varyans Analizi)

Eğer $R = \left[0 \quad I_{k-1}\right]$, $r = 0$ alırsak, bu takdirde $H_0 : R\beta = r$ hipotezi aşağıdaki sıfır hipotezine indirgenir:

$$H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0 \qquad (2.61)$$

Bu durumda alternatif (karşıt) hipotez, en az bir $j = 2, 3, \ldots, k$ için, $H_1 : \beta_j \neq 0$ dır. Bu hipotez y ve $X_2, X_3, \ldots, X_k$ açıklayıcı değişkenlerinin herhangi bir kümesi arasında bir lineer ilişkinin var olup olmadığını belirler. Burada X_1 in modeldeki regresyon sabiti terimine karşılık geldiğine ve her $i = 1, 2, \ldots, n$ için $x_{i1} = 1$ olduğuna dikkat edelim. Bu, model uygunluğunun bir geniş kapsamlı testidir. Sıfır hipotezinin reddedilmesi $X_2, X_3, \ldots, X_k$ arasındaki açıklayıcı değişkenlerin en az birinin modele anlamlı olarak katkıda bulunduğunu gösterir. Bu durum varyans analizi olarak adlandırılır. Burada $\varepsilon \sim N\left(0, \sigma^2 I\right)$ olduğundan,

$$y \sim N\left(X\beta, \sigma^2 I\right)$$

ve

$$b = \left(X'X\right)^{-1} X'y \sim N\left[\beta, \sigma^2\left(X'X\right)^{-1}\right]$$

olacaktır. Aynı zamanda,

$$\begin{aligned}\hat{\sigma}^2 &= \frac{SS_{hata}}{n-k} = \frac{(y-\hat{y})'(y-\hat{y})}{n-k} \\ &= \frac{y'\left[I - X\left(X'X\right)^{-1}X'\right]y}{n-k} = \frac{y'\bar{H}y}{n-k} = \frac{y'y - b'X'y}{n-k}\end{aligned} \tag{2.62}$$

dır. $\left(X'X\right)^{-1} X'\bar{H} = 0$ olduğundan b ve $\hat{\sigma}^2$ bağımsız olarak dağılır. $y'\bar{H}y = \varepsilon'\bar{H}\varepsilon$ olup $\bar{H}$ bir idempotent matris olduğundan, $SS_{hata} \sim \chi^2_{(n-k)}$ olacaktır, yani $n-k$ serbestlik dereceli merkezi ki–kare dağılımına sahiptir.

X_2^* alt matrisi; $X_2, X_3, \ldots, X_k$ açıklayıcı değişkenlerini içermek üzere, X matrisini $X = \left[X_1, \quad X_2^*\right]$ olarak ve β_2^* alt vektörü; $\beta_2, \beta_3, \ldots, \beta_k$ regresyon katsayılarını içermek üzere β vektörünü, $\beta = \left[\beta_1, \quad \beta_2^*\right]$ olarak parçalayalım. Buna paralel olarak *y* lere bağlı genel kareler toplamını

$$SS_T = y'Ay = SS_{reg} + SS_{hata} \tag{2.63}$$

olarak parçalayabiliriz. Burada $SS_{reg} = b_2^{*\prime} X_2^{*\prime} A X_2^* b_2^*$ regresyona bağlı kareler toplamıdır ve tahmin edicilerin hata kareler toplamı

$$SS_{hata} = (y - Xb)'(y - Xb) = y'\bar{H}y = SS_T - SS_{reg} \tag{2.64}$$

ile verilir. Ayrıca,

$$\frac{SS_{reg}}{\sigma^2} \sim \chi^2_{k-1}\left(\frac{\beta_2^{*\prime} X_2^{*\prime} A X_2^* \beta_2^*}{2\sigma^2}\right)$$

yani, $\beta_2^{*\prime} X_2^{*\prime} A X_2^* \beta_2^* \big/ 2\sigma^2$ merkezi olmama parametreli, merkezi olmayan χ^2 dağılımı,

$$\frac{SS_T}{\sigma^2} \sim \chi^2_{n-1}\left(\frac{\beta_2^{*\prime} X_2^{*\prime} A X_2^* \beta_2^*}{2\sigma^2}\right)$$

yani, $\beta_2^{*\prime} X_2^{*\prime} A X_2^* \beta_2^* \big/ 2\sigma^2$ merkezi olmama parametreli, merkezi olmayan χ^2 dağılımıdır. $X_2 \bar{H} = 0$ olduğundan, SS_{reg} ve SS_{hata} bağımsız olarak dağılırlar. Bu durumda regresyona bağlı kareler ortalaması

$$MS_{reg} = \frac{SS_{reg}}{k-1}$$

ve hata kereler ortalaması

$$MS_{hata} = \frac{SS_{hata}}{n-k}$$

olacağından

$$\frac{MS_{reg}}{MS_{hata}} \sim F_{k-1,n-k}\left(\frac{\beta_2^{*\prime} X_2^{*\prime} A X_2^* \beta_2^*}{2\sigma^2}\right) \tag{2.65}$$

yazılabilir, yani $(k-1, n-k)$ serbestlik dereceli ve $\beta_2^{*\prime} X_2^{*\prime} A X_2^* \beta_2^* \big/ 2\sigma^2$ merkezi olmama parametreli bir merkezi olmayan *F*–dağılımına sahiptir. Böylece $H_0 : \beta_2 = \beta_3 = \ldots = \beta_k$ hipotezi altında,

$$F = \frac{MS_{reg}}{MS_{hata}} \sim F_{k-1,n-k}$$

dağılmıştır. Bu durumda $F \geq F_\alpha(k-1, n-k)$ olduğunda, α anlam düzeyinde kararımız H_0 hipotezini reddetmek olacaktır.

F–istatistiğinin hesabı aşağıdaki gibi verilen bir varyans analizi (ANOVA) tablosu biçiminde özetlenebilir:

Çizelge 2.1 ANOVA Tablosu

Değişimin kaynağı	Kareler toplamı	Serbestlik dereceleri	Kareler ortalaması	F
Regresyon	SS_{reg}	$k-1$	$MS_{reg} = SS_{reg} / k-1$	F
Hata	SS_{hata}	$n-k$	$MS_{hata} = SS_{hata} / (n-k)$	
Toplam	SS_T	$n-1$		

H_0 hipotezinin reddedilmesi en azından bir $\beta_i \neq 0$; $(i = 1, 2, \ldots, k)$ durumunun olası olduğunu gösterir.

2.3.3 Ayrı Ayrı Regresyon Katsayıları Üzerindeki Hipotezin Testi

Bu kısımda, eğer varyans analizinde test reddedilir ise, sıfır hipotezinin reddedilmesi için regresyon katsayılarının hangisinin sorumlu olduğu/oldukları başka bir soru ortaya çıkar. Böyle regresyon katsayılarına karşılık gelen açıklayıcı değişkenler model için oldukça önemlidir. Böyle açıklayıcı değişkenleri eklemek aynı zamanda uydurulan $\hat{y}$ değerlerinin varyansını da artırır, bu nedenle tepkimeyi açıklamada sadece gerçek değere sahip olan açıklayıcıları eklemeye dikkat edilmelidir. Önemli olmayan açıklayıcı değişkenleri eklemek modelin kullanışlılığını azaltabilir ve hata kare ortalamasını artırabilir. $H_1 : \beta_j \neq 0$ alternatif hipotezine karşı $H_0 : \beta_j = 0$ sıfır hipotezini test etmek basit lineer regresyon modeli durumunda daha evvel tartışılmıştı. Bu durumda eğer H_0 kabul edilir ise bu, X_j açıklayıcı değişkeninin modelden silinebileceğini ifade eder. Karşılık gelen test istatistiği, H_0 altında

$$t = \frac{b_j}{se(b_j)} \sim t(n-k-1)$$

dir. Burada β_j nin b_j OLSE sinin standart hatası C_{jj}; ye karşılık gelen $(X'X)^{-1}$ in b_j ye karşılık gelen $j-$yinci köşegen elemanını göstermek üzere, $se(b_j) = \sqrt{\hat{\sigma}^2 C_{jj}}$

dir. Eğer $|t| > t_{\alpha/2,n-k-1}$ ise, α anlam düzeyinde kararımız H_0 hipotezini reddetmektir. Öte yandan $\hat{\beta}_j$; modelde olan diğer tüm açıklayıcı değişkenlere bağlı olduğundan, bunun sadece bir kısmi test olduğuna dikkat edilmelidir. Bu, diğer açıklayıcı değişkenler modelde verildiğinde X_i nin katkısının bir testidir.

2.4 Güven Aralığı Tahmini

Çoklu regresyon modelinde güven aralıkları ortaklaşa olarak olduğu gibi tek tek regresyon katsayıları için de inşa edilebilir. Bu iki durumu aşağıdaki gibi göz önüne alabiliriz.

2.4.1 Bir Tek Regresyon Katsayısı Üzerinde Güven Aralığı

$y = X\beta + \varepsilon$ modelinde ε_i lerin $N\left(0,\sigma^2\right)$ sınıfından bağımsız ve özdeş olarak dağıldığını kabul ederek,

$$y \sim N\left(X\beta, \sigma^2 I\right)$$

ve

$$b \sim N\left(\beta, \sigma^2\left(X'X\right)^{-1}\right)$$

yazılabilir. Bu nedenle, herhangi bir regresyon katsayı tahmininin marjinal dağılımı, C_{jj}; $\left(X'X\right)^{-1}$ in $j-$yinci köşegen elemanı olmak üzere,

$$b_j \sim N\left(\beta_j, \sigma^2 C_{jj}\right)$$

şeklindedir. Bu nedenle,

$$\hat{\sigma}^2 = \frac{SS_{hata}}{n-k} = \frac{y'y - b'X'y}{n-k}$$

olmak üzere, H_0, $j = 1, 2, \ldots$ altında

$$t_j = \frac{b_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \sim t\left(n-k\right) \tag{2.66}$$

dir. Bu nedenle β_j $(j=1,2,\ldots,k)$ için $100\left(1-\alpha\right)\%$ güven aralığı aşağıdaki gibi elde edilir:

$$P\left[-t_{\frac{\alpha}{2},n-k} \le \frac{b_j-\beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \le t_{\frac{\alpha}{2},n-k}\right]=1-\alpha$$

$$P\left[b_j - t_{\frac{\alpha}{2},n-k}\sqrt{\hat{\sigma}^2 C_{jj}} \le \beta_j \le b_j + t_{\frac{\alpha}{2},n-k}\sqrt{\hat{\sigma}^2 C_{jj}}\right]=1-\alpha$$

Bu nedenle güven aralığı

$$\left(b_j - t_{\frac{\alpha}{2},n-k}\sqrt{\hat{\sigma}^2 C_{jj}}, b_j + t_{\frac{\alpha}{2},n-k}\sqrt{\hat{\sigma}^2 C_{jj}}\right) \tag{2.67}$$

olarak elde edilir.

2.4.2 Regresyon Katsayıları Üzerinde Eş Anlı Güven Aralıkları

$1-\alpha$ olasılıkla eşzamanlı olarak doğru olan bir güven aralığı kümesi eşzamanlı veya ortak güven aralıkları olarak adlandırılır. Çoklu regresyon modelinde β için ortak bir güven bölgesi tanımlamak nispeten kolaydır.

$$\frac{(b-\beta)'X'X(b-\beta)}{kMS_{res}} \sim F_{k,n-k}$$

olduğundan

$$P\left[\frac{(b-\beta)'X'X(b-\beta)}{kMS_{res}} \le F_\alpha(k,n-k)\right]=1-\alpha$$

elde edilir. Bu nedenle β daki tüm parametreler için $100(1-\alpha)\%$ ortak güven aralığı, eliptik bir bölgeyi tanımlayan

$$\frac{(b-\beta)'X'X(b-\beta)}{kMS_{res}} \le F_\alpha(k,n-k) \tag{2.68}$$

biçiminde yazılır.

2. 5 Determinasyon Katsayısı (R^2) ve Ayarlanmış R^2

R; y ve $X_1, X_2, \ldots, X_k$ arasındaki çoklu korelasyon katsayısı olsun. Bu takdirde çoklu korelasyonun karesi R^2 determinasyon katsayısı olarak adlandırılır. R^2 nin değeri genel olarak örneklem regresyon doğrusunun gözlenen veriye ne kadar iyi uyduğunu belirler. Bu aynı zamanda modelin iyiliğinin bir ölçüsü olarak da uygulanır. Regresyon sabiti teriminin

$$y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_k X_{ik} + u_i, \; i = 1, 2, \ldots, n$$

olarak modelde mevcut olduğunu kabul edelim. Bu takdirde; SS_{hata}, tahmin edilen hatalara ait karelerin toplamı; SS_T, genel kareler toplamı ve SS_{reg}, regresyona bağlı karelerin toplamı olmak üzere,

$$R^2 = 1 - \frac{e'e}{\sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2} = 1 - \frac{SS_{hata}}{SS_T} = \frac{SS_{reg}}{SS_T} \tag{2.69}$$

dir. R^2 modelin açıklayıcı gücünü ölçer ki bu dolayısıyla modelin uyumunun iyiliğini yansıtır. O aynı zamanda açıklayıcı değişkenin açıklayıcı gücünün ne kadar olduğu anlamında model uygunluğunu yansıtır.

$$e'e = y' \left[I - X \left(X'X \right)^{-1} X' \right] y = y' \bar{H} y$$

olduğundan,

$$\sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$$

dir. Burada $l = \left(1, 1, \ldots 1\right)'$, $y = \left(y_1, y_2, \ldots, y_n \right)'$ olmak üzere,

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{n} l'y \tag{2.70}$$

dir. Bu nedenle, $A = I - l \left(l'l \right)^{-1} l'$ olmak üzere,

$$
\begin{aligned}
\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2 &= y'y-n\left(\frac{1}{n^2}l'yy'l\right) \\
&= y'y-y'l\frac{1}{n}l'y \\
&= y'y-y'l\left(l'l\right)^{-1}l'y \\
&= y'\left[I-l\left(l'l\right)^{-1}l'\right]y \\
&= y'Ay
\end{aligned}
\qquad (2.71)
$$

biçimindedir. Bu nedenle,

$$R^2=1-\frac{y'\bar{H}y}{y'Ay}$$

olarak yazılabilir. R^2 nin sınırları 0 ve 1 dir, yani, $0\le R^2\le 1$ dir.

$R^2=0$ olması modelin uyumunun zayıf olduğunu gösterir. $R^2=1$ olması modelin en iyi uyumda olduğunu gösterir. $R^2=0.95$, y deki değişimin 95% lik kısmının R^2 tarafından açıklandığını gösterir. Basit bir söyleyişle, model 95% iyidir. Benzer şekilde, R^2 nin 0 ve 1 arasındaki diğer herhangi bir değeri uydurulan modelin uygunluğunu gösterir.

Eğer modele daha fazla açıklayıcı değişken eklenirse, bu takdirde R^2 değeri artar. Değişkenlerin ilgisiz olması durumunda, R^2 yine artacak ve aşırı derecede iyimser bir görüntü verecektir. Aşırı derecede iyimser görüntüde bir düzeltme amacıyla $\bar{R}^2$ veya $adjR^2$ ile gösterilen ayarlanmış R^2

$$\bar{R}^2=1-\frac{SS_{hata}/(n-k)}{SS_T/(n-1)}=1-\left(\frac{n-1}{n-k}\right)\left(1-R^2\right) \qquad (2.72)$$

olarak tanımlanır. Bu durumda $n-k$ ve $n-1$ in SS_{hata} ve SS_T nin dağılımları ile ilgili serbestlik dereceleri olduklarını daha sonra göreceğiz. Ayrıca, e ve y nin ilgili varyanslarının yansız tahmin edicilerine dayanan $SS_{hata}/n-k$ ve $SS_T/n-1$ nicelikleri varyans analizinin kapsamındadır. Eğer fazladan bir değişken $1-R^2$ de çok küçük bir azalmaya neden olursa $n-1/n-k$ artış ile ayarlanmış R^2 küçülecektir. Ayarlanmış R^2 nin başka bir sınırlandırması onun negatif de olabilmesidir. Örneğin, $k=3$, $n=10$, $R^2=0.16$ alınırsa, bu takdirde

$$\bar{R}^2 = 1 - \frac{9}{7} \times 0.97 = -0.25 < 0$$

dır ki bu, yoruma sahip değildir. Bu durumda aşağıdaki sınırlandırmalar verilebilir.

1. Eğer modelde sabit terim yok ise, bu takdirde R^2 tanımlanamaz. Böylesi durumlarda R^2 negatif olabilir.

$y = X\beta + \varepsilon$ modelinde, β nın alışılmış en küçük kareler tahmin edicisi $b = \left(X'X\right)^{-1} X'y$ dir. Uydurulmuş modeli, e tahmin edilen hata olmak üzere,

$$y = Xb + \left(y - Xb\right) = Xb + e$$

olarak göz önüne alalım. $\hat{y} = Xb$ uygun değer ve $l = (1,1,\ldots,1)'$ $n \times 1$ tipinde bir vektör olmak üzere,

$$y - l\,\bar{y} = Xb + e - l\,\bar{y} = \hat{y} + e - l\,\bar{y}$$

olduğuna dikkat edelim. $TSS = \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$ genel kareler toplamı

$$\begin{aligned} TSS &= \left(y - l\,\bar{y}\right)'\left(y - l\,\bar{y}\right) = \left[\left(\hat{y} - l\,\bar{y}\right) + e\right]'\left[\left(\hat{y} - l\,\bar{y}\right) + e\right] \\ &= \left(\hat{y} - l\,\bar{y}\right)'\left(\hat{y} - l\,\bar{y}\right) + e'e + 2\left(\hat{y} - l\,\bar{y}\right)'e \\ &= SS_{reg} + SS_{hata} + 2\left(Xb - l\,\bar{y}\right)'e \\ &= SS_{reg} + SS_{hata} - 2\bar{y}l'e \end{aligned}$$

olarak elde edilir. Fisher–Cochran teoremi varyans analizinin kapsamında $TSS = SS_{reg} + SS_{hata}$ nın gerçeklenmesini gerektirir ve üstelik R^2 nin tanımlanmasını gerektirir. $TSS = SS_{reg} + SS_{hata}$ nın gerçeklenmesi için $l'e$ nin sıfır olmasına, yani, sadece modelde sabit terim var olduğunda mümkün olan $l'e = l'(y - \hat{y}) = 0$ olmasına, gereksinim vardır. Bu iddiayı aşağıdaki gibi gösterebiliriz. İlk olarak, β_1 parametresi $b_1^* = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$ olarak tahmin edilmek üzere,

$$y_i = \beta_1 x_i + \varepsilon_i\,,\ i = 1, 2, \ldots n$$

sabit terimsiz basit lineer regresyon modelini göz önüne alalım. Bu takdirde genel olarak,

$$l'e = \sum_{i=1}^{n} e_i = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right) = \sum_{i=1}^{n} \left(y_i - b_1^* x_i \right) \neq 0$$

dır. Benzer şekilde, bir $y = X\beta + \varepsilon$ sabit terimsiz çoklu lineer regresyon modelinde

$$l'e = l'\left(y - \hat{y} \right) = l'\left(X\beta + \varepsilon - Xb \right) = -l'X\left(b - \beta \right) + l'\varepsilon \neq 0$$

olduğu bulunur. Sonra da, β_0 ve β_1 sırasıyla $b_0 = \bar{y} - b_1\bar{x}$ ve $b_1 = s_{xy} / s_{xx}$ olarak tahmin edilir ve $s_{xy} = \sum_{i=1}^{n} \left(x_i - \bar{x} \right)\left(y_i - \bar{y} \right)$, $s_{xx} = \sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2$, $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ ve $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ olmak üzere, bir $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \left(i = 1, 2, \ldots, n \right)$ sabit terimli basit lineer regresyon modelini göz önüne alınır. Bu durumda

$$\begin{aligned}
l'e &= \sum_{i=1}^{n} e_i = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right) \\
&= \sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i \right) \\
&= \sum_{i=1}^{n} \left(y_i - \bar{y} + b_1\bar{x} - b_1 x_i \right) \\
&= \sum_{i=1}^{n} \left[\left(y_i - \bar{y} \right) - b_1 \left(x_i - \bar{x} \right) \right] \\
&= \sum_{i=1}^{n} \left(y_i - \bar{y} \right) - b_1 \sum_{i=1}^{n} \left(x_i - \bar{x} \right) \\
&= 0
\end{aligned}$$

bulunur. β_0 ve β nın tahmini sırasıyla $\hat{\beta}_0 = \bar{y} - b\bar{x}$ ve $b = \left(X'X \right)^{-1} X'y$ olmak üzere, bir sabit terimine sahip bir

$$y = \beta_0 l + X\beta + \varepsilon \tag{2.73}$$

çoklu lineer regresyon modelinde,

$$\begin{aligned}
l'e &= l'\left(y - \hat{y} \right) \\
&= l'\left(y - \hat{\beta}_0 - Xb \right)
\end{aligned}$$

$$= l'\left(y - \bar{y} + \bar{X}b - Xb\right)$$
$$= l'\left(y - \bar{y}\right) + l'\left(X - \bar{X}\right)b$$
$$= 0$$

olduğunu görülür. Bu nedenle, Fisher–Cochran teoreminin genel kareler toplamının iki ortogonal tümleyene bölünmesi, yani, regresyona bağlı kareler toplamı ve hatalara bağlı kareler toplamına bölünebilmesi anlamında gerçeklenmesi için, $l'e = l'\left(y - \hat{y}\right) = 0$ olması gerektiği ve bunun sadece sabit terim modelde mevcut olduğunda mümkün olduğu sonucuna varılır.

2. R^2 uç değerlere duyarlıdır, bu nedenle sağlamlıktan yoksundur.

3. R^2 modeldeki açıklayıcı değişkenlerin sayısındaki artış ile daima artar. Bu özelliğin asıl mahsuru konuyla ilgili açıklayıcı değişkenler modele eklendiğinde bile R^2 nin yine arttığıdır. Bu aslında doğru olmayan daha iyi bir modelin elde edildiğini gösterir. Aşağıdaki iki modele sahip olduğumuz bir durumu göz önüne alalım.

$$y_i = \beta_1 + \beta_2 X_{i2} + \ldots \beta_k X_{ik} + u_i,\ i = 1, 2, \ldots, n \tag{2.74a}$$

$$\log y_i = \gamma_1 + \gamma_2 X_{i2} + \ldots + \gamma_k X_{ik} + v_i \tag{2.74b}$$

Şimdi sorumuz, hangi modelin daha iyi olduğudur. Birinci model için bir seçenek olarak R^2 yi

$$R_1^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$$

ve ikinci model için bir seçenek R^2 yi

$$R_2^2 = 1 - \frac{\sum_{i=1}^{n}\left(\log y_i - \log \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(\log y_i - \log \bar{y}\right)^2}$$

olarak tanımlamaktır. Böyle olunca da R_1^2 ve R_2^2 kıyaslanabilir değildir. Eğer hala iki modelin kıyaslanmasına gerek duyulursa, R^2 yi tanımlamak için daha iyi bir öneri aşağıdaki gibi olabilir. $\hat{y}_i^* = \log y_i$ olmak üzere,

$$R_3^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - anti\log \hat{y}_i^*\right)}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$$

yazılabilir. Şimdi, karşılaştırmada R_1^2 ve R_3^2 iki modelin uygunluğu hakkında bir fikir verebilir.

2.5.1 Varyans Analizi Testi ve Determinasyon Katsayısının İlişkisi

β_1 in bir sabit terim olduğunu kabul ederek, bu durumda $H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0$ için, varyans analizi testindeki F–istatistiği

$$F = \frac{MS_{reg}}{MS_{hata}} = \frac{(n-k)}{(k-1)} \frac{SS_{reg}}{SS_{hata}} = \left(\frac{n-k}{k-1} \right) \frac{R^2}{1-R^2} \tag{2.75}$$

dir. Burada R^2 determinasyon katsayısıdır. Bu nedenle F ve R^2 yakından ilişkilidir. $R^2 = 0$ olduğunda $F = 0$ dır. Sınırda, $R^2 = 1$ olduğunda $F = \infty$ dur. Bu nedenle F ve R^2 nin her ikisi direkt olarak farklı olur. Yani daha büyük R^2, daha büyük F değerini belirtir. Bu nedenle varyans analizi altında F testi tahmin edilen regresyonun genel öneminin bir ölçüsü olarak adlandırılır. Eğer F hayli anlamlı ise, bu H_0 hipotezini reddedebileceğimizi ifade eder, yani y vektörü X vektörlerine lineer olarak bağlıdır.

2.5.2 İnceleme Değişkeninin Değerlerinin Tahmini

Çoklu regresyon modelinde tahmin iki görünüme sahiptir. Bunlardan birincisi inceleme değişkeninin ortalamasının veya ortalama tepkimenin tahmini, ikincisi ise inceleme değişkeninin gerçek değerinin tahminidir. y nin ortalamasının tahmininde $E(y)$ yi verilen bir $x_0 = (x_{01}, x_{02}, \ldots, x_{0k})'$ de tahmin etmemiz gerekir. Bir nokta tahmini olarak tahmin edici $E(p) = x_0'\beta$ olmak üzere

$$p = x_0'b = x_0'(X'X)^{-1} X'y$$

dır. Bu nedenle, p ifadesi $E(y)$ için bir yansız tahmin edicidir. Bunun varyansı

$$Var(p) = E[p - E(y)]'[p - E(y)] = \sigma^2 x_0'(X'X)^{-1} x_0$$

olacaktır. Bu takdirde

$$E(\hat{y}_0) = x_0'\beta = E(y \mid x_0)$$

ve

$$Var(\hat{y}_0)=\sigma^2 x_0'(X'X)^{-1}x_0$$

olur. Dolayısıyla $x_{01},x_{02},\ldots,x_{0k}$ gibi özel bir noktada ortalama tepkime üzerindeki güven aralığı aşağıdaki gibi bulunabilir:

$x_0=(x_{01},x_{02},\ldots,x_{0k})'$ yü tanımlayalım. x_0 daki uygun değer $\hat{y}_0=x_0'b$ dir. Bu takdirde

$$P\left[-t_{\frac{\alpha}{2},n-k}\le\frac{\hat{y}_0-E(y\,|\,x_0)}{\sqrt{\hat{\sigma}^2x_0'(X'X)^{-1}x_0}}\le t_{\frac{\alpha}{2},n-k}\right]=1-\alpha$$

ve

$$P\left[\hat{y}_0-t_{\frac{\alpha}{2},n-k}\le\sqrt{\hat{\sigma}^2x_0'(X'X)^{-1}x_0}\le E(y\,|\,x_0)\le\hat{y}_0+t_{\frac{\alpha}{2},n-k}\le\sqrt{\hat{\sigma}^2x_0'(X'X)^{-1}x_0}\right]=1-\alpha$$

dır. $x_{01},x_{02},\ldots,x_{0k}$ noktasında ortalama tepkime üzerindeki $100(1-\alpha)\%$ güven aralığı

$$\left[\hat{y}_0-t_{\frac{\alpha}{2},n-k}\sqrt{\hat{\sigma}^2x_0'(X'X)^{-1}x_0},\hat{y}_0+t_{\frac{\alpha}{2},n-k}\sqrt{\hat{\sigma}^2x_0'(X'X)^{-1}x_0}\right] \tag{2.76}$$

olarak ifade edilir. y nin gerçek değerinin tahmininde bir $x_0=(x_{01},x_{02},\ldots,x_{0k})'$ de y yi tahmin etmemiz gerekir. Bir nokta tahmini olarak tahmin edici $E(p_f)=x_0'\beta$ olmak üzere

$$p_f=x_0'b$$

dir. Bu nedenle p_f; y için bir yansız tahmin edicidir. Bunun varyansı ise

$$Var(p_f)=E\left((p_f-y)(p_f-y)'\right)=\sigma^2\left[1+x_0'(X'X)^{-1}x_0\right]$$

olacaktır. Bu gelecek gözlem için $100(1-\alpha)\%$ lık güven aralığı

$$\left[p_f-t_{\frac{\alpha}{2},n-k}\sqrt{\hat{\sigma}^2\left[1+x_0'(X'X)^{-1}x_0\right]},p_f+t_{\frac{\alpha}{2},n-k}\sqrt{\hat{\sigma}^2\left[1+x_0'(X'X)^{-1}x_0\right]}\right] \tag{2.77}$$

olacaktır.

2.6 Lineer En Küçük Kareler

Lineer model regresyon problemlerinde temel tekniktir ve bunun için en önemli araç en küçük kareler uyumudur. Burada, hata karelerinin bir toplamını veya eşdeğer olarak hata karelerinin örneklem ortalamasını minimum yaparız. Beklenen hata karesini minimum yapan regresyon fonksiyonunu bulmayla ilgilendiğimizde, lineer en küçük kareler doğal bir seçimdir. Bu kısımda buna analiz, lineer cebir ve geometrik açıdan bakarak, lineer en küçük kareler tahmininin temel teorisi incelenecektir. Aynı zamanda lineer en küçük kareler için bir dağılım teorisi ve lineer regresyonun hesapsal görünümlerini de irdelenecektir.

2.6.1 En Küçük Kareler Tahminleri

Bu kısımda $y_1, y_2, \ldots, y_n$ gözlenen tepkime değerleri ve $i = 1, \ldots, n$ ve $j = 1, \ldots, p$ için z_{ij} açıklayıcıları ile işe koyulalım ve

$$S\left(\beta_1, \ldots, \beta_p\right) = \sum\nolimits_{i=1}^{n} \left(y_i - \sum\nolimits_{j=1}^{p} z_{ij} \beta_j \right)^2 \tag{2.77}$$

toplamını minimum yapan $\beta_1, \ldots, \beta_p$ parametre değerlerini arayalım. Minimum yapan değerler $\hat{\beta}_1, \ldots, \hat{\beta}_p$ ile gösterilir. S yi minimumlaştırmak için, onun her bir β_j ye göre kısmi türevleri sıfıra eşitlenir. Bu durum

$$\frac{\partial}{\partial \beta_j} S = -2 \sum\nolimits_{i=1}^{n} \left(y_i - \sum\nolimits_{j=1}^{p} z_{ij} \hat{\beta}_j \right)^2 z_{ij} = 0, \quad j = 1, \ldots, p \tag{2.78}$$

eşitliğini sağlar. Daha sonra bunun gerçekten minimum değeri verdiğini maksimum veya bir eyer noktasını vermediğini göstereceğiz. (2.78) deki p tane denklem normal denklemler olarak bilinir. $j-$yinci özellik için değerlerin $Z_{\cdot j} = \left(z_{1j}, \ldots, z_{nj} \right)' \in \mathbb{R}^n$ vektörünü göz önüne alalım. (2.78) bağıntısı bu özellik vektörünün $i-$yinci elemanı $\hat{\varepsilon}_i = y_i - \sum\nolimits_{j=1}^{p} z_{ij} \hat{\beta}_j$ sahip olan hata vektörünün sıfırın bir nokta çarpımına sahip olduğunu söyler.

Her bir özellik vektörü hataların (hata tahminlerinin) vektörüne ortogonaldir (normaldir). Bu durum; (2.77) ve (2.78) bağıntılarını matris notasyonuyla yazarken değerlidir. Bununla birlikte, (2.77) den (2.78) e geçiş koordinatlar ile daha basittir ve sonraki işlemler vektör formunda daha kolaydır.

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \text{ ve } Z = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{pmatrix}$$

ile tepkimeleri ve özellik değerlerini Y vektörü ve Z matrisi içinde gösterelim. Aynı zamanda $\beta = (\beta_1, \ldots, \beta_p)'$ koyalım ve $\hat{\beta}$, kareler toplamının minimumlaştırıcısını göstersin. Son olarak, hataların n vektörü $\hat{\varepsilon} = Y - Z\hat{\beta}$ olsun. Şimdi, (2.77) bağıntısı $S(\beta) = (Y - Z\beta)'(Y - Z\beta)$ olarak yazılabilir ve (2.78) normal denklemlerini – 2 ile böldükten sonra

$$\hat{\varepsilon}' Z = 0 \tag{2.79}$$

olur. Bu normal denklemler

$$Z'Z\hat{\beta} = Z'Y \tag{2.80}$$

olarak da yazılabilir. Bu durumda eğer $Z'Z$ tersinir ise, bu takdirde

$$\hat{\beta} = (Z'Z)^{-1} Z'Y \tag{2.81}$$

olduğu görülür. İlk olarak, $Z'Z$ nin tersinir olduğunu kabul edeceğiz ve daha sonra tekil durumu göz önüne alacağız. (2.81) bağıntısı lineer regresyonda “lineer” sıfatının başka bir anlatımının altını çizer. Tahmin edilen $\hat{\beta}$ katsayıları onu $(Z'Z)^{-1} Z'$ matrisi ile Y yi çarparak elde etmemiz anlamında, Y nin sabit bir lineer kombinasyonudur. $z_0 = \phi(x_0)$ özellikli herhangi bir yeni x_0 noktasında Y nin tahmin edilen değeri de Y ye göre lineerdir. Bu ise $z_0'(Z'Z)^{-1} Z'Y$ dir.

Şimdi, $\hat{\beta} = (Z'Z)^{-1} Z'Y$ nin gerçekten minimumlaştırıcı olduğunu, sadece hata kareler toplamının gradyentinin sıfır olduğu bir nokta olmadığını ispatlayalım. (2.77) gibi bir kareler toplamının, bir maksimum yapan β ya sahip olamadığı açık bir şekilde görülür. Ancak eyer (semer) noktalarına da bakmalıyız ve böylece $\hat{\beta}$ nın bir tek en küçük kareler tahmin edicisi olduğunu da göstermeliyiz.

$\hat{\beta}$ için bir ifade daha evvel bulunduğundan dolayı onun, $\hat{\beta}+(\tilde{\beta}-\hat{\beta})$ gibi bir $\tilde{\beta}\in\mathbb{R}^p$ genelleyicisini ifade ederek ve sonra da $S(\hat{\beta}+(\tilde{\beta}-\hat{\beta}))$ yı açarak, doğru olduğunu gösterelim. Bu ekleme çıkarma tekniği ekseriyetle hata karelerinin öne çıktığı problemlerde uygulanabilir.

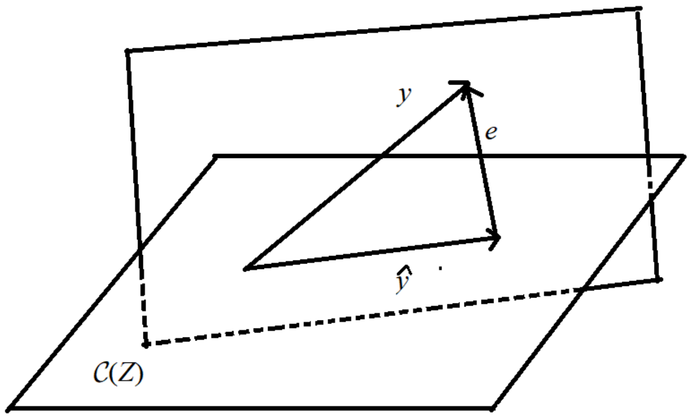
Teorem 2.3. $Z'Z$ tersinir olmak üzere Z $n\times p$ tipinde bir matris olsun ve Y bir $n\times 1$ vektör olsun. $S(\beta)=(Y-Z\beta)'(Y-Z\beta)$ yı tanımlayalım ve $\hat{\beta}=(Z'Z)^{-1}Z'Y$ alalım. Bu takdirde $S(\beta)>S(\hat{\beta})$ olması $\beta\neq\hat{\beta}$ olduğunda sağlanır.

İspat. $Z'\left(Y-Z\hat{\beta}\right)=0$ olduğu bilinmektedir. Ayrıca $\tilde{\beta}$ $\mathbb{R}^p$ de herhangi bir nokta olsun ve $\gamma=\tilde{\beta}-\hat{\beta}$ alınsın. Bu takdirde

$$
\begin{aligned}
S\left(\tilde{\beta}\right)&=\left(Y-Z\tilde{\beta}\right)'\left(Y-Z\tilde{\beta}\right)\\
&=\left(Y-Z\hat{\beta}-Z\gamma\right)'\left(Y-Z\hat{\beta}-Z\gamma\right)\\
&=\left(Y-Z\hat{\beta}\right)'\left(Y-Z\hat{\beta}\right)-\gamma'Z'\left(Y-\hat{\beta}\right)-\left(Y-\hat{\beta}\right)Z\gamma+\gamma'Z'Z\gamma\\
&=S\left(\hat{\beta}\right)+\gamma'Z'Z\gamma
\end{aligned}
$$

olacaktır. Bu nedenle, $S\left(\tilde{\beta}\right)=S\left(\hat{\beta}\right)+\|Z\gamma\|^2\geq S\left(\hat{\beta}\right)$ yazılabilir. $\hat{\beta}$ nın, S nin bir minimumlaştırıcısı olduğu görülür. Teklik için, $\gamma\neq 0$ ifadesinin $Z\gamma\neq 0$ olduğunu belirttiği gösterilmelidir. $\gamma\neq 0$ için aksine olarak $Z\gamma=0$ olduğunu farz edelim. Bu takdirde $\gamma\neq 0$ için $Z'Z\gamma=0$ a sahip olmalıyız, fakat bu $Z'Z$ nin tersinir olduğu varsayımıyla çelişir. Bu nedenle, eğer $\tilde{\beta}=\hat{\beta}$ ise, bu takdirde $S\left(\tilde{\beta}\right)=S\left(\hat{\beta}\right)+\left\|Z\left(\tilde{\beta}-\hat{\beta}\right)\right\|^2>0$ olmak zorundadır. Bu ise ispatı tamamlar.

2.6.2 En Küçük Karelerin Geometrisi



Şekil 2.1 En Küçük Karelerin Geometrisi

Şekil 2.1, lineer en küçük kareleri açıklamak için bir şekli gösterir. $y = (y_1, \ldots, y_n)'$, $\mathbb{R}^n$ deki bir nokta olarak temsil edilir.

$$\mathcal{M} = \{Z\beta \mid \beta \in \mathbb{R}^p\} \tag{2.82}$$

kurgusu $\mathbb{R}^n$ nin p–boyutlu bir lineer alt kümesidir. $Z'Z$ nin tersinir olduğunu ve bu nedenle Z nin p rankına sahip olduğunu kabul ettiğimizden burada $\mathcal{M}$ uzayı tam olarak p boyutludur. Modelimiz altında, $E(Y) = Z\beta \in \mathcal{M}$ dir. En küçük karelerin ardındaki düşünce $\hat{\beta}$ yı ve dolayısıyla $\hat{y} = Z\hat{\beta}$ yı $\mathcal{M}$ içinden y ye en yakın olan nokta olarak bulmaktır. Bu nedenle, $\mathcal{M}$ yi bir dik doğruya iz düşürerek y ye bu en yakın noktayı bulmak amaçlanır. Yani, $\hat{\varepsilon} = y - \hat{y}$ hatası $\mathcal{M}$ içindeki herhangi bir doğruya dik olmalıdır. Bu durumda (2.79) normal denklemlerinden, $\hat{\varepsilon}'Z = 0$ olduğu görülür. Bu nedenle $\hat{\varepsilon}$ her $Z\beta \in \mathcal{M}$ noktasına dolayısıyla $\mathcal{M}$ deki her doğruya diktir.

Herhangi bir $\tilde{\beta} \in \mathbb{R}^p$ için üç $y, \hat{y}$ ve $Z\tilde{\beta}$ noktasını kullanarak bir dik üçgen oluşturulabilir. $\tilde{\beta} = 0$ alarak ve Pisagor teoremini kullanarak

$$\|y\|^2 = \|\hat{\varepsilon}\|^2 + \left\|z\hat{\beta}\right\|^2$$

elde edilebilir. Z nin birinci sütunu 1 lerden oluşuyor ise, bu takdirde $(\bar{y}, \ldots, \bar{y})' = Z(\bar{y}, 0, \ldots, 0) \in \mathcal{M}$ dir ve bu durumda Pisagor teoremi,

$$\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2=\sum_{i=1}^{n}\left(\hat{y}_i-\bar{y}\right)^2+\sum_{i=1}^{n}\left(y_i-\hat{y}\right)^2 \tag{2.83}$$

olduğunu belirtir. (2.83) ün sol tarafına y_i lerin merkezileştirilmiş kareler toplamı denir ve bu $\left(n-1\right)\times Y_i$ nin ortak varyansının alışılmış tahminidir. Bağıntı, karelerinin toplamını iki kısma ayırır, ilki $\hat{y}_i$ uydurulmuş değerlerinin hata karelerinin merkezileştirilen toplamıdır. İkincisi model hatalarının karelerinin toplamıdır. Eğer Z nin birinci sütunu 1 lerden ibaret ise, bu takdirde $\left(1/n\right)\sum_{i=1}^{n}\hat{y}_i=\bar{y}$ olacaktır. Yani, $\bar{\hat{y}}=\bar{y}$ dır. Bu durumda (2.83) deki iki terim $\hat{y}_i$ uygun değerlerinin onların ortalaması civarındaki kareler toplamına ve hataların kareleri toplamına karşılık gelir. Bu durum $SS_{TOT}=SS_{FIT}+SS_{RES}$ olarak yazılabilir. Bu durummda Y deki değişimin bazısı model vasıtasıyla açıklanır ve bazısı açıklanmadan geri kalır. Açıklanan kısım,

$$R^2=\frac{SS_{FIT}}{SS_{TOT}}=1-\frac{SS_{RES}}{SS_{TOT}}$$

ile gösterilir. R^2 niceliği determinasyon (belirleme) katsayısı olarak bilinir. Bu durumda R^2; Y nin $Z\hat{\beta}$ ile ne kadar iyi tahmin edildiğini veya belirlendiğini ölçer. Onun negatif olmayan R kareköküne çoklu korelasyon katsayısı denir. Dolayısıyla R değeri Z de toplu olarak alınan p açıklayıcı ile Y nin ne kadar iyi ilişkili olduğunun bir ölçüsüdür. $z_i=\left(1,x_i\right)\in\mathbb{R}^2$ olmak üzere, basit lineer regresyonun özel durumu için, R^2 değeri x ve y nin alışılagelen Pearson korelasyonunun karesidir. Öte yandan (2.83) bağıntısı ANOVA ayrışımının bir örneğidir.

2.6.3 En Küçük Karelerin Cebiri

En küçük kareler tahminlerini kullanarak, y_i için tahmin edilen değer, $\hat{y}_i=Z_i\hat{\beta}$ dır. Bu durumda uygun değerlerin vektörü $\hat{y}=Z\hat{\beta}=Z\left(Z'Z\right)^{-1}Z'Y$ olarak yazılabilir. Yani, $H=Z\left(Z'Z\right)^{-1}Z'$ olmak üzere, $\hat{y}=Hy$ dir. Burada H matrisi y üzerine şapka koyduğundan, Tukey H için “şapka matrisi” sözcüğünü türetmiştir. Şapka matrisinin bazı basit özellikleri en küçük kareleri yorumlamada önemlidir.

$H^2 = HH$ yazarak, $H^2 = H$ olduğunu görülür, yani H matrisi idempotenttir. Geriye dönüp bakıldığında $H^2 = H$ elde etmemiz gerektiği geometrik olarak açıktır. Herhangi $y \in \mathbb{R}^n$ için, $\mathcal{M}$ nin içinde y ye en yakın nokta Hy dir. Hy önceden $\mathcal{M}$ de olduğundan, $H(Hy) = Hy$ dir. Yani, herhangi bir y için $H^2 y = Hy$ ve bu nedenle $H^2 = H$ dir. Aşikar olarak herhangi bir $k \geq 1$ tam sayısı için $H^k = H$ dir. Öte yandan, $Z'Z$ matrisi simetriktir ve bu nedenle $(Z'Z)^{-1}$ de simetriktir ve buradan H matrisinin de simetrik olduğu görülür. H gibi bir simetrik idempotent matrise bir dik izdüşüm matrisi denir.

Teorem 2.4. H reel değerli, simetrik ve idempotent bir matris olsun. Bu takdirde H nin özdeğerlerinin tümü ya 0 ya da 1 dir.

İspat. x in H nin λ özdeğerli bir öz vektörü olduğunu, bu nedenle, $Hx = \lambda x$ olduğunu farz edelim. H idempotent olduğundan, $H^2 x = Hx = \lambda x$ dir. Aynı zamanda $H^2 x = H(Hx) = H(\lambda x) = \lambda^2 x$ elde edilir. Bu nedenle, $\lambda x = \lambda^2 x$ dir. Özvektörün tanımı $x = 0$ olmasına izin vermez ve bu nedenle $\lambda^2 = \lambda$ elde edilmelidir. Buna göre λ ya 0 ya da 1 dir. $i = 1, \ldots, n$ için P nin sütunları H nin p_i özvektörleri olmak üzere, $H = P'\Lambda P$ olsun. Bu takdirde, Teorem 2.2 ye göre, her bir λ_i özdeğeri 0 veya 1 olmak üzere, $H = \sum_{i=1}^{n} \lambda_i p_i p_i'$ dür. Genelliği yitirmemek üzere, 0 lardan önce gelmek üzere 1 ler için düzenleme yapabiliriz. r sayıda 1 olduğunu farz edelim. Bu takdirde $H = \sum_{i=1}^{r} p_i p_i'$ yazılabilir. Burada kesin olarak r nin p ye eşit olmasını bekleriz ve bu gerçekten sağlanır. H nin özdeğerleri toplamı r dir, bu nedenle $iz(H) = r$ dir. Aynı zamanda,

$$iz(H) = iz\left(Z(Z'Z)^{-1} Z'\right) = iz\left(Z'Z(Z'Z)^{-1}\right) = iz\left(I_p\right) = p$$

yazılabilir. Bu nedenle, $r = p$ ve p_i ler karşılıklı olarak ortogonal n sayıda vektör olmak üzere, $H = \sum_{i=1}^{p} p_i p_i'$ yazılabilir. $i-$yinci gözlem için tahmin, $H_{i\cdot}$; şapka matrisinin $i-$yinci satırı olmak üzere, $\hat{y}_i = H_{i\cdot} y$ olarak yazılabilir. Bu durumda H nin $i-$yinci satırı basit olarak, $z_i'(Z'Z)^{-1} Z'$ olacaktır ve şapka matrisinin ij–yinci elemanı $H_{ij} = z_i'\left(Z'Z\right)^{-1} z_j$ dir. Öte yandan $H_{ij} = H_{ji}$ olduğundan, y_i nin $\hat{y}_j$ ya katkısı y_j nin $\hat{y}_i$ ya katkısına eşittir. Burada şapka matrisinin köşegen elemanları da çok önemlidir. Bu durumda,

$$H_{ii} = z_i'\left(Z'Z\right)^{-1} z_i \tag{2.84}$$

olarak ifade edilir. Şimdi, $\hat{\varepsilon}_i = y_i - \hat{y}_i$ hataları (hata tahminleri) ile de ilgilenelim. Bu durumda hataların tam vektörü $\hat{\varepsilon} = y - \hat{y} = \left(I - H\right)y$ olarak yazılabilir. Eğer H bir dik izdüşüm matrisi ise, $I - H$ de bir dik izdüşüm matrisidir. $\mathcal{M} = \left\{Z\beta \mid \beta \in \mathbb{R}^p\right\}$ model uzayı Z nin sütunlarının lineer kombinasyonlarının kümesidir. $\mathcal{M}$ nin tipik elemanı $\sum_{j=1}^{p} \beta_j Z_{\cdot j}$ dir. Bu nedenle, $\mathcal{M}$; $\mathcal{C}\left(Z\right)$ ile gösterilen Z nin sütun uzayı olarak bilinen uzaydır. H şapka matrisi vektörleri $\mathcal{C}\left(Z\right)$ üzerine dik iz düşürür. $Zv = 0$ olmak üzere, Z ye ortogonal vektörlerin kümesi $\mathbb{R}^n$ nin, Z nin sıfır uzayı olarak bilinen bir lineer alt uzaydır. Bunu $\mathcal{M}^{\perp}$ veya $null\left(Z\right)$ olarak yazabiliriz. Sütun uzayı ve sıfır uzayları ortogonal tümleyenlerdir. Herhangi bir $v \in \mathbb{R}^n$, $v_1 \in \mathcal{M}$ ve $v_2 \in \mathcal{M}^{\perp}$ olmak üzere, $v_1 + v_2$ olarak bir tek şekilde yazılabilir. Şapka matrisine dayanarak, $v_1 = Hv$ ve $v_2 = \left(I - H\right)v$ dir.

2.6.4 Dağılımsal Sonuçlar

X ve *Z* nin bir rasgele olmayan matris ve *Z* nin tam ranka sahip olduğunu farz etmeyi sürdürelim. En küçük kareler modeli $Y = Z\beta + \varepsilon$ biçimine sahiptir. Bu takdirde

$$\hat{\beta} = \left(Z'Z\right)^{-1} Z'Y = \left(Z'Z\right)^{-1}\left(Z'Z\beta + \varepsilon\right) = \beta + \left(Z'Z\right)^{-1} Z'\varepsilon \tag{2.85}$$

olacaktır. (2.85) in sağ yanındaki yegane rasgelelik ε dan gelir. Bu ise $\hat{\beta}$ nın ortalama ve varyansının *x* sabitine göre çalışılmasını kolaylaştırır.

Lemma 2.1. Z matrisi rasgele olmamak üzere, eğer $Y = Z\beta + \varepsilon$, $Z'Z$ tersinir ve $E\left(\varepsilon\right) = 0$ ise, bu takdirde

$$E\left(\hat{\beta}\right)=\beta \tag{2.86}$$

dır.

İspat. Bu durumda $E\left(\hat{\beta}\right)=\beta+\left(Z'Z\right)^{-1}Z'E\left(\varepsilon\right)=\beta$ elde edilir. Bu ise, ispatı tamamlar. ε sıfır ortalamaya sahip olduğu sürece, $\hat{\beta}$ en küçük kareler tahminleri β için yansızdır. Lemma 2.1 hataların normal olarak dağılmış olduğuna ihtiyaç duymaz ve $\operatorname{var}\left(\varepsilon\right)$ hakkında herhangi varsayım bile yapmaz. Bu nedenle $\hat{\beta}$ nın varyansını incelemek için $\operatorname{var}\left(\varepsilon\right)$ üzerinde varsayımlara ihtiyacımız olacak, fakat onun ortalaması hakkında varsayımlara ihtiyacımız olmayacaktır.

Lemma 2.2. Z matrisi rasgele olmamak üzere, eğer $Y=Z\beta+\varepsilon$, $Z'Z$ tersinir ve $0\leq\sigma^2<\infty$ için, $\operatorname{var}\left(\varepsilon\right)=\sigma^2 I$ ise, bu takdirde

$$\operatorname{var}\left(\hat{\beta}\right)=\left(Z'Z\right)^{-1}\sigma^2 \tag{2.87}$$

dir.

İspat. (2.85) bağıntısını kullanarak,

$$\begin{aligned}\operatorname{var}\left(\hat{\beta}\right)&=\left(Z'Z\right)^{-1}Z'\operatorname{var}\left(\varepsilon\right)Z\left(Z'Z\right)^{-1}\\&=\left(Z'Z\right)^{-1}Z'\left(\sigma^2 I\right)Z\left(Z'Z\right)^{-1}=\left(Z'Z\right)^{-1}\sigma^2\end{aligned} \tag{2.88}$$

olduğu görülür.

Birçok özel lineer model için, (2.88) ifadesine bakacağız ve yorumlayacağız. Şimdilik, ε un normal dağılıp dağılmadığına veya onun 0 ortalamaya sahip olup olmadığına bakmaksızın onun gerçeklendiğine dikkat edeceğiz. Şimdiye kadar β nın tahmininin ortalama varyansını inceledik. Bundan sonra dikkatimizi σ^2 ye vereceğiz. σ^2 yi tahmin etmek için anahtar $\hat{\varepsilon}$ hata tahminleri vektörüdür.

Lemma 2.3. Lemma 2.2 nin şartları altında, $E\left(\hat{\varepsilon}'\hat{\varepsilon}\right)=\left(n-p\right)\sigma^2$ dir. $p<n$ için

$$s^2=\tfrac{1}{n-p}\sum_{i=1}^{n}\left(Y_i-z_i'\hat{\beta}\right)^2 \tag{2.89}$$

tahmini $E\left(s^2\right)=\sigma^2$ yi sağlar.

İspat. $\hat{\varepsilon} = (I - H)\varepsilon$ olduğunu hatırlayalım. Bu nedenle,

$$E(\hat{\varepsilon}'\hat{\varepsilon}) = E(\varepsilon'(I-H)\varepsilon) = iz((I-H)I\sigma^2) = (n-p)\sigma^2$$

yazılabilir. Sonuç olarak $s^2 = \hat{\varepsilon}'\hat{\varepsilon}/(n-p)$ dir. Bu ise ispatı tamamlar. Şimdi $\varepsilon \sim N(0, \sigma^2 I)$ olan güçlü varsayım eklentisini yapalım. Normal dağılan hatalar, normal dağılan en küçük kareler tahminlerini, uyumlarını ve hata tahminlerini ortaya koyar.

Teorem 2.5. Z nin $n \times p$ tipinde reel değerli bir matris, $Z'Z$ nin tersinir ve $\varepsilon \sim N(0, \sigma^2 I)$ olmak üzere, modelin $Y = Z\beta + \varepsilon$ olduğunu farz edelim. Bu takdirde, $H = Z(Z'Z)^{-1} Z'$ olmak üzere,

$$\begin{aligned}
&\hat{\beta} \sim N\left(\beta, (Z'Z)^{-1}\sigma^2\right)\\
&\hat{y} \sim N\left(Z\beta, H\sigma^2\right)\\
&\hat{\varepsilon} \sim N\left(0, (I-H)\sigma^2\right)
\end{aligned}$$

dağılımlarına sahiptir. Ayrıca, $\hat{\varepsilon}$; $\hat{\beta}$ dan ve $\hat{y}$ dan bağımsızdır.

İspat. Bunun için

$$v = \begin{pmatrix} \hat{\beta} \\ \hat{y} \\ \hat{\varepsilon} \end{pmatrix} = \begin{pmatrix} (Z'Z)^{-1} Z' \\ H \\ I-H \end{pmatrix} Y \in \mathbb{R}^{p+2n}$$

vektörünü göz önüne alalım. *Y* tepkimesi bir çok değişkenli normal dağılıma sahiptir ve bu nedenle, *v* de öyledir. Bundan dolayı $\hat{\beta}$, $\hat{y}$ ve $\hat{\varepsilon}$ tahminlerinin her biri çok değişkenli normaldir. $HZ = Z$ olduğundan, $\hat{\beta}$, $\hat{y}$ ve $\hat{\varepsilon}$ için yukarıda yazılan ortalamaları saptayan *v* nin beklenen değeri

$$E(v) = \begin{pmatrix} (Z'Z)^{-1} Z' \\ H \\ I-H \end{pmatrix} Z\beta = \begin{pmatrix} \beta \\ HZ\beta \\ (I-H)Z\beta \end{pmatrix} = \begin{pmatrix} \beta \\ Z\beta \\ 0 \end{pmatrix}$$

olarak elde edilir. Öte yandan Lemma 2.2 ye göre, $\hat{\beta}$ nın varyansı $\left(Z'Z\right)^{-1}\sigma^2$ olacaktır. *H* simetrik ve idempotent olduğundan, $\hat{y}$ nın varyansı

$$\operatorname{var}\left(\hat{y}\right)=H\operatorname{var}\left(Y\right)H'=H\left(\sigma^2 I\right)H=H\sigma^2$$

dir. Aynı şekilde, $\hat{\varepsilon}=\left(I-H\right)\left(Z\beta+\varepsilon\right)=\left(I-H\right)\varepsilon$ ve bu nedenle, $I-H$ simetrik ve idempotent olduğundan, $\operatorname{var}\left(\hat{\varepsilon}\right)=\left(I-H\right)\left(\sigma^2 I\right)\left(I-H\right)'=\left(I-H\right)\sigma^2$ yazılabilir. Böylece gösterilen üç dağılım saptandı, fakat iddia edilen bağımsızlık henüz ispatlanmadı. Bunu sona erdirmek için,

$$kov\left(\hat{\beta},\hat{\varepsilon}\right)=\left(Z'Z\right)^{-1}Z'\left(\sigma^2 I\right)\left(I-H\right)=\left(Z'Z\right)^{-1}\left(Z-HZ\right)'\sigma^2=0$$

olduğu ifade edilebilir, çünkü $Z=HZ$ dir. Bu nedenle, $\hat{\beta}$ ve $\hat{\varepsilon}$ ilişkisizdir ve buradan *v* bir normal dağılıma sahip olduğundan dolayı bağımsızdır. Benzer şekilde, ikinci ve son bağımsızlık iddiasını saptayan

$$kov\left(\hat{y},\hat{\varepsilon}\right)=H\left(I-H\right)'\sigma^2=\left(H-HH'\right)\sigma^2=0$$

ifade edilebilir. Öte yandan Teorem 2.3 den şapka matrisi hakkında bir sezgiye göz atabiliriz. Öncelikle $\operatorname{var}\left(\hat{y}_i\right)=\sigma^2 H_{ii}$ olduğundan, $H_{ii}\geq 0$ olduğunu belirtelim. Ayrıca $\operatorname{var}\left(\hat{\varepsilon}\right)=\sigma^2\left(1-H_{ii}\right)$ olduğundan, $1-H_{ii}\geq 0$ elde edilir. Bu nedenle, $i=1,\ldots,n$ için $0\leq H_{ii}\leq 1$ olduğu gerçeklenir.

Teorem 2.6. Teorem 2.5 in şartlarını ve $p<n$ olduğunu kabul edelim. $s^2=\frac{1}{n-p}\sum_{i=1}^{n}\left(Y_i-z_i'\hat{\beta}\right)^2$ olsun. Bu takdirde

$$\left(n-p\right)s^2\sim\sigma^2\chi^2_{(n-p)} \tag{2.90}$$

dir.

İspat. İlk olarak,

$$\left(n-p\right)s^2=\left(Y-Z\hat{\beta}\right)'\left(Y-Z\hat{\beta}\right)=Y'\left(I-H\right)Y=\varepsilon'\left(I-H\right)\varepsilon$$

olduğunu belirtelim. Bu durumda P ortogonal ve $\Lambda = köşeg\left(\lambda_1,\ldots,\lambda_n\right)$ olmak üzere, $I-H$ matrisi, $I-H = P\Lambda P'$ olarak yazılabilir. $I-H=\left(I-H\right)^2$ olduğundan $\lambda_i \in \{0,1\}$ elde edilir. $\tilde{\varepsilon} = P'\varepsilon \sim N\left(0,\sigma^2 I\right)$ olsun. Bu takdirde, k; $\lambda_i = 1$ in sayısı, yani, $k = iz\left(I-H\right)$ olmak üzere,

$$\varepsilon'(I-H)\varepsilon = \tilde{\varepsilon}'\Lambda\tilde{\varepsilon} \sim \sigma^2\chi^2_{(k)}$$

dir. Bu nedenle,

$$k = n - iz\left(H\right) = n - iz\left(Z\left(Z'Z\right)^{-1}Z'\right) = n - iz\left(\left(Z'Z\right)^{-1}Z'Z\right) = n - iz\left(I_p\right) = n-p$$

elde edilir ve böylece ispat tamamlanır.

$\sigma = 0$ olsa bile (2.90) bağıntısının hala geçerli olduğu açıktır. Bu durumda s^2 iyi tanımlı olmadığından teorem ifadesi $n=p$ ile kenar durumunu içerir. Çoğu kez dikkatimizi β nın bileşenlerinin özel bir lineer kombinasyonu üzerinde toplarız ve c bir $p\times 1$ satır vektörü olmak üzere, böyle bir kombinasyonu $c\beta$ olarak yazarız. 1 $j-$yinci hanede olmak üzere, $c=\left(0,\ldots,0,1,0,\ldots 0\right)$ almak $c\beta=\beta_j$ yi verir. Örneğin, eğer $c\beta_j = 0$ ise, bu takdirde j özelliğinin Y yi tahmin etmemize yardım edip etmediği sorusunu sorarız. Daha genel olarak, $\beta_2-\beta_1$ veya $\beta_1-2\beta_2+\beta_3$ gibi parametre kombinasyonları, c nin bir mantıklı seçimini yaparak incelenebilir. $c = z_0'$ almak özellik vektörü z_0 a eşit olduğunda Y nin beklenen değerini $c\beta$ yapar ve $c = \tilde{z}_0' - z_0'$ almak bize $\tilde{z}_0$ ve z_0 özelliklerinde $E\left(Y\right)$ arasındaki farkı inceleme imkanını verir. Böyle herhangi bir c vektörü için, Teorem 2.5 $c\hat{\beta} \sim N\left(c\beta, c\left(Z'Z\right)^{-1}c'\sigma^2\right)$ olduğunu belirtir. Genellikle σ^2 yi bilmeyiz fakat onu (2.89) dan s^2 vasıtasıyla tahmin edebiliriz. Daha sonraki kısımlarda $c\beta$ nın özel bir $c\beta_0$ değerini alıp almadığını test edeceğiz ve Teorem 2.7 yi kullanarak $c\beta$ için güven aralıklarını oluşturacağız.

Teorem 2.7. Z $n\times p$ tipinde bir matris, $Z'Z$ tersinir ve $\sigma > 0$ olmak üzere, $Y \sim N\left(Z\beta,\sigma^2 I\right)$ olduğunu farz edelim $\hat{\beta} = (Z'Z)^{-1}Z'Y$ ve

$s^2 = \frac{1}{n-p}\left(Y - Z\hat{\beta}\right)'\left(Y - Z\hat{\beta}\right)$ olsun. Bu takdirde sıfırdan farklı herhangi bir c $1\times p$ vektörü için,

$$\frac{c\hat{\beta} - c\beta}{s\sqrt{c\left(Z'Z\right)^{-1}c'}} \sim t_{(n-p)} \tag{2.91}$$

dir.

İspat. Bunun için

$$U = \left(c\hat{\beta} - c\beta\right) \Big/ \left(\sigma\sqrt{c\left(Z'Z\right)^{-1}c'}\right)$$

olsun. Bu takdirde, $U \sim N\left(0,1\right)$ dir. $V = s^2/\sigma^2$, bu nedenle $V \sim \chi^2_{(n-p)}\big/\left(n-p\right)$ olsun. Böylece U rasgele değişkeni $\hat{\beta}$ nın bir fonksiyonudur ve V rasgele değişkeni $\hat{\varepsilon}$ nın bir fonksiyonudur. Bu nedenle, U ve V bağımsızdır. Sonuç olarak, t dağılımının tanımından,

$$\frac{c\hat{\beta} - c\beta}{s\sqrt{c\left(Z'Z\right)^{-1}c'}} = \frac{\dfrac{\left(c\hat{\beta} - c\beta\right)}{\sqrt{c\left(Z'Z\right)^{-1}c'\sigma}}}{s/\sigma} = \frac{U}{\sqrt{V}} \sim t_{(n-p)}$$

yazılabilir ve böylece ispat tamamlanır.

Teorem 2.7 t testleri için bir temel niteliğindedir ve lineer modeller için güven aralıkları ile ilgilidir. Formül, $c\hat{\beta}$ tahmininden varsayılan $c\beta$ parametre değerini çıkarmak ve sonra da $c\hat{\beta}$ nın standart sapmasının bir tahminiyle bölmektir. Bu adımları geçtikten sonra bir $t_{(n-p)}$ dağılımlı nicelikle kalırız. Bu adımları çeşitli lineer model kurgularında tekrar tekrar kullanacağız. Kimi zaman Teorem 2.7 yeteri kadar güçlü değildir. β nın bir C $r\times p$ matrisinde dışa vuran r sayıda farklı lineer kombinasyonuna sahip olabiliriz ve $C\beta = C\beta_0$ olup olmadığını test etmeyi isteriz. C nin r sayıda satırını ayrı ayrı test edebiliriz, fakat onların tümünü aynı anda test etmek oldukça farklıdır ve bazı problemler böyle bir eş anlı testi gerektirebilir. Aşağıda verilen Teorem 2.8 Teorem 2.7 nin bir r boyutlu genelleştirmesidir.

Teorem 2.8. Z $n\times p$ tipinde bir matris, $Z'Z$ tersinir ve $\sigma>0$ olmak üzere, $Y \sim N\left(Z\beta,\sigma^2 I\right)$ olduğunu farz edelim ve $\hat{\beta}=\left(Z'Z\right)^{-1}Z'Y$ olsun. Bu takdirde lineer bağımsız satırlara sahip $r\times p$ tipinde herhangi bir C matrisi için

$$\frac{1}{r}\left(\hat{\beta}-\beta\right)'C'\left[C\left(Z'Z\right)^{-1}C'\right]^{-1}C\left(\hat{\beta}-\beta\right)\Big/s^2 \sim F_{r,n-p} \tag{2.92}$$

dir.

İspat. $U=C\left(\hat{\beta}-\beta\right)\big/\sigma$ olsun. Bu takdirde, $\Sigma=C\left(Z'Z\right)^{-1}C'$ tekil olmayan bir $r\times r$ matris olmak üzere, $U\sim N\left(0,\Sigma\right)$ ve bu nedenle, $U'\Sigma^{-1}U\sim\chi^2_{(r)}$ dir. $V=s^2\big/\sigma^2$ ve bu nedenle $V\sim\chi^2_{(n-p)}\big/(n-p)$ olsun. Bu durumda (2.92) nin sol yanı $V/(n-p)\sim\chi^2_{(n-p)}\big/(n-p)$ ile bölünen $(1/r)U'\Sigma^{-1}U\sim\chi^2_{(r)}\big/r$ dir. Öte yandan $\hat{\beta}$ s^2 den bağımsız olduğundan, bu iki χ^2 değişkeni bağımsızdır. Teorem 2.6 yı kullanmak için, bir $H_0: C\beta=C\beta_0$ hipotezini formülleyelim ve (2.92) nin sol yanına $\beta=\beta_0$ koyalım. Eğer sonuç $F^{1-\alpha}_{r,n-p}$ dan daha büyük çıkar ise, bu durumda H_0 hipotezini α düzeyinde reddederiz. Eğer (2.92) de r = 1 koyar ve C yi c olarak yazarsak, bu takdirde ifade

$$\frac{\left(c\hat{\beta}-c\beta\right)^2}{s^2 c\left(Z'Z\right)^{-1}c'}\sim F_{1,n-p}$$

ye sadeleşir. Burada (2.91) bağıntısını daha da iyileştirmek umulabilir. Aslında, sonuçlar güçlü bir şekilde ilişkilidirler. (2.92) nin sol yanındaki istatistik (2.91) in sol yanındakinin karesidir. Dolayısıyla (2.92) nin sağ yanındaki $F_{1,n-p}$ dağılımı (2.91) in sağ yanındaki $t_{(n-p)}$ dağılımının karesidir. Fark sadece (2.92) olmasa da (2.91) in $c\hat{\beta}-c\beta$ nın işaretini açık olarak hesaba katmasıdır.

C matrisini oluşturmak ve bunu Teorem 2.8 in içinde kullanmak biraz gariptir. Uygulamada hipotezleri test etmek için daha basit ve doğrudan bir yöntem kullanılır. β nın bileşenlerinin verilen bir takım alt kümesinin tümünün sıfır olduğu hipotezi test etmek istediğimizi farz edelim. Bu takdirde $\tilde{Z}$; Z nin $q<r$ tanesine sahip ve

$\gamma \in \mathbb{R}^q$ sadece bu q sayıda sütun için katsayıları ihtiva etmek üzere, $E(Y) = \tilde{Z}\gamma$ dır. Tek yapmamız gereken, ilk olarak Z ve sonra da $\tilde{Z}$ üzerinde her iki biçimde regresyon gidişatıdır ve ortaya çıkan hata kareler toplamlarından alt modelin bir testini oluşturabiliriz.

$E(Y) = Z\beta$ ya tam model ve $E(Y) = \tilde{Z}\gamma$ ya alt model diyelim. Bu durumda hata kareler toplamları, sırasıyla SS_{FULL} ve SS_{SUB} olacaktır. Tam model alt modelin yapabileceği herhangi bir modeli çoğaltabildiğinden, her zaman $SS_{FULL} \leq SS_{SUB}$ eşitsizliği sağlanır. Ancak, eğer alt model altında karelerin toplamı çok artar ise, bu takdirde o alt modele karşı bir delildir. Fazladan kareler toplamı $SS_{EXTRA} = SS_{SUB} - SS_{FULL}$ olur. Bu durumda SS_{EXTRA} / SS_{FULL} tesadüfen oluşamayacak kadar büyük olduğunda, alt model reddedilir.

Teorem 2.9. Z $n \times p$ tipinde bir matris, $Z'Z$ tersinir ve $\sigma > 0$ olmak üzere, $Y \sim N(Z\beta, \sigma^2 I)$ olduğunu farz edelim. $q < p$ olmak üzere, $\tilde{Z}$; Z nin bir $n \times q$ alt matrisi olsun. Tam model $\hat{\beta} = (Z'Z)^{-1} Z'Y$ en küçük kareler tahminine ve $SS_{FULL} = (Y - Z\hat{\beta})'(Y - Z\hat{\beta})$ hata kareler toplamına sahip olsun. Benzer şekilde, alt model $\hat{\gamma} = (\tilde{Z}'\tilde{Z})^{-1} \tilde{Z}'Y$ en küçük kareler tahminine ve $SS_{SUB} = (Y - \tilde{Z}\hat{\gamma})'(Y - \tilde{Z}\hat{\gamma})$ hata kareler toplamına sahip olsun. Eğer bir $\gamma \in \mathbb{R}^q$ için alt model $E(Y) = \tilde{Z}\gamma$ gerçeklenir ise, bu takdirde

$$\frac{\frac{1}{p-q}(SS_{SUB} - SS_{FULL})}{\frac{1}{n-p} SS_{FULL}} \sim F_{p-q, n-p} \tag{2.93}$$

dir.

İspat. $H = Z(Z'Z)^{-1} Z'$ ve $\tilde{H} = \tilde{Z}(\tilde{Z}'\tilde{Z})^{-1} \tilde{Z}'$ sırasıyla, tam model ve alt model için şapka matrisleri olsun. Bu takdirde $I - H$ simetrik ve idempotent olduğundan,

$$SS_{FULL} = Y'(I - H)Y \sim \sigma^2 \chi^2_{rank(I-H)}$$

yazılabilir. Öte yandan fazladan kareler toplamı

$$SS_{SUB} - SS_{FULL} = Y'\left(I - \tilde{H}\right)Y - Y'\left(I - H\right)Y = Y'\left(H - \tilde{H}\right)Y$$

olarak yazılır. Ayrıca $H - \tilde{H}$ matrisi simetriktir. Herhangi bir $Y \in \mathbb{R}^n$ için, $H\tilde{H}Y = \tilde{H}Y$ elde edilir. Çünkü, $\tilde{H}Y$; $\tilde{Z}$ nın sütun uzayında, bu nedenle Z nin de sütun uzayındadır. Böylece, $H\tilde{H} = \tilde{H}$ ve transpoze ile $\tilde{H}H = \tilde{H}$ dır. Bu nedenle,

$$\left(H - \tilde{H}\right)\left(H - \tilde{H}\right) = H - H\tilde{H} - \tilde{H}H + \tilde{H} = H - \tilde{H}$$

elde edilir. Sonuç olarak $H - \tilde{H}$ matrisi de idempotenttir. Bu nedenle, $SS_{SUB} - SS_{FULL} \sim \sigma^2 \chi^2_{rank\left(H - \tilde{H}\right)}$ dır. H ve $\tilde{H} - H$ matrisleri ortogonaldir. HY ve $\left(\tilde{H} - H\right)Y$ nin bağımsız oldukları, bu nedenle SS_{FULL} ün $SS_{SUB} - SS_{FULL}$ den bağımsız olduğu görülür. Böylece

$$\frac{\frac{1}{rank\left(H - \tilde{H}\right)}\left(SS_{SUB} - SS_{FULL}\right)}{\frac{1}{rank\left(I - H\right)} SS_{FULL}} \sim F_{rank\left(H - \tilde{H}\right), rank\left(I - H\right)}$$

olduğu gösterilmiş olur. İspatı tamamlamak için, Z tam p rankına sahip olduğunda H nin de p rankına sahip olduğuna dikkat edelim. Aynı zamanda $\tilde{Z}$, q rankına sahiptir ve $\tilde{H}$ da q ranklıdır. $I - H$ ve $H - \tilde{H}$ gibi simetrik ve idempotent matrislerinin rankları izlerine eşittir. Bu durumda, $iz\left(I - H\right) = n - iz\left(H\right) = n - p$ ve $iz\left(H - \tilde{H}\right) = iz\left(H\right) - iz\left(\tilde{H}\right) = p - q$ olur. Böylece ispat tamamlanır.

Burada (2.93) bağıntısına bir geometrik yorum verilebilir. $\varepsilon \sim N\left(0, \sigma^2 I\right)$ olmak üzere, eğer $Y = \tilde{Z}\gamma + \varepsilon$ ise, bu takdirde Y yi $Y_{SUB} + Y_{FULL} + Y_{RES}$ olarak yazabiliriz. Burada Y_{SUB}, $\tilde{Z}$ nın sütunları tarafından gerilen q boyutlu bir uzayda olup, Y_{FULL} vektörü $\tilde{Z}$ ya ortogonal fakat Z nin sütun uzayındaki vektörlerin $p - q$ boyutlu uzayında ve Y_{RES} ise Z ye ortogonal vektörlerin $n - p$ boyutlu uzayındadır. ε gibi bir küresel Gauss vektörü bu uzayların herhangi biri için hiçbir tercih göstermez. Onun her bir uzaydaki izdüşümünün norm karesi ortalaması sadece σ^2 ile uzayın boyutunun çarpımıdır. F testi, Y_{FULL} e karşılık gelen boyutların izdüşüm paylarından daha fazlasını elde edip etmediğini ölçer. Önemli bir özel durumda, tam model bir sabit terimi içerir ve alt model sadece sabit terime sahiptir. Bu takdirde, $SS_{SUB} = \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$ dir. (2.93) bağıntısındaki ANOVA ayrışımı $SS_{SUB} = SS_{FULL} + SS_{RES}$ e sahiptir. Bu durumda F testi R^2 yeteri derecede büyük olduğunda tam modelin lehinde alt modeli reddeder.

2.6.5 Rasgele Açıklayıcılar

Bu kısımda “rasgele X ve rasgele Y “ kurgulamasını göz önüne alacağız. Bir şartlanma argümanı her bir X_i yi sabit nicelik sanki gözlenen x_i değerine eşitmiş gibi işleme sokabildiğimizi söyler ve bu takdirde bu analizin X_i rasgele olduğunda bile bir rasgele X analizinden daha iyi olduğunu hissettirir. Eğer (X_i, Y_i) özdeş bağımsız dağılan çiftler ise, bu takdirde (Z_i, Y_i) ler de aynıdır. Burada $X = x$ verildiğinde Y nin beklenen değerine gereksinim duyacağız. Buna $\mu(x)$ diyelim. Benzer şekilde, $\sigma^2(x) = \text{var}\left(Y \mid X = x\right)$ eşitliğini tanımlayalım. Bu durumda β nın en küçük kareler değeri $E\left(\left(Y_1 - Z_1\beta\right)^2\right)$ yi minimum yapan değer olacaktır. Burada kolaylık olsun diye, $i = 1$ gözlemini seçtik. En küçük kareler türetmesine benzer yöntemleri kullanarak, β yı tanımlayan bir bağıntı olarak, $E\left(Z_1\left(Y_1 - Z_1'\beta\right)\right) = 0$ elde edilir. Başka bir deyişle, normal denklemlerin bir kitle versiyonu elde edilir. $Y_i - Z_i'\beta$ hatası onların çarpımı sıfır beklenen değerine sahip olsun diye Z_i özellik vektörüne ortogonaldir. Bilindiği üzere, $E\left(Z_1 Z_1'\right)$ nün tersinir olduğu durumda

$$\beta = E\left(Z_1 Z_1'\right)^{-1} E\left(Z_1 Y_1\right)$$

elde etmek için, normal denklemleri tekrar düzenleyebiliriz. Bu durumda tekil olmayan $Z'Z$ için, en küçük kareler tahmin edicisi

$$\hat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} Z_i Y_i$$

olarak yazılabilir. Başka bir deyişle, en küçük kareler tahmin edicisi β için pay ve paydanın tahminlerini dikkate alarak elde edilir. Burada $E(\hat{\beta}) = E((Z'Z)^{-1} Z'Y)$ dir. Öte yandan X rasgele olduğunda Z de rasgeledir. Çarpımların beklenen değerleri genel olarak beklenen değerlerin karşılık gelen çarpımları olmadığından, beklenen değer $E\left((Z'Z)^{-1}\right)\times E(Z'Y)$ ye sadeleşmez. Sadeleşse bile, bir beklenen ters, beklenenin tersi olmadığından, $E((Z'Z)^{-1})$, $E(Z_1Z_1')^{-1}/n$ ye sadeleşmez.

Şimdi $X_1,\ldots,X_n$ üzerinde bazı şartlar koşacağız. χ, $X_1,\ldots,X_n$ leri temsil etsin. χ üzerindeki şarta bağlı olarak, Z ile birlikte, X_i ve Z_i sabittir. χ üzerinde şart koşarak ve sonra da sadece şeklen çalışarak, $\mu(X)$ ifadesi $E(Y \mid X = X_i)$ *i*–yinci elemanına sahip rasgele bir vektör olmak üzere,

$$E(\hat{\beta}) = E\left(E(\hat{\beta} \mid \chi)\right) = E\left((Z'Z)^{-1} Z'E(Y \mid \chi)\right) = E\left((Z'Z)^{-1} Z'\mu(X)\right)$$

elde edilir. Şimdiye kadar, $\mu(x)$ fonksiyonunu lineer regresyon modeline bağlamadık. Bu durumu sabit x ve rasgele Y durumu ile karşılaştırılabilir kılmak için, $\mu(x) = z'\beta$ olduğunu farz edelim. Burada $z = \phi(x)$ dir. Ayrıca $\sigma^2(x) = \sigma^2$ nin de sabit olduğunu farz edelim. Bu varsayımların ilkinden

$$E(\hat{\beta}) = E\left((Z'Z)^{-1} Z'Z\beta\right) \tag{2.94}$$

yazılabilir. Bu durumda (2.94) bağıntısı $E(\hat{\beta}) = \beta$ ya oldukça yakındır. Eğer $Z'Z$ nin tekil olma olasılığı 0 ise, bu takdirde istenildiği gibi $E(\hat{\beta}) = \beta$ olacaktır. $Z'Z$ fazlaca yakın fakat pek tekil sayılmazsa bile, (2.94) bağıntısı yine de sorun çıkarmaz. Örneğin, küçük $\varepsilon > 0$ için $Z'Z = \begin{pmatrix} 1 & 0 \\ 0 & \varepsilon \end{pmatrix}$ olsa dahi, bu takdirde $(Z'Z)^{-1} Z'Z\beta = \beta$ olur. En büyük endişe, bazen $Z'Z$ nin tekil olabilmesi veya bir algoritmanın onu tekil olarak işleme sokmasından ziyade sayısal olarak tekile çok yakın olmasıdır. $E(Z_iZ_i')$ tekil olmasa dahi $Z'Z$ tekil olabilir. Örneğin, $1/2$ olasılıkla $X = 1$ ve $1/2$ olasılıkla $X = 0$ olduğunu ve regresyonun $\phi(X) = (1, X)'$ özelliklerine sahip olduğunu farz edelim. Bu takdirde 2^{-n} olasılıkla her $x_i = 1$ dir ve $Z'Z$ tekildir. x_i lerin tümü sıfır olabilse bile bir tekil $Z'Z$ nin böyle olasılığı 2^{1-n} dir. Eğer X_i ler kesikli ve özdeş bağımsız dağılımlı iseler, $Z'Z$ nin tekil olmasının sıfır olmayan bir olasılığı daima vardır. Bu takdirde $\hat{\beta}$ değeri $1-2^{1-n}$ olasılıkla mevcuttur ve 2^{1-n} olasılıkla mevcut değildir. Bu olsaydı muhtemelen özellik 2 yi bırakacak ve dolaylı ya da açık bir şekilde $\hat{\beta}=(\bar{Y},0)'$ yü kullanacaktık. Bu takdirde *n* ye göre üstel olarak küçük olan bir yana sahip oluruz. Bazen nitelik olarak farklı bir şekilde davranan bir algoritmaya sahip olmak bir sıkıntı olabilir, ancak kendi başına böylesine küçük bir ön yargı (yan) nadiren pratik öneme sahip olacaktır.

Sonuç olarak, lineer model $\mu(x)$ e uygun bir şekilde eşleşir ve $Z'Z$ hiç bir şekilde tekil olmaz ise, bu takdirde rasgele *X* ve rasgele *Y* modeli için yanlılık yoktur ya da ihmal edilebilirdir. Şimdi $Z'Z$ nin tekilliğinin olabilirliğinin ihmal edilebilir olduğunu farz edelim. Bu takdirde, her x için $\operatorname{var}(Y \mid X=x)=\sigma^2$ varsaydığımızı akılda tutarak,

$$\begin{aligned}
\operatorname{var}(\hat{\beta}) &= \operatorname{var}\left(E\left(\hat{\beta} \mid \chi\right)\right)+E\left(\operatorname{var}\left(\hat{\beta} \mid \chi\right)\right) \\
&= E\left(\operatorname{var}\left(\hat{\beta} \mid \chi\right)\right) \\
&= E\left((Z'Z)^{-1} Z'\left(\sigma^2 I\right) Z\left(Z'Z\right)^{-1}\right) \\
&= E\left((Z'Z)^{-1}\right)\sigma^2
\end{aligned} \tag{2.95}$$

yazılabilir. Varyans “sabit *X* rasgele *Y*“ durumunda elde ettiğimiz $(Z'Z)^{-1}\sigma^2$ ifadesine benzerdir. Rasgele X durumundaki varyans $(Z'Z)^{-1}$ in beklenen değerini kullanır, halbuki sabit durumda $(Z'Z)^{-1}$ in gerçek gözlenen değerine sahiptik. Daha büyük bir $Z'Z$ matrisi bize sabit X durumunda β nın daha doğru bir tahminini verir. Sabit *X* kurgulamasında $Z'Z$ kullanılır ve bu durumda onun alabildiği diğer değerlerden daha fazlası hesaba katılır. Bu durumda sahip olduğumuz gerçek doğruluğu hesaba katmak, sahip olabileceğimiz doğruluk seviyelerini hesaba katmaktan daha iyidir.

2.6.6 Rank Kusurlu Özellikler

Bazen $Z'Z$ terse sahip değildir, bu nedenle $\left(Z'Z\right)^{-1}Z'Y$ bir tahmin olarak mevcut değildir. Eğer *Z* nin sütunlarından biri diğerinin bir lineer kombinasyonu ise bu durum ortaya çıkacaktır. Aslına bakılırsa bu bir tekil $Z'Z$ matrisini ortaya koyabilmenin yegane yöntemidir.

Teorem 2.10. Eğer $Z'Z$ tersinir değil ise, bu takdirde *Z* nin bir sütunu diğerlerinin bir lineer kombinasyonudur.

İspat. Eğer $Z'Z$ tersinir değil ise, bu takdirde sıfırdan farklı herhangi bir $v \in \mathbb{R}^p$ için $Z'Zv = 0$ dır. Bu durumda $v'Z'Zv = 0$ olacaktır. Ancak $v'Z'Zv = \left\|Zv\right\|^2$ olup bu nedenle $Zv = 0$ dır. Şimdi v_j; *v* nin sıfır olmayan bir elemanı olsun. Bu takdirde $Z_{\cdot j} = \sum_{k \neq j} Z_{\cdot k} v_k / v_j$ dir.

Özelliklerimizden birini diğerinin bir lineer kombinasyonu olarak bulabildiğimiz birden fazla yol vardır. Fahrenheit derecelerindeki sıcaklığı Celcius olarak ölçtüğümüz bir durumda $C = 1.8F + 32$ olduğunu hatırlayarak, eğer alışıldığı gibi model bir sabit terimi içerir ise *Z* de lineer bağımlılık bulacağız. Benzer şekilde, eğer bir finansal model her bölgenin satış dışı özelliklerini ve tüm bölgeler üzerindeki toplam satışları da ortaya koyar ise, bu takdirde *Z* içindeki "toplam" sütunu bölgeseller toplamı olacaktır. Sonunda *Z* nin bir satırı 1_M ve 1_F sırasıyla erkek ve kadın deneklerin gösterge fonksiyonları olmak üzere ve noktaları ilave özellikleri göstermek üzere, $\left(1, 1_M 1_F, \ldots\right)'$ biçimini alabilir.

Bir tekil *Z* matrisi ekseriyetle regresyonu düzenlemede bir yanlış yaptığımızın bir işaretidir. O durumda $Z'Z$ tersinir oluncaya kadar gereksiz özellikleri uzaklaştırmak ve ondan sonra tersinir durumu için yöntemleri ve sonuçları kullanmak suretiyle onu onarabiliriz. Geri kalan kısım, aslına bakılırsa bir tekil düzenleme ile çalışmayı tercih edebildiğimiz kurgulamaları tanımlar. Bir gereksiz düzenekleme ile çalışmak için bir neden, simetridir. Bir–yönlü varyans analizi ile ilgili bölümdeki örneklerde bunu göreceğiz. Başka bir neden ise bir organizmadaki her gen için bir, bir şirketin her müşterisi için bir veya bir elektronik araçta saniyede bir diyeceğimiz çok sayıda regresyonu yapmayı planlayabilmemizdir. Bu regresyonların hepsi aynı değişkenler üzerinde olabilir, ancak bazısı onlarca verilen veri kümelerinin bir sonucu olarak tekil olabilir. Tekillik için sorumlu olan daima aynı değişken olmayabilir. Aynı zamanda, özellik ne olursa olsun, eğer $p > n$ ise, tersinir olmayan bir $Z'Z$ matrisine sahip olmak kaçınılmazdır. Fazladan özellikleri uzaklaştırmak için her bir regresyona ayrı ayrı ince ayar yapılamaz, bu takdirde regresyoları lineer bağımlı veya lineer bağımsız sütunlarla yapmak için iyi bir yönteme gereksinim vardır.

Tekil kurgulamada $\mathcal{M} = \left\{ Z\beta \mid \beta \in \mathbb{R}^p \right\}$, $r < p$ olmak üzere, $\mathbb{R}^n$ nin *r*–boyutlu bir alt kümesidir. Bu durumda *p* boyutlu bir β vektörüyle indekslenen *r*–boyutlu bir alt uzaya sahip oluruz. Bu nedenle, $\mathcal{M}$ deki her bir nokta için β etiketlerinin (isimlendirmelerinin) bir *p–r* boyutlu kümesi vardır. Örneğin, $Z_{i1} = 1$ ve $Z_{i2} = F_i$ ve $Z_{i3} = C_i$ nin $i = 1, \ldots, n$ veri noktalarının Fahrenhait ve Celcius derecelerindeki sıcaklıklar olduklarını farz edelim. Bu takdirde herhangi bir $t \in \mathbb{R}$ için

$$\beta_0 + \beta_1 F_i + \beta_2 C_i = \left(\beta_0 - 32t\right) + \left(\beta_1 - 1{,}8t\right) F_i + \left(\beta_2 + t\right) C_i$$

dir. Eğer Y_i değeri $\beta_0 + \beta_1 F_i + \beta_2 C_i$ ye yakın ise, bu takdirde *t* nin herhangi bir değeri için, $(\beta_0 - 32t) + (\beta_1 - 1.8t) F_i + (\beta_2 + t) C_i$ ye aynı ölçüde yakındır. Problem çekindiğimiz gibi, eksik ranklı Z matrisi ile değil, $Z'Z\hat{\beta} = Z'y$ denklemine çözümün var olmaması ile ilgilidir. Bu durumda çözümlerin sonsuz bir ailesi vardır ve bu aile $\mathbb{R}^p$ nin bir lineer alt uzayıdır. Böyle bir durumda makul bir yaklaşım sadece en kısa çözüm vektörünü seçmektir. Yani, $Z'Z\beta = Z'y$ kısıtlaması altında $\beta'\beta$ nın bir tek minimumlaştırıcısını bulur ve ona $\hat{\beta}$ deriz.

$Z'Z$ tekil olduğunda, $\hat{\beta}$ nın kendisi olmamakla birlikte bazı $c\hat{\beta}$ lineer kombinasyonlarının $\left(Z'Z\right)\hat{\beta} = Z'y$ normal denklemleriyle bir tek şekilde belirlendiği bulunabilir. Eğer β ile ilgili incelememizi $c\beta$ nın bütün lineer kombinasyonları ile sınırlandırabilirsek, bu takdirde $Z'Z$ nin tekilliği o kadar zararlı olmayacaktır.

Geometrik olarak, $\mathcal{M}$ de y ye en yakın bir tek noktanın var olması gerektiği açıktır. Bu nedenle, $\hat{y} = Z\hat{\beta}$ en küçük kareler şartlarıyla belirlenir. Böylece her bir $\hat{y}_i = z_i'\hat{\beta}$ bileşeninin de belirlendiği görülür. Diğer bir deyişle, eğer örneklenmiş özellik vektörlerimizden biri için $c = z_i'$ belirli ise, bu takdirde $c\hat{\beta}$ da belirlenebilir. c değeri z_i' nün bir lineer kombinasyonu, yani Z tasarım matrisinin satırlarının bir lineer kombinasyonu ise, bu takdirde $c\hat{\beta}$ tanımlıdır. Böyle c yi $\gamma \in \mathbb{R}^n$ için $c = \sum_{i=1}^{n} \gamma_i z_i' = \gamma' Z$ olarak yazabiliriz. Aslında sadece tahmin edilebilir $c\beta$ bu biçimi alır.

Tanım 2.1. (Tahmin Edilebilirlik): Eğer herhangi bir $\beta \in \mathbb{R}^p$ için $E(\lambda' Y) = c\beta$ nın gerçeklendiği tepkimelerin bir $\lambda' Y = \sum_{i=1}^{n} \lambda_i Y_i$ lineer kombinasyonu mevcut ise $c\beta$ lineer kombinasyonları tahmin edilebilirdir denir. Tahmin edilebilirliğin bu tanımı Y_i nin $c\beta$ için yansız olan bir lineer kombinasyonunu bulabildiğimiz şarta bağlar. Bundan sonraki teorem aynı zamanda tanım olarak da kullanılabilen iki eşdeğer özelliği verir. Birincisi tahmin edilebilir $c\beta$ lineer kombinasyonlarının bir tek en küçük kareler tahmin edicilerine sahip olmalarıdır. Diğeri ise tahmin edilebilir lineer kombinasyonların Z tasarım matrisinin satırlarının bir lineer kombinasyonu olmalarıdır. Eğer gerçekten tahmin etmek istediğimiz herhangi bir $c\beta$ var ise, veri toplamak için herhangi x_i yi seçiyor olduğumuzda onun tahmin edilebilir olduğundan emin olabiliriz.

Teorem 2.11. Aşağıdaki ifadeler eşdeğerdir.

1. $c\beta$ tahmin edilebilirdir.
2. Herhangi bir $\gamma \in \mathbb{R}^n$ için, $c = \gamma' Z$ dir.
3. Eğer $Z'Z\hat{\beta} = Z'Y$ ve $Z'Z\tilde{\beta} = Z'Y$ ise, bu takdirde $c\hat{\beta} = c\tilde{\beta}$ olacaktır.

2.6.7 En Çok Olabilirlik ve Gauss Markov Teoremi

Bu kısımda en küçük kareleri seçmek için motivasyonlar veya gerekçeler gözden geçirilecektir. İlki en çok olabilirlik vasıtasıyladır. Eğer hataların normal dağıldıkları kabul edilirse, bu takdirde en küçük kareler β regresyon katsayısının en çok olabilirlik tahmin edicisini ortaya koyar. En küçük kareleri kullanmak hataları normal kabul etmekle aynı değildir. En küçük kareleri kullanmaya sevk edebilen başka varsayımlar da vardır. ε hata vektörünün sıfır ortalama ve $\sigma^2 I$ varyans matrisine sahip olduğu daha güçsüz varsayımı altında, en küçük kareler tahmin edicileri lineer yansız tahmin edicilerin sınıfı arasında en küçük varyansa sahiptir. Bu optimallik Gauss Markov teoremi vasıtasıyla hatasız olarak yapılır.

$\hat{\beta}$ en küçük kareler tahminleri, bir normal dağılım modeli altında β nın en çok olabilirlik tahminleri olarak da elde edilebilir. Yani, β nın, gözlenen verinin olasılığını maksimum yapan değerini arayacağız ve onun en küçük kareler ile bulduğumuz ve izdüşümler vasıtasıyla incelediğimiz $\hat{\beta}$ nın aynısı olduğunu bulacağız. σ^2 nin en çok olabilirlik tahmin edicisi (MLE)

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - z_i'\hat{\beta}\right)^2 \tag{2.96}$$

dir. Bu MLE yi elde etmek için, sabit ve tam ranklı Z için $Y \sim N\left(Z\beta, \sigma^2 I\right)$ olduğunu farz edelim. $i = 1,\ldots,n$ için $Y_i = y_i$ yi gözlemenin olasılığı şüphesiz sıfırdır ki bu durum yararlı değildir. Ancak gözlemlerimiz sadece sonlu duyarlılıkta ölçülmüş ve $\Delta > 0$ ölçüm doğruluğu olmak üzere, y_i gözlemini $y_i - \Delta \le Y_i \le y_i + \Delta$ olayına karşılık geliyor gibi yorumlayabiliriz. Aslında Δ yı bilemeyiz, fakat bu değer σ ya kıyasla küçük olmalıdır. y etrafındaki küçük bir n boyutlu kutuda Y yi gözlemlemenin olasılığı

$$\Pr\left(\left|Y_i - y_i\right| \le \Delta,\ 1 \le i \le n\right) = \frac{\left(2\Delta\right)^n}{\left(2\pi\right)^{n/2}\sigma^n}\exp\left(-\frac{1}{2\sigma^2}\left(y - Z\beta\right)'\left(y - Z\beta\right)\right) \tag{2.97}$$

dır. Herhangi bir $0 < \sigma < \infty$ için, (2.97) eşitliğini maksimum yapan β aynı zamanda $\left(y - Z\beta\right)'\left(y - Z\beta\right)$ kareler toplamını minimum yapan β dır. Bu nedenle MLE en küçük kareler tahminidir ve normal denklemleri çözerek bulunabilir. Şimdi σ nın MLE sini bulalım. (2.97) eşitliğinin sağ yanının logaritması

$$-n\log\Delta - \frac{n}{2}\log\left(2\pi\right) - n\log\sigma - \frac{1}{2\sigma^2}\left(y - Z\beta\right)'\left(y - Z\beta\right) \tag{2.98}$$

dır. (2.98) ifadesini σ ya göre diferansiyelleyerek

$$-\frac{n}{\sigma}+\frac{1}{2\sigma^3}(y-Z\beta)'(y-Z\beta) \tag{2.99}$$

bağıntısı elde edilir ki bunun sıfır olması için gerek ve yeter şart

$$\sigma^2=\frac{1}{n}(y-Z\beta)'(y-Z\beta) \tag{2.100}$$

olmasıdır. (2.100) denklemini $\beta=\hat{\beta}$ için çözmek (2.96) dan $\hat{\sigma}^2$ nin MLE sini verir. Aynı zamanda log–olabilirliğin $\sigma=\hat{\sigma}$ noktasında ikinci türevini de incelemek gerekir. Bu değer $-2n/\sigma^2<0$ dır ve bu nedenle (2.96) gerçekten bir MLE dir ve bir minimum değildir. Kesin olarak $y=Z\hat{\beta}$ olduğu durumda log olabilirlik dejeneredir. Bu durumda $\hat{\beta}$ ve $\hat{\sigma}$ için MLE formülleri hala yorumlanabilirdir. Bunlardan ilki hatayı sıfır kılar ve ikincisi ise 0 a indirgenir.

En küçük kareler için ikinci gerçekleme Gauss–Markov teoremi vasıtasıyladır. Eğer her β için $E(\lambda'Y)=c\beta$ gerçeklenir ise, $\lambda'Y$ ye $c\beta$ için bir lineer yansız tahmin edicidir denir. $c\beta$ nın en iyi lineer yansız tahmin edicisini minimum varyansa erişen biri olarak tanımlayalım. Onun bir tek olması gerekmez. Gauss–Markov teoremi en küçük karelerin, en iyi lineer yansız tahmin edicileri verdiğini ifade eder.

Teorem 2.12. (Gauss – Markov Teoremi) Z rasgele olmayan bir $n\times p$ matris, β $\mathbb{R}^p$ de bir bilinmeyen nokta ve ε, 0 ortalamalı ve $\sigma^2 I_n$ varyanslı bir rasgele vektör olmak üzere, $Y=Z\beta+\varepsilon$ olsun. $c\beta$ tahmin edilebilir ve $\hat{\beta}$ bir en küçük kareler tahmini olsun. Bu takdirde $c\hat{\beta}$ ifadesi $c\beta$ nın bir en iyi lineer yansız tahminidir.

İspat. Teoremi $Z'Z$ nin tersinir olduğu tam rank durumu için ispatlayalım. $c\hat{\beta}$ nın yansız ve lineer olduğunu önceden biliyoruz. $E(\lambda'Y)=c\beta$ ve bu nedenle $\lambda'Z=c$ olduğunu farz edelim. Bu takdirde

$$\begin{aligned}\operatorname{var}(\lambda'Y)&=\operatorname{var}\left(c\hat{\beta}+\left(\lambda'Y-c\hat{\beta}\right)\right)\\&=\operatorname{var}\left(c\hat{\beta}\right)+\operatorname{var}\left(\lambda'Y-c\hat{\beta}\right)+2kov\left(c\hat{\beta},\lambda'Y-c\hat{\beta}\right)\end{aligned}$$

yazılabilir. Öte yandan $Z'\lambda=c'$ olduğundan,

$$
\begin{aligned}
kov\left(c\hat{\beta},\lambda'Y-c\hat{\beta}\right) &= kov\left(c\left(Z'Z\right)^{-1}Z'Y,\left(\lambda'-c\left(Z'Z\right)^{-1}Z'\right)Y\right)\\
&= c\left(Z'Z\right)^{-1}Z'\left(\sigma^2 I\right)\left(\lambda-Z\left(Z'Z\right)^{-1}c'\right)\\
&= \sigma^2\left(c\left(Z'Z\right)^{-1}Z'\lambda-c\left(Z'Z\right)^{-1}c'\right)\\
&= 0
\end{aligned}
$$

olacaktır. Bu nedenle, $\mathrm{var}(\lambda'Y)=\mathrm{var}(c\hat{\beta})+\mathrm{var}(\lambda'Y-c\hat{\beta})\geq \mathrm{var}(c\hat{\beta})$ olduğu görülür ki bu da teoremin ispatını tamamlar.

Eksik ranklı Z matrisi için, yani tam sütun ranklı olmayan Z için de benzer bir ispat yapılabilir, ancak biraz daha fazla dikkat gerektirir. Sezgisel olarak, rank eksikliği durumunda tahmin edilebilir fonksiyonlar için Gauss–Markov teoreminin direnmesini bekleyebiliriz. Böyle durumlarda $Z'Z$ tam ranklı oluncaya dek Z nin sütunlarını düşürebilir, sonuçta ortaya çıkan tam rank probleminde bir en iyi lineer yansız tahmin edici elde edebilir ve sonra da sadece gereksiz sütunları düşürdüğümüzden dolayı düşürdüklerimizi eski yerine geri getirebiliriz.

Tanım 2.2. Eğer $rank\left(A\right)$ A matrisini sütunların sayısına eşit ise, A matrisi tam sütun ranklıdır denir.

Sonuç 2.3. $\mathcal{C}\left(A'A\right)=\mathcal{C}\left(A'\right)$ dir. Bu durumda A nın tam sütun ranklı olması için gerek ve yeter şart $A'A$ iç çarpımının tersinir olmasıdır.

İspat. Önce $\mathcal{C}\left(A'A\right)\subseteq\mathcal{C}\left(A'\right)$ olduğunu gösterelim. Eğer, $v\in\mathcal{C}\left(A'A\right)$ ise, bu takdirde herhangi bir w için $v=\left(A'A\right)w$ olacaktır. $x=Aw$ alınarak $v=A'x$ olduğu görülür ki bu da $v\in\mathcal{C}\left(A'\right)$ olduğunu ifade eder. Şimdi de $\mathcal{C}\left(A'\right)\subseteq\mathcal{C}\left(A'A\right)$ olduğunu gösterelim. Bunun için bir $v\in\mathcal{C}\left(A'\right)$ seçelim. Bu takdirde herhangi bir w vektörü için $v=Aw$ yazılabilir. Şimdi $w'\in K\left(A'\right)$ ve $w''\in K\left(A'\right)^{\perp}=\mathcal{C}\left(A\right)$ olmak üzere, w yı $w=w'+w''$ olarak ayıralım. w', A nın sıfır uzayında olduğundan $v=A'w''$ yazılır. Fakat herhangi bir x vektörü için $w''=Ax$ olacağından $v=A'Ax\Rightarrow v\in\mathcal{C}\left(A'A\right)$ dır.

Şimdi normal denklemlere geri dönelim ve $\hat{\beta}_{LS}$ en küçük kareler tahminlerinin $X'X\hat{\beta}_{LS}=X'Y$ denklemini sağladığını hatırlayalım. $\mathcal{C}\left(X'X\right)=\mathcal{C}\left(X'\right)$

olduğundan, $\hat{\beta}_{LS}$ daima mevcut olmalıdır. Bu durumda aşağıdaki durumlar söz konusudur.

1. Eğer $X'X$ tersinir ise (yani, X tam ranklı ise), bu takdirde $\hat{\beta}_{LS} = \left(X'X\right)^{-1} X'Y$ bir tek en küçük kareler tahminidir.

2. Eğer X tam ranklı değil ise, sonsuz sayıda çözüm vardır. Örneğin, eğer $\hat{\beta}_{LS}$ bir en küçük kareler tahmini ve *k*, X in sıfır uzayında ise (*k* yı sıfırdan farklı olarak seçebiliriz), bu takdirde $\hat{\beta}'_{LS} = \hat{\beta}_{LS} + k$ başka bir en küçük kareler tahminidir, çünkü bu da normal denklemleri sağlar.

2.6.8 Lineer Modellerde Belirlenebilirlik

Belirlenebilirlik, istatistiksel modellerde parametrelerin bir özelliğidir. Aşikar olarak, veriden hangi parametrelerin tahmin edilebilir olduğunu ve olmadığını bilmek için, belirlenebilirlik yararlıdır. Belirlenebilirlik, bu belirlemeyi kılmaya yardım eder.

Amaç, parametreler ve istatistiksel modeller arasındaki eşleştirmeyi incelemektir. Böyle bir eşleştirmenin birebir olup olmadığını bilmek isteriz. Bunun nedenini anlamak için, θ nın bir parametre ve $P\left(x,\theta\right)$ nın bir istatistiksel model (bir olasılık dağılımı) olduğunu farz edelim. Aşikar olarak eğer $\theta_1 = \theta_2$ ise, bu takdirde her x için, $P\left(x,\theta_1\right) = P\left(x,\theta_2\right)$olacaktır. Bununla beraber, eğer birkaç $\theta_1 \neq \theta_2$ mevcut ise, fakat $P\left(x,\theta_1\right) = P\left(x,\theta_2\right)$ ise, bu takdirde $P\left(\cdot,\theta_1\right)$ e göre dağılan veri $P\left(\cdot,\theta_2\right)$ ye göre dağılan veri ile aynı dağılıma sahip olacaktır. Bu durumda, gerçek dağılımın parametresinin θ_1 veya θ_2 olup olmadığını veriden belirlemek mümkün olmaz. Bu takdirde, böyle modellerde, θ parametresini tahmin etmek daima mümkün değildir.

Lineer modellerde β nın belirlenebilir olması için, $\mu\left(\beta\right) = E\left(Y\right) = X\beta$ ortalama parametresinin β nın bir birebir fonksiyonu olması yeterlidir. Herhangi bir simetrik pozitif–yarı tanımlı Σ (β ya bağlı olmayan) matrisi için $kov\left(\varepsilon\right) = \Sigma$ olduğu sürece bu doğrudur.

Tanım 2.3. (Lineer modellerde β nın belirlenebilirliği): Eğer, $\mu(\beta_1)=\mu(\beta_2)$ olacak şekilde $\forall\beta_1,\beta_2$ için bu eşitlik $\beta_1=\beta_2$ olduğunda doğru ise, β parametresi belirlenebilirdir. Buna göre eğer X matrisi tam sütun ranklı ise, β belirlenebilirdir.

Eğer β_1 ve β_2 parametreleri için $X\beta_1=X\beta_2$ eşitliği sağlanırsa, bu takdirde $X(\beta_1-\beta_2)=0$ olacaktır ve bu nedenle, bu $\beta_1-\beta_2\in K(X)$ olması demektir. Ancak X in tam ranklı olması $K(X)=\{0\}$ olduğunu ifade eder. Bu nedenle, $\beta_1-\beta_2=0$, yani $\beta_1=\beta_2$ dir. Eğer X tam ranklı değil ise, bu takdirde β belirlenebilir değildir ve $k\neq 0$ olmak üzere $k\in K(X)$ mevcuttur. Bir β_1 parametresini seçelim ve $\beta_2=\beta_1+k$ alalım. Bu takdirde, $X\beta_2=X(\beta_1+k)=X\beta_1+Xk=X\beta_1$ dir fakat $\beta_1\neq\beta_2$ dir. Öte yandan β nın kendisi belirlenebilir olmayabilir iken, β nın belirlenebilir olan fonksiyonları mevcut olabilir.

Tanım 2.4. (β nın belirlenebilir fonksiyonları): Eğer $\forall\beta_1,\beta_2$ için $\mu(\beta_1)=\mu(\beta_2)$ olduğunda $g(\beta_1)=g(\beta_2)$ olacak şekilde $g(\beta)$ mevcut ise, bu takdirde $g(\beta)$ ya β nın bir belirlenebilir fonksiyonu denir.

Teorem 2.13. (β nın belirlenebilir fonksiyonlarının tanımlanması): $g(\beta)$ nın β nın bir belirlenebilir fonksiyonu olması için gerek ve yeter şart g nin sadece $\mu(\beta)$ vasıtasıyla β ya bağlı olmasıdır, yani,

$$g(\beta)=g_*(\mu(\beta))=g_*(X\beta)$$

olacak şekilde bir g_* fonksiyonunun var olmasıdır.

Şimdi belirlenebilirlik ve en küçük kareler tahmini arasındaki ilişkiyi inceleyelim. Eğer X tam sütun ranklı ise, yani $X'X$ tersinir ise ve $\hat{\beta}_{LS}=(X'X)^{-1}X'Y$ belirlenebilir β parametresinin yegane en küçük kareler tahmini ise bu takdirde herhangi bir g fonksiyonu için $g(\hat{\beta}_{LS})$ ifadesi $g(\beta)$ nın bir tek en küçük kareler tahminidir.

Eğer X tam sütun ranklı değil ise, bir tek en küçük kareler tahmini olmayacaktır ve biz bu durumda β yı tahmin etmek için hangisini kullanmalıyız? Ayrıca, β parametresi belirlenebilir değildir ve bu nedenle $\hat{\beta}_{LS}$ en küçük kareler çözümünün hangi parametrenin tahmini olacağı bile açık değildir. Bununla beraber, eğer $\hat{\beta}_{LS}^{(1)}$ ve $\hat{\beta}_{LS}^{(2)}$ iki en küçük kareler tahmini ise, bu takdirde $X\hat{\beta}_{LS}^{(1)} = X\hat{\beta}_{LS}^{(2)}$ dir. Bu nedenle, β nın belirlenebilir fonksiyonları bir tek en küçük kareler çözümlerine sahiptir. Çünkü, eğer $g(\beta) = g_*(X\beta)$ ise, bu takdirde herhangi bir $\hat{\beta}_{LS}$, en küçük kareler tahmini için $g(\hat{\beta}_{LS}) = g_*(X\hat{\beta}_{LS})$ fonksiyonu $g(\beta)$ nın yegane en küçük kareler tahminidir.

Öte yandan β nın lineer fonksiyonları için, yani, λ ağırlıklarının bir kümesi için, $g(\beta) = \lambda'\beta$ fonksiyonları için bir karşıt vardır. Eğer $\hat{\beta}_{LS}^{(1)}$ ve $\hat{\beta}_{LS}^{(2)}$; $\lambda'\hat{\beta}_{LS}^{(1)} = \lambda'\hat{\beta}_{LS}^{(2)}$ yi sağlayan iki en küçük kareler tahmini ise, bu takdirde ρ ağırlıklarının herhangi bir kümesi için $\lambda = X'\rho$ olacaktır. Yani, $g(\beta) = \lambda'\beta = \rho' X\beta = g_*(X\beta)$ yazılabilir ve bu nedenle g ; β nın bir belirlenebilir lineer fonksiyonudur.

3. BULGULAR ve TARTIŞMA

3.1 Kısıtlanmış Lineer Modelde Tahminin Geometrisi ve Hipotez Test Etme: Tam Rank Durumu

Lineer modellere geometrik yaklaşımda, dikkatler gözlem uzayı üzerine odaklanır. Burada tahmin ve hipotez test etme herkesçe bilinen vektör uzayı tekniklerini kullanarak yerine getirilebilir. Eğer Y vektörü Ω da uzandığı bilinen $E(Y)$ ortalama vektörüne sahip olan gözlemlerin bir $n \times 1$ vektörü ise, bu takdirde ortalama vektörü; $\hat{Y} = P_{\Omega} Y$ ile tahmin edilir. Burada Ω, n – boyutlu E^n Öklid uzayının q – boyutlu bir alt uzayı ve P_{Ω} da E^n nin Ω üzerine ortogonal izdüşümüdür. Sıfır hipotezleri, ω Ω nın r – boyutlu bir alt uzayı olmak üzere, $E(Y) \in \omega$ olarak ifade edilir. γ nın, ω nın Ω daki ortogonal tümleyeni olmasına izin vererek, $\omega \oplus \gamma = \Omega$ yazarsak (yani, $v \in \gamma$ ve $w \in \omega$ için, $v'\omega = 0$ olacak şekilde, γ; Ω nın bir $q-r$ boyutlu alt uzayıdır) bu durumda

$$\left\|P_{\gamma}Y\right\|^2 \div (q-r) \Big/ \left\|P_{\Omega^{\perp}}Y\right\|^2 \div (n-q) = Y'P_{\gamma}Y \div (q-r) / Y'P_{\Omega^{\perp}}Y \div (n-q)$$

oranı $E(Y) \in \omega$ sıfır hipotezini test için alışılmış orandır. Bununla beraber, eğer η γ nın herhangi bir alt uzayı ve δ; $\Omega^{\perp}$ nin herhangi bir alt uzayı ise,

$$\left(\left\|P_{\eta}Y\right\|^2 \div b\right) \Big/ \left(\left\|P_{\delta}Y\right\|^2 \div d\right)$$

oranı da sıfır hipotezi doğru olduğunda bir merkezi F dağılımına sahiptir (burada b ve d sırasıyla η ve δ nın boyutlarıdır). Burada $v \in E^n$ için, $\|v\|^2 = v'v$ dir.

Klasik yaklaşım (ki o tamamıyla cebirseldir) $E(Y) = X\beta$ parametrelenmiş modeli ile başlar. Bir ANOVA modelinde, β pekala hücre ortalamalarının vektörü olabilir. Sıfır hipotezleri, $H_0 : H'\beta = 0$ olarak parametre vektörüne göre ifade edilir. β nın $\hat{\beta}$ tahmini β cinsinden bir kuadratik formu minimumlaştırarak hesaplanır ve $E(Y)$ ifadesi $\hat{Y} = X\hat{\beta}$ ile tahmin edilir. Sıfır hipotezini test etmek için pay kareler toplamı $\hat{\beta}' H \left\{H'(X'X)^{-1} H\right\}^{-1} H'\hat{\beta}$ ile verilir.

Kısıtlanmış modelde, parametre vektörünün $C'\beta = 0$ şartını sağlaması istenir. β nın $\hat{\beta}_c$ tahminini hesaplamak için β değişkenli kuadratik form $C'\beta = 0$ kısıtlamasına bağlı olarak minimumlaştırılır. Bu ise Lagrange çarpanları tekniğinin kullanımıyla başarılır. Ortalama vektörü $\hat{Y}_c = X\hat{\beta}_c$ ile tahmin edilir. Kısıtlanmış modelde $H'\beta = 0$ hipotezini test etmek için C ve H matrisleri yeni bir $T = (C, H)$ matrisini oluşturmak için eklenir. Bu test için, pay kareler toplamı

$$SS(C,H) = \hat{\beta}'\left[T\left\{T'(X'X)^{-1}T\right\}^{-1}T' - C\left\{C'(X'X)^{-1}C\right\}^{-1}C'\right]\hat{\beta}$$

(3.1)

olacaktır. Bu durumda sıfır hipotezi doğru olduğunda, kısıtlanmış modelde β nın $\hat{\beta}_{C,H}$ tahmini sanki kısıtlanmış modeldeki β nın tahminiymiş gibi hesaplanır. $T'\beta = 0$ altında $E(Y) = X\beta$ olacaktır. Hesaplama formulü

$$\hat{\beta}_{C,H} = \left[I - (X'X)^{-1}T\left\{T'(X'X)^{-1}T\right\}^{-1}T'\right]\hat{\beta} \tag{3.2}$$

şeklindedir. Böylece $\hat{Y}_{C,H} = X\hat{\beta}_{C,H}$ ifadesi sıfır hipotezi doğru olduğunda kısıtlanmış modelde $E(Y)$ yi tahmin eder. Eğer C ve H sırasıyla $q \times t$ ve $q \times s$ tipinde matrisler ise, bu takdirde T matrisi $q \times (t+s)$ tipinde olup $T'(X'X)^{-1}T$, $(t+s)\times(t+s)$ matrisi (3.1) ve (3.2) de tersinir olmalıdır.

Bu kısımda kısıtlanmış model ile ilgili tüm cebirsel hesaplamaları geometrik olarak ifade edeceğiz. Bunun için önce model ve ilgili hipotez parametrik olarak $C'\beta = 0$ ile kısıtlanan $E(Y) = X\beta$ ve $H_0 : H'\beta = 0$ gibi ifade edilir ve sonra da geometrik analiz için gözlem uzayına dönüştürülür.

Geometrik yaklaşım hem kısıtlanmış hem de kısıtlanmamış modellere uygulanan tahmin ve hipotez test etme için tutarlı bir teoriyi temin eder. Lagrange Çarpanları yöntemi kısıtlanmamış modelde kullanılmaz. Ayrıca, geometrik mülahazalar, (3.1) bağıntısıyla verilen $\hat{\beta}$ cinsinden ikisinin farkından hesapsal olarak daha etkin olan $\hat{\beta}_c$ cinsinden bir tek kuadratik form olarak $SS(C,H)$ için daha doğal bir ifadeye götürür, çünkü bu kuadratik form, $X'X$ hariç, sadece $t \times t$ ve $s \times s$ matrislerinin terslerini içerir.

Tahminin teorisi üç ardışık $\hat{Y}$, $\hat{Y}_C$ ve $\hat{Y}_{C,H}$ tahmininin herhangi birinin uygun bir ortogonal izdüşüm vasıtasıyla herhangi bir önce gelenden elde edilebildiğini ortaya çıkarır. Aynı sonuç, izdüşümlerin ortogonal olmaması hariç olmak üzere, parametre vektörünün tahmini için geçerlidir. Altı tahminin (ortalama ve parametre vektörlerinin) tümünü hesaplamak için etkin bir hesaplama şeması da verilir.

$$\left.\begin{array}{r} C'\beta = m \neq 0 \text{ ile kısıtlanan } E(Y) = X\beta \\ H_0 : H'\beta = h \neq 0 \end{array}\right\} \tag{3.3}$$

durumu zorluk sunmaz. C ve H nin sütunları lineer bağımsız ise, bu takdirde $C'\beta = m$ ve $H'\beta = h$ ortak bir çözüme sahiptir (Gerçekten, ortak çözümlerin kümesi aslında C ve H cinsinden karakterize edilebilen dönüştürülmüş bir alt uzaydır) ve veri dönüştürülebilir, bu nedenle (3.3) ifadesi

$$\left.\begin{array}{r} C'\tilde{\beta} = 0 \text{ ile kısıtlanan } E(\tilde{Y}) = X\tilde{\beta} \\ H_0 : H'\tilde{\beta} = 0 \end{array}\right\} \tag{3.4}$$

olarak yazılabilir.

Bu kısımda, *Y*; $E(Y) = X\beta$ ortalama vektörüne ve $\sigma^2 I$ kovaryans matrisine sahip olan bir $n \times 1$ çok değişkenli normal rasgele vektördür. Burada X tam (sütun) ranklı bir $n \times q$ matris, β bir $q \times 1$ sabit vektör ve *I*, $n \times n$ birim matristir. Ω; E^n nin X in sütunları tarafından gerilen alt uzayını gösterecektir. $r + s < q$ olmak üzere, $q \times t$ tipindeki kısıtlama matrisi *C* ile gösterilecek ve $q \times s$ tipindeki hipotez matrisi ise H ile gösterilecektir. C ve H nin sütunları birlikte E^q da $t + s$ sayıda lineer bağımsız vektörlerin bir kümesini oluşturur. P_A sembolü, A üzerine ortogonal izdüşüm ve $Q_A = I - P_A$ ise $A^\perp$ üzerine ortogonal izdüşümü gösterecektir. Burada $A^\perp$; *A* nın ortogonal tümleyenidir. *A* nın E^n veya E^q nun bir alt uzayı olup olmamasına bağlı olarak *I* matrisi *n* veya *q* boyutlu birim matristir.

3.1.1 Kısıtlanmamış Model

Eğer $E(Y)=X\beta$ ise, bu takdirde $E(Y)$ Ω da uzanır. $E(Y)$ nin GM (Gauss-Markov) tahmini, $P_\Omega = X\left(X'X\right)^{-1}X'$ matrisi E^n nin Ω üzerine ortogonal izdüşümü olmak üzere, $\hat{Y}=P_\Omega Y$ ile verilir. Ortalama vektörü bir kereye mahsus tahmin edilmiş olduğunda, $\beta=\hat{\beta}=\left(X'X\right)^{-1}X'Y$ tahmini $X\beta=\hat{Y}$ ya tek bir çözümü verir. β için tanımlanan denklemlerde $\hat{Y}$ için $P_\Omega Y$ yi yerine koymak daha fazla bilinen $\hat{\beta}=\left(X'X\right)^{-1}X'Y$; sonucunu verir.

$H'\beta=0$ sıfır hipotezini test etmek için, $\gamma_H = X\left(X'X\right)^{-1}H$ olmak üzere, $H'\beta=0$ olması için gerek ve yeter şartın $\gamma_H{}'E(Y)=0$ olması olduğuna dikkat edelim. ω_H, γ_H nin Ω daki ortogonal tümleyeni olsun, bu takdirde $\omega_H \oplus \gamma_H = \Omega$ dır. Bu kısımda γ_H hem $X\left(X'X\right)^{-1}H$ matrisini ve hem de E^n nin $X\left(X'X\right)^{-1}H$ nin sütunları tarafından gerilen alt uzayını göstermek için kullanılmıştır. Eğer ortalama vektörümüz $\gamma_H{}'E(Y)=0$ bağıntısını sağlar ise, bu takdirde $E(Y)$, ω_H üzerinde uzanmalıdır ve bu nedenle sıfır hipotezi $E(Y)\in\omega_H$ olarak ifade edilebilir. $P_{\gamma(H)}$, $\gamma(H)=\gamma_H$ üzerine ortogonal izdüşüm olmak üzere, sıfır hipotezini test etmek için pay kareler toplamı $\left\|P_{\gamma(H)}Y\right\|^2$ dir. Bu aşağıda verilmiştir.

$$\begin{aligned}
\left\|P_{\gamma(H)}Y\right\|^2 &= Y'P_{\gamma(H)}Y \\
&= Y'\gamma_H\left(\gamma_H{}'\gamma_H\right)^{-1}\gamma_H{}'Y \\
&= Y'X\left(X'X\right)^{-1}H\left\{H'\left(X'X\right)^{-1}H\right\}^{-1}H'\left(X'X\right)^{-1}X'Y \\
&= \hat{\beta}'H\left\{H'\left(X'X\right)^{-1}H\right\}^{-1}H'\hat{\beta}
\end{aligned} \tag{3.5}$$

Q_Ω; E^n nin $\Omega^\perp$ üzerine ortogonal izdüşümü olmak üzere, hata kareler toplamı $SS(kısıtlanmamış)=\left\|Q_\Omega Y\right\|^2 = Y'Q_\Omega Y$ ile verilir. $Q_\Omega = I-P_\Omega$ olduğundan,

$$SS(\text{kısıtlanmamış}) = Y'\left\{I - X\left(X'X\right)^{-1}X'\right\}Y$$

bilinen formülü elde edilir.

3.1.2 Kısıtlanmış Model

Paramatre vektörü $C'\beta = 0$ ile kısıtlandığında $E(Y) = X\beta$ ortalama vektörü $\gamma_C = X(X'X)^{-1} C$ ye ortogonal olmalıdır. Çünkü $C'\beta = 0$ olması için gerek ve yeter şart $\gamma_C' E(Y) = 0$ olmasıdır. Ω_C yi γ_C nin Ω daki ortogonal tümleyeni ($\gamma_C \oplus \Omega_C = \Omega$ olacak şekilde) olarak alarak $C'\beta$ kısıtlaması $E(Y)$ ortalama vektörünün Ω_C de uzanması kısıtlamasını getirir. Bu nedenle, $\Omega(C) = \Omega_C$ olmak üzere, $E(Y)$ nin GM tahmini $\hat{Y}_C = P_{\Omega_C} Y$ şeklindedir.

Eğer $C'\beta = 0$ ise, bu takdirde β ; C nin sütunlarına ortogonal olmalıdır, yani, β ; C nin E^q daki ortogonal tümleyeni olan $C^\perp$ de uzanmalıdır. Bu nedenle γ_C ile C ve Ω_C ile $C^\perp$ alt uzaylarının doğal bir eşlemesi vardır. Alt uzay çiftlerini bağlamak için uygun dönüşümler aşağıdaki gibidir.

$$X(X'X)^{-1} : C \rightleftarrows \gamma_C : X'$$

$$X : C^\perp \rightleftarrows \Omega_C : (X'X)^{-1} X'$$

Burdick ve Ark. sırasıyla “en iyi tahmin köprüsü ve “en iyi uygun köprü” olarak $X(X'X)^{-1}$, X' ve X, $(X'X)^{-1} X'$ ifadelerine başvurur. $P_{\Omega(C)}$ yi hesaplamak için $\gamma(C) = \gamma_C$ olmak üzere, $P_{\Omega(C)} = P_\Omega - P_{\gamma(C)}$ yazalım. Bu durumda

$$P_\Omega = X(X'X)^{-1} X'$$

ve

$$\begin{aligned} P_{\gamma(C)} &= \gamma_C (\gamma_C' \gamma_C)^{-1} \gamma_C' \\ &= X(X'X)^{-1} C \left\{ C'(X'X)^{-1} C \right\}^{-1} C'(X'X)^{-1} X' \end{aligned}$$

olduğundan,

$$P_{\Omega(C)} = XA_C (X'X)^{-1} X' \tag{3.6}$$

olacaktır, burada

$$A_C = I - (X'X)^{-1} C \left\{ C'(X'X)^{-1} C \right\}^{-1} C' \tag{3.7}$$

E^q nun $(X'X)^{-1} C$ boyunca $C^{\perp}$ üzerine (ortogonal olmayan) izdüşümüdür. Bundan böyle, (3.7) ifadesi ile tanımlanan A_C sembolü E^q nun $(X'X)^{-1} C$ boyunca $C^{\perp}$ üzerine izdüşümünü gösterecektir. Bir kez $\hat{Y}_C = P_{\Omega(C)} Y = X A_C (X'X)^{-1} X'Y$ hesaplanmış olduğunda, β

$$\hat{\beta}_C = (X'X)^{-1} X'\hat{Y}_C = A_C (X'X)^{-1} X'Y \tag{3.8}$$

ile tahmin edilir. Bu durumda $\beta = \hat{\beta}_C$ nin $X\beta = \hat{Y}_C$ ye bir tek çözüm olduğu ve $\hat{\beta}_C \in C^{\perp}$ olduğu kolayca gösterilebilir.

Ω_C nin, E^n nin $E(Y) \in \Omega_C$ özelliğine sahip olan alt uzayı olması için gerek ve yeter şartın aşağıdaki iki denklemin eş anlı olarak sağlanması gerektiğini hatırlayalım:

$$E(Y) = X\beta \text{ ve } C'\beta = 0$$

İlk işimiz $E(Y) \in \Omega_C$ ve $H'\beta = 0$ olması için gerek ve yeter şartın $E(Y) \in \omega_{C,H}$ olması için, Ω_C nin $\omega_{C,H}$ alt uzayını bulmaktır. Bundan sonra $\omega_{C,H}$ nin Ω_C deki $\Pi_{C,H}$ ortogonal tümleyenini oluşturacağız. Bu durumda kısıtlanmış modelde $H'\beta = 0$ hipotezini test etmek için pay ve payda kareler toplamları, $\Pi(C,H) = \Pi_{C,H}$ olmak üzere, sırasıyla $Y'P_{\Pi(C,H)}Y$ ve $Y'Q_{\Omega(C)}Y$ ile verilir.

Sıfır hipotezi kısıtlanmış modelde doğru olduğunda, $C'\beta = 0$ ve $H'\beta = 0$ dır. Bu iki denklem, $T = (C,H)$ olmak üzere, tek bir $T'\beta = 0$ denklemi olarak yeniden yazılabilir. Ancak, $T'\beta = 0$ olması için gerek ve yeter şart, $\gamma_T = X(X'X)^{-1} T$ olmak üzere, $\gamma_T' E(Y) = 0$ olmasıdır. Bu nedenle, sıfır hipotezi kısıtlanmış modelde doğru olduğunda, $E(Y) \in \omega_{C,H}$ olacaktır. Burada $\omega_{C,H}$ γ_T nin Ω daki ortogonal tümleyenidir, yani $\omega_{C,H} \oplus \gamma_T = \Omega$ dır. Bu nedenle $\Pi_{C,H}$ yi, $\omega_{C,H}$

nin Ω_C deki ortogonal tümleyeni olarak alarak, dolayısıyla $\omega_{C,H} \oplus \Pi_{C,H} = \Omega_C$ alarak,

$$P_{\Pi(C,H)} = P_{\Omega(C)} - P_{\omega(C,H)} \tag{3.9}$$

olduğu anlaşılır. $P_{\Omega(C)}$ ve $P_{\omega(C,H)}$ için (3.9) da sırasıyla $P_\Omega - P_{\gamma(C)}$ ve $P_\Omega - P_{\gamma(T)}$ yi yerine koyarak

$$\begin{aligned} P_{\Pi(C,H)} &= P_{\gamma(T)} - P_{\gamma(C)} \\ &= X\left(X'X\right)^{-1}\left[T\left\{T'\left(X'X\right)^{-1}T\right\}^{-1}T' - C\left\{C'\left(X'X\right)^{-1}C\right\}^{-1}C'\right]\left(X'X\right)^{-1}X' \end{aligned}$$

olduğu görülür.

$Y'P_{\Pi(C,H)}Y$ içinde yerine koyulduğunda, $P_{\Pi(C,H)}$ için bu formül T artırılmış matrisini ihtiva eder ve (3.1) bağıntısını üretir. Öte yandan $P_{\Pi(C,H)}$ için H ve C cinsinden ayrı ayrı ifade etmek için

$$\gamma_T = \gamma_C + \gamma_H = \gamma_C \oplus Q_{\gamma_{(C)}}\gamma_H \tag{3.10}$$

yazılabilir. Bu durumda (3.10) bağıntısı $\Omega_C = \omega_{C,H} \oplus Q_{\gamma(C)}\gamma_H$ olduğunu gösterir. Bu nedenle,

$$\Pi_{C,H} = Q_{\gamma_{(C)}}\gamma_H = XA_C\left(X'X\right)^{-1}H$$

yazılabilir ve dolayısıyla $A_C\left(X'X\right)^{-1}H$ ve $\Pi_{C,H}$ arasında doğal bir tekabül vardır, yani,

$$X : A_C\left(X'X\right)^{-1}H \rightleftarrows \Pi_{C,H} : \left(X'X\right)^{-1}X'$$

dir, burada her iki dönüşüm birebir ve örten olmaktadır. $A_C\left(X'X\right)^{-1}H$, C ye ortogonal olmasına rağmen T de ihtiva edilmez ve bu nedenle $T^{\perp}$ ye ortogonal değildir. E^n nin $\Pi_{C,H}$ üzerine ortogonal izdüşümü

$$P_{\Pi(C,H)} = Q_{\gamma(C)}\gamma_H\left(\gamma'_H Q_{\gamma(C)}\gamma_H\right)^{-1}\gamma'_H Q_{\gamma(C)} \tag{3.11}$$

ile verilir. Öte yandan

$$\begin{aligned}\gamma_H' Q_{\gamma(C)} &= H'\left(X'X\right)^{-1} X'Q_{\gamma(C)} = H'\left(X'X\right)^{-1} X'X\left(X'X\right)^{-1} X'Q_{\gamma(C)} \\ &= \gamma_H' P_{\Omega} Q_{\gamma(C)} = \gamma_H' P_{\Omega(C)}\end{aligned}$$

olduğundan (3.11) bağıntısı

$$P_{\Pi(C,H)} = P_{\Omega(C)}\gamma_H\left(\gamma_H' P_{\Omega(C)}\gamma_H\right)^{-1}\gamma_H' P_{\Omega(C)} \tag{3.12}$$

$$= XA_C\left(X'X\right)^{-1} H\left\{H'A_C\left(X'X\right)^{-1} H\right\}^{-1} H'A_C\left(X'X\right)^{-1} X' \tag{3.13}$$

olarak yeniden yazılabilir. Bu durumda test için pay kareler toplamı

$$\begin{aligned}Y'P_{\Pi(C,H)}Y &= Y'P_{\Omega(C)}\gamma_H\left(\gamma_H' P_{\Omega(C)}\gamma_H\right)^{-1}\gamma_H' P_{\Omega(C)}Y \\ &= \hat{\beta}_C' H\left\{H'A_C\left(X'X\right)^{-1} H\right\}^{-1} H'\hat{\beta}_C\end{aligned} \tag{3.14}$$

ye eşittir, burada

$$\gamma_H' P_{\Omega(C)}Y = \gamma_H'\hat{Y}_C = H'\left(X'X\right)^{-1} X'\hat{Y}_C = H'\hat{\beta}_C$$

gerçeği kullanılmıştır. Bu nedenle $Y'P_{\Pi(C,H)}Y$, $\hat{\beta}_C$ cinsinden bir kuadratik form olarak yazılabilir, ki onun $H\left\{H'A_C\left(X'X\right)^{-1} H\right\}^{-1} H'$ çekirdeği sadece ($X'X$ hariç) $t\times t$ ve $s\times s$ tipindeki matrislerinin terslerini ihtiva eder. (3.14) ve (3.5) ifadeleri arasındaki benzerliğe dikkat edelim. Bunların her ikisi de model için parametre vektörünün tahminine göre kuadratik formlardır. Bu durumda

$$SSE_{kısıtlanmış} = \left\|Q_{\Omega(C)}Y\right\|^2 = Y'Q_{\Omega(C)}Y$$

yazılabilir. Öte yandan $\Omega\left(C\right)^{\perp} = \gamma_C \oplus \Omega^{\perp}$ olduğundan,

$$\begin{aligned}SSE_{kısıtlanmış} &= \left\|Q_{\Omega(C)}Y\right\|^2 = \left\|P_{\gamma(C)}Y\right\|^2 + \left\|Q_{\Omega}Y\right\|^2 \\ &= Y'P_{\gamma(C)}Y + SSE_{kısıtlanmamış}\end{aligned} \tag{3.15}$$

olacaktır. Burada $Y'P_{\gamma(C)}Y = Y'X\left(X'X\right)^{-1}C\left\{C'\left(X'X\right)^{-1}C\right\}^{-1}C'\left(X'X\right)^{-1}X'Y$ kuadratik formunun $E(Y) = X\beta$ kısıtlanmamış modelinde, $H_0 : C'\beta = 0$ hipotezini test etmek için pay kareler toplamı olması gerektiğine dikkat edelim.

3.1.3 Sıfır Hipotezi Doğru Olduğunda Kısıtlanmış Modelde Tahmin

Bu kısımda, ortalama (gözlem) vektörünün tahmini, Y yi $\omega(C,H)$ üzerine izdüşürerek elde edilecektir. Bunun için $\omega(C,H)\oplus\Pi(C,H)=\Omega(C)$ olduğundan,

$$P_{\omega(C,H)} = P_{\Omega(C)} - P_{\Pi(C,H)} \tag{3.16}$$

yazılabilir. Bu durumda (3.6) ve (3.13) ifadelerini (3.16) da yerine koyarak,

$$P_{\omega}(C,H) = X\left[I - A_C\left(X'X\right)^{-1}H\left\{H'A_C\left(X'X\right)^{-1}H\right\}^{-1}H'\right]A_C\left(X'X\right)^{-1}X'$$

elde edilir. Bu nedenle,

$$\hat{Y}_{C,H} = X\left[I - A_C\left(X'X\right)^{-1}H\left\{H'A_C\left(X'X\right)^{-1}H\right\}^{-1}H'\right]A_C\left(X'X\right)^{-1}X'Y$$

sıfır hipotezi doğru olduğunda kısıtlanmış modelde ortalama vektörünü tahmin eder ve $\hat{\beta}_{C,H} = \left(X'X\right)^{-1}X'\hat{Y}_{C,H}$ parametre vektörünün tahminidir, burada $\beta = \hat{\beta}_{C,H}$ nin $X\beta = \hat{Y}_{C,H}$ nin bir tek çözümü olduğuna ve $\hat{\beta}_{C,H} \in T^{\perp}$ olduğuna dikkat edelim.

3.1.4 Tahminler Arasında İlişkiler

Ortalama vektörünün tahminlerinin her biri daha önce hesaplananlardan elde edilebilir. Örneğin,

$$\begin{aligned}\hat{Y}_C &= P_{\Omega(C)}Y \\ &= Q_{\gamma(C)}P_{\Omega}Y \\ &= Q_{\gamma(C)}\hat{Y}\end{aligned}$$

$$\hat{Y}_{C,H} = Q_{\Pi(C,H)}\hat{Y}_C$$

ve

$$\hat{Y}_{C,H} = Q_{\gamma(C,H)}\hat{Y}$$

olduğu da gösterilebilir. Benzer sonuçlar parametre vektörü için geçerlidir. Bununla beraber, izdüşümler ortogonal değildirler. Örneğin,

$$
\begin{aligned}
\hat{\beta}_C &= \left(X'X\right)^{-1} X'Y_C \\
&= \left(X'X\right)^{-1} X'Q_{\gamma(C)}\hat{Y} \\
&= \left(X'X\right)^{-1} X'Q_{\gamma(C)}X\hat{\beta}
\end{aligned}
$$

dır. Cebirsel olarak $\left(X'X\right)^{-1} X'Q_{\gamma(C)}X$ dönüşümünün A_c ye, yani, E^q nun $\left(X'X\right)^{-1}C$ boyunca $C^{\perp}$ üzerine izdüşümüne eşit olacağı kolayca gösterilebilir. Ayrıca,

$$\hat{\beta}_{C,H} = \left(X'X\right)^{-1} X'Q_{\Pi(C,H)}X\hat{\beta}_C$$

ve

$$\hat{\beta}_{C,H} = \left(X'X\right)^{-1} X'P_{\omega(C,H)}X\hat{\beta}$$

yazılabilir. Öte yandan

$$D_{C,H} = \left(X'X\right)^{-1} X'Q_{\Pi(C,H)}X = I - A_C\left(X'X\right)^{-1} H\left\{H'A_C\left(X'X\right)^{-1} H\right\}^{-1} H'A_C$$

dönüşümü $A_C\left(X'X\right)^{-1}$ boyunca $\left(X'X\right)^{-1}C + T^{\perp}$ üzerine ortogonal olmayan bir izdüşümdür. Halbuki $\left(X'X\right)^{-1} X'P_{\omega(C,H)}X$ ifadesi $T = (C,H)$ olmak üzere $\left(X'X\right)^{-1}T$ boyunca $T^{\perp}$ üzerine ortogonal olmayan izdüşümdür.

Altı tahmini hesaplamak için etkin bir şema aşağıda ortaya konmuştur. $D_{C,H}\hat{\beta}_C = F_{C,H}\hat{\beta}_C$ olduğuna dikkat edelim, burada $A_C\hat{\beta}_C = \hat{\beta}_C$ olduğundan, $F_{C,H} = I - A_C\left(X'X\right)^{-1} H\left\{H'A_C\left(X'X\right)^{-1} H\right\}^{-1} H'$ olacaktır.

$$
\begin{tikzcd}
 & \hat{Y} & & \\
Y \arrow[r, "(X'X)^{-1}X'"] & \hat{\beta} \arrow[u, "X"] \arrow[r, "A_C"] & \hat{\beta}_C \arrow[r, "X"] \arrow[d, "F_{C,H}"'] & \hat{Y}_C \\
 & & \hat{\beta}_{C,H} \arrow[r, "X"] & \hat{Y}_{C,H}
\end{tikzcd}
$$

Şekil 3.1 Tahminin Şeması

$C'\beta = m \neq 0$, $H'\beta = h \neq 0$ Durumu: Bu kısımda, (3.3) sisteminin (3.4) formunda nasıl yazılabildiğini gösterelim. α_h (α_m); $H'\beta = h$ ($C'\beta = m$) denkleminin herhangi bir çözümü olsun. Bu denklemin tüm çözümlerinin kümesi aşağıdaki gibi tanımlanabilir.

$H'\beta = h$ ($C'\beta = m$) olması için gerek ve yeter şart $\beta \in \alpha_H + H^{\perp}\left(\alpha_m + C^{\perp}\right)$ olmasıdır. Bu nedenle, $S = \{\beta : C'\beta = m$ ve $H'\beta = h\} = \left(\alpha_m + C^{\perp}\right) \cap \left(\alpha_h + H^{\perp}\right)$ yazılabilir. Bu durumda S boş değildir. Gerçekten, herhangi bir $\alpha_0 \in S$ için, S kümesi $S = \alpha_0 + \left(C^{\perp} \cap H^{\perp}\right)$ olarak tanımlanabilir. β_0 S deki herhangi bir vektör olsun. Bu takdirde $C'\beta_0 = m$ ve $H'\beta_0 = h$ yazılabilir. $\tilde{\beta} = \beta - \beta_0$ ve $\tilde{Y} = Y - X\beta_0$ alarak (3.3) bağıntısı (3.4) şeklini alır.

3.1.5. Örnekler

Bu kısımda baştan sona A faktörü üç düzeyli ve B faktörü iki düzeyli olmak üzere iki faktörlü bir tasarımı göz önüne alalım. Hücre ortalamaları aşağıdaki gibidir.

Çizelge 3.1 Hücre Ortalamaları

	B_1	B_2
A_1	μ_{11}	μ_{12}
A_2	μ_{21}	μ_{22}
A_3	μ_{31}	μ_{32}

$\mu = (\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}, \mu_{31}, \mu_{32})'$, $n_{ij} = i, j$ kutucuğundaki Y_{ijk} gözlemlerinin sayısı, $Y_{ij} = \left(Y_{ij1}, \ldots, Y_{ijn_{ij}}\right)'$, $Y = (Y'_{11}, Y'_{12}, Y'_{21}, Y'_{22}, Y'_{31}, Y'_{32})'$, $j_{ij} = 1$ lerin $n_{ij} \times 1$ sütun vektörü ve $E(Y) = X\mu$ olsun, burada

$$X = \begin{bmatrix} j_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & j_{12} & 0 & 0 & 0 & 0 \\ 0 & 0 & j_{21} & 0 & 0 & 0 \\ 0 & 0 & 0 & j_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 & j_{31} & 0 \\ 0 & 0 & 0 & 0 & 0 & j_{32} \end{bmatrix}$$

ve X de görünen her bir 0; 0 ların uygun boyuta sahip olan bir sütun vektörüdür.

Tam Olarak Rasgeleleştirilmiş Tasarım (Kısıtlanmamış Model): Bu model için hücre ortalamalarının vektörü

$$\hat{\mu} = \left(X'X\right)^{-1} X'Y = \left(\bar{Y}_{11\cdot}, \bar{Y}_{12\cdot}, \cdots, \bar{Y}_{32\cdot}\right)'$$

ile tahmin edilir. Etkileşimin olmadığının sıfır hipotezi

$$H' = \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix}$$

olmak üzere, $H_0 : H'\mu = 0$ olarak yazılabilir. Pay kareler toplamı $\hat{\beta}$ parametre vektörü için $\hat{\mu}$ koyarak, (3.5) bağıntısını kullanarak hesaplanabilir.

Rasgeleleştirilmiş Blok Tasarım (Kısıtlanmış Model): Bu modelde bir rasgelleleştirilmiş blok tasarımında işlem etkileşimi ile hiçbir blok mevcut değildir. Bu nedenle, kutucuk ortalamaları, C matrisi hemen bir önceki kısımdaki H matrisi olmak üzere, $C'\mu = 0$ denklemiyle kısıtlanır. Bu durumda $X = I$ olduğundan, (3.8) bağıntısı μ nün tahmini olarak

$$\hat{\mu}_C = \left\{I - C\left(C'C\right)^{-1} C'\right\} Y$$

yi verir. Genel olarak, (3.8) bağıntısı tahminler olarak örnek (hücre) ortalamaları gibi basit ifadeleri vermez. İşlem etkisi için test etmek için, her i, i' için $H_0 : \mu_i = \mu_{i'}$ sıfır hipotezini test edelim. Matris notasyonu ile H_0 hipotezi

$$H' = \begin{bmatrix} 1 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 & -1 \end{bmatrix}$$

olmak üzere, $H'\mu = 0$ olarak yazılabilir. (3.15) bağıntısı payda kareler toplamı olarak $Y'C\left(C'C\right)^{-1} C'Y$ yi verir iken, (3.14) bağıntısı test için

$$\hat{\mu}_C' H \left[H' \left\{I - C\left(C'C\right)^{-1} C'\right\} H \right]^{-1} H' \hat{\mu}_C$$

pay kareler toplamını verir.

3.2 Çoklu Lineer Regresyonda Lineer Kısıtlamalar: Varyans Analizi

Bir $Y = X\beta + \varepsilon$ çoklu lineer regresyonunu göz önüne alalım ve s sayıda bir lineer denklem takımıyla verilen bir hipotezi test etmek istediğimizi farz edelim. A bir $s \times p$ matris ve c bir $s \times 1$ vektör olmak üzere, bir matris formunda

$$H_0 : A\beta = c$$

dir. $s \leq p$ olduğunu ve A matrisinin s rankına sahip olduğunu kabul edeceğiz. Parametrelerin sadece bir lineer kombinasyonunu (s = 1 durumu) göz önüne aldığımızda veya eş anlı olarak tüm parametreler hakkındaki hipotezi (s = p durumu) test ettiğimizde bu, daha önce verilen iki tip hipotezi genelleştirir. Bu genel hipotezi test etmek için doğal bir düşünce $A\hat{\beta}$ nın c den ne kadar uzak olduğunu karşılaştırmaktır ve bunu yapmak için $A\hat{\beta}$ nın dağılımını bulmamız gerekir. Aşikar olarak, bu dağılım $A\beta$ ortalamalı ve

$$D := A\left(X'X\right)^{-1} A'$$

olmak üzere,

$$EA\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)^T A^T = Akov\left(\hat{\beta}\right)A^T = \sigma^2 A\left(X^T X\right)^{-1} A^T = \sigma^2 D$$

kovaryanslı normal dağılımdır. D matrisi simetrik pozitif tanımlı tersinir bir $s \times s$ matristir ve bu nedenle, onun $D^{1/2}$ karekökünü alabiliriz. $D^{-1/2}A\left(\hat{\beta}-\beta\right)$ nın kovaryansı $\sigma^2 I$ olacaktır. Bu ise

$$\frac{1}{\sigma^2}\left|D^{-1/2}A\left(\hat{\beta}-\beta\right)\right|^2 = \frac{1}{\sigma^2}\left(A\left(\hat{\beta}-\beta\right)\right)^T D^{-1}A\left(\hat{\beta}-\beta\right) \sim \chi_s^2$$

olduğunu belirtir. Buradan sıfır hipotezi altında, yani $A\beta = c$ altında,

$$\frac{1}{\sigma^2}\left(A\hat{\beta}-c\right)^T D^{-1}\left(A\hat{\beta}-c\right) \sim \chi_s^2 \qquad (3.17)$$

elde edilir. Bu durumda $n\hat{\sigma}^2/\sigma^2 \sim \chi_{n-p}^2$ olup $\hat{\beta}$ dan bağımsız olduğundan,

$$\begin{aligned}&\frac{1}{s\sigma^2}\left(A\hat{\beta}-c\right)^T D^{-1}\left(A\hat{\beta}-c\right)\Big/\frac{n\hat{\sigma}^2}{(n-p)\sigma^2}\\&=\frac{n-p}{ns\hat{\sigma}^2}\left(A\hat{\beta}-c\right)^T D^{-1}\left(A\hat{\beta}-c\right)\sim F_{s,n-p}\end{aligned} \tag{3.18}$$

elde edilir. Bu ise H_0 hipotezini test etmek için yeterlidir. Bununla beraber, çeşitli uygulamalarda, (3.17) nin farklı bir eşdeğer gösterimini kullanmak daha yararlıdır. Bu β nın H_0 daki kısıtlamayı sağlayan $\hat{\beta}_A$ en çok olabilirlik tahmini (MLE) ile verilir. Başka bir deyişle,

$$|Y-X\beta|^2 \to A\beta = c \text{ kısıtlaması altında } \min_{\beta} \tag{3.19}$$

yı çözerek elde edilir.

Lemma 3.1. Eğer $\hat{\beta}_A$ (3.19) un çözümü ise, bu takdirde (3.17) nin sol yanı

$$\frac{1}{\sigma^2}\left|X\left(\hat{\beta}_A-\hat{\beta}\right)\right|^2 \tag{3.20}$$

ye eşittir.

İspat. İlk olarak açık bir şekilde kısıtlanmış $\hat{\beta}_A$ MLE sini bulalım. Lagrange çarpanları yöntemi vasıtasıyla, λ bir $s\times 1$ vektör olmak üzere, aşağıdaki iki denklemli sistemi çözmeliyiz:

$$A\beta = c,\ \frac{\partial}{\partial\beta}\left(|Y-X\beta|^2+(A\beta-c)^T\lambda\right)=0$$

Buradan ikinci denklem

$$-2X'Y+2X'X\beta+A'\lambda=0$$

şeklindedir. Bu durumda β için bunu çözmek

$$\hat{\beta}_A=\left(X^TX\right)^{-1}X^TY-\frac{1}{2}\left(X^TX\right)^{-1}A^T\lambda=\hat{\beta}-\frac{1}{2}\left(X^TX\right)^{-1}A^T\lambda$$

yı verir. $\hat{\beta}_A$ nın lineer kısıtlamayı sağlaması gerektiğinden,

$$c=A\hat{\beta}_A=A\hat{\beta}-\frac{1}{2}A\left(X^TX\right)^{-1}A^T\lambda=A\hat{\beta}-\frac{1}{2}D\lambda$$

elde edilir. Bunu λ için çözersek $\lambda = 2D^{-1}\left(A\hat{\beta} - c\right)$ elde edilir. Buradan

$$\hat{\beta}_A = \hat{\beta} - \left(X^T X\right)^{-1} A^T D^{-1}\left(A\hat{\beta} - c\right)$$

ve böylece

$$X\left(\hat{\beta}_A - \hat{\beta}\right) = -X\left(X^T X\right)^{-1} A^T D^{-1}\left(A\hat{\beta} - c\right)$$

elde edilir. Öte yandan

$$\begin{aligned}
\left|X\left(\hat{\beta}_A - \hat{\beta}\right)\right|^2 &= \left(X\left(A\hat{\beta} - \hat{\beta}\right)\right)' X\left(\hat{\beta}_A - \hat{\beta}\right) \\
&= \left(A\hat{\beta} - c\right)'\left(X\left(X'X\right)^{-1} A'D^{-1}\right)' X\left(X'X\right)^{-1} A'D^{-1}\left(A\hat{\beta} - c\right) \\
&= \left(A\hat{\beta} - c\right)' D^{-1}A\left(X'X\right)^{-1} X'X\left(X'X\right)^{-1} A'D^{-1}\left(A\hat{\beta} - c\right) \\
&= \left(A\hat{\beta} - c\right)' D^{-1}A\left(X'X\right)^{-1} A'D^{-1}\left(A\hat{\beta} - c\right) \\
&= \left(A\hat{\beta} - c\right)' D^{-1}DD^{-1}\left(A\hat{\beta} - c\right) \\
&= \left(A\hat{\beta} - c\right)' D^{-1}\left(A\hat{\beta} - c\right)
\end{aligned}$$

yı hesaplamak için, bu formülü kullanabiliriz. Bunu (3.17) ile karşılaştırırsak Lemma 3.1 in ispatı tamamlanır. Bu durumda (3.18) ve Lemma 3.1 i kullanarak, H_0 sıfır hipotezi altında

$$\frac{n-p}{ns\hat{\sigma}^2}\left|X\left(\hat{\beta}_A - \hat{\beta}\right)\right|^2 \sim F_{s,n-p} \tag{3.21}$$

elde edilir. Bu çoklu lineer regresyonun özel durumları olan birçok farklı model vardır ve bu model hakkındaki birçok hipotez genel lineer kısıtlamalar olarak yazılabilir. Böyle bir modeli varyans analizinde bir–yönlü planlamayı ayrıntılı bir şekilde tanımlayacağız. Bu takdirde düşüncemiz açık olacağından, ayrıntılara girmeksizin diğer modellerin bir çiftini de ele alabiliriz. $n_1,\ldots,n_p$ büyüklüklerine sahip *p* sayıda bağımsız

$$\begin{gathered}
Y_{11},\ldots,Y_{1n_1} \sim N\left(\mu_1, \sigma^2\right) \\
\vdots \\
Y_{p1},\ldots,Y_{pn_p} \sim N\left(\mu_p, \sigma^2\right)
\end{gathered}$$

örneklemlerinin uygun bir şekilde verildiğini farz edelim. Ayrıca tüm dağılımların varyansının aynı olduğunu kabul edelim. Tüm dağılımların ortalamalarının eşit olduğu

$$H_0 : \mu_1 = \ldots = \mu_p$$

hipotezini test etmek isteyelim. Bu problem aslında bir çoklu regresyonun ve lineer denklemler vasıtasıyla verilen hipotezi test etmenin bir özel durumudur. $k = 1, \ldots, p$ ve $i = 1, \ldots, n_i$ için, $g_{ki} \sim N\left(0, \sigma^2\right)$ olmak üzere, $Y_{ki} = \mu_k + \varepsilon_{ki}$ yazılabilir. $n = n_1 + \cdots + n_p$ olmak üzere,

$$Y = (Y_{11}, \ldots, Y_{1n_1}, \ldots, Y_{p1}, \ldots, Y_{pn_p})'$$

$n \times 1$ vektörünü ve

$$\mu = (\mu_1, \ldots, \mu_p)'$$

$p \times 1$ parametre vektörünü göz önüne alalım. Bu takdirde X

$$X = \begin{pmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \cdots & 0 & \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 1 \end{pmatrix}'$$

gibi bir $n \times p$ matris olmak üzere tüm denklemleri

$$Y = X\mu + \varepsilon$$

matris formunda yazabiliriz. Bloklar $n_1, \ldots, n_p$ satırlarına sahiptir. Aslında X açıklayıcı matris, gözlemin hangi gruba ait olduğu ile ilgili göstergelerden oluşur.

$$A = \begin{pmatrix} 1 & 0 & \cdots & 0 & -1 \\ 0 & 1 & \cdots & 0 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{pmatrix}$$

$\left(p-1\right) \times p$ matrisi için, H_0 hipotezi bir matris formunda $A\mu = 0$ olarak yazılabilir. Bu takdirde $F_{p-1,n-p}$ dağılımına sahip olacak olan (3.21) deki istatistiği hesaplamamız gerekir. Herşeyden önce,

$$X^T X = \begin{pmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n_p \end{pmatrix}$$

olacaktır. $\hat{\mu} = \left(X^T X\right)^{-1} X^T Y$ olduğundan, her bir $i \le p$ için, $\hat{\mu}_i = \frac{1}{n_i}\sum_{k=1}^{n_i} Y_{ik} = \overline{Y}_i$ nin *i*–yinci örneklemin ortalaması olduğunu görmek kolaydır. Aynı zamanda

$$\hat{\sigma}^2 = \frac{1}{n}\left|Y - X\mu\right|^2 = \frac{1}{n}\sum_{i=1}^{p}\sum_{k=1}^{n_i}\left(Y_{ik} - \overline{Y}_i\right)^2$$

eşitliği de elde edilir. $A\mu = 0$ lineer kısıtlaması altında $\hat{\mu}_A$ MLE sini bulmak için, sadece $\left|Y - X\mu\right|^2$ yi eşit koordinatlara sahip $\mu = (\mu_1, \dots \mu_1)'$ vektörleri üzerinden minimumlaştırmamız gerekir. Ancak, $\mu = (\mu_1, \dots \mu_1)'$ olmak üzere $X\mu$ $n \times 1$ boyutlu bir vektör olduğundan,

$$\sum_{i=1}^{p}\sum_{k=1}^{n_i}\left(Y_{ik} - \mu_1\right)^2 \min_{\mu_1}$$

ifadesini minimumlaştırmalıyız ve bu durumda $\mu_1 = \frac{1}{n}\sum_{i=1}^{p}\sum_{k=1}^{n_i} Y_{ik} = \overline{Y} -$ örneklemlerin genel (yekün) ortalamasını elde ederiz. Bu nedenle,

$$\hat{\mu}_A - \hat{\mu} = \left(\overline{Y} - \overline{Y}_1, \dots, \overline{Y} - \overline{Y}_p\right)^T$$

ve

$$\left|X\left(\hat{\mu}_A - \hat{\mu}\right)\right|^2 = \sum_{i=1}^{p}\sum_{k=1}^{n_i}\left(\overline{Y}_i - \overline{Y}\right)^2 = \sum_{i=1}^{p} n_i\left(\overline{Y}_i - \overline{Y}\right)^2$$

olacaktır. Bu durumda (3.21) e göre, H_0 hipotezi altında,

$$F := \frac{n-p}{p-1}\frac{\sum_{i=1}^{p} n_i\left(\overline{Y}_i - \overline{Y}\right)^2}{\sum_{i=1}^{p}\sum_{k=1}^{n_i}\left(Y_{ik} - \overline{Y}_i\right)^2} \sim F_{p-1,n-p} \tag{3.22}$$

yazılabilir. H_0 hipotezini test etmek için, $F_{p-1,n-p}\left(c_\alpha, +\infty\right) = \alpha$ olmak üzere, bir

$$\delta = \begin{cases} H_0, & F \le c_\alpha \\ H_1, & F > c_\alpha \end{cases}$$

karar kuralını tanımlayalım. (3.22) nin payındaki toplam her bir kitlenin $\overline{Y}_i$ örneklem ortalamalarının $\overline{Y}$ genel ortalaması etrafındaki toplam değişimini gösterir. Paydadaki toplam Y_{ik} gözlemlerinin onların özel $\hat{Y}_i$ örneklem ortalamaları etrafındaki toplam değişimini gösterir. Test istatistiğinin bu yorumu varyans analizi veya ANOVA adını açıklar.

Şimdi de iki–yönlü tasarımı göz önüne alalım. Yine farklı gruplardan örneklemlere sahip olduğumuzu, ancak şimdi grupların iki faktörle tanımlanan iki kategoriye sahip olacağını farz edelim. Örneğin, eğer farklı devletlerde fakat ayrı halk ve özel okullardaki SAT puanlarını karşılaştırmak istersek, bu takdirde iki faktör – devlet ve okul tipi – vasıtasıyla tanımlanan gruplara sahip olacağız. $i = 1,\ldots,a,\ j = 1,\ldots,b$ ve $k = 1,\ldots,n_{ij}$ için verinin aşağıdaki modelini göz önüne alalım.

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

Yani, birinci faktörün a sayıda kategorisine, ikinci faktörün b sayıda kategorisine ve (i, j) grubunda n_{ij} sayıda gözleme sahip olduğumuzu göz önüne alalım. Bu modelin bir–yönlü ANOVA dan hiçbir farkı yoktur, sadece gruplar iki parametreyle/faktörle indislenir, ancak parametrelerin tahmini bir–yönlü ANOVA daki gibi gerçekleştirilebilir. Bununla beraber, bu iki faktörün etkileri hakkındaki çeşitli hipotezleri test etmek için modeli aşağıdaki gibi eşdeğer bir biçimde yazmak daha uygundur.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

Burada

$$\sum_{i=1}^{a} \alpha_i = 0, \sum_{j=1}^{b} \beta_j = 0, \sum_{i=1}^{b} \gamma_{ij} = \sum_{j=1}^{b} \gamma_{ij} = 0$$

olduğu kabul edilir. Bu kısıtlamalar tüm parametreleri orijinal μ_{ij} parametrelerinden bir tek şekilde tanımlar. Burada μ parametresine genel ortalama denir. İki faktörün

α_i ve β_j toplamsal γ_{ij} en genel etkileşim etkisinden ayırmamızın nedeni çeşitli hipotezleri bu parametrelere dayanarak formüllemek daha kolay olmasıdır. Örneğin;

- $H_0 : \alpha_1 = \ldots = \alpha_a = 0$ birinci faktörün toplamsal etkisi önemsizdir.
- $H_0 : \beta_1 = \ldots = \beta_b = 0$ ikinci faktörün toplamsal etkisi önemsizdir.
- $H_0 :$ her $\gamma_{ij} = 0$ – her iki faktörün etkileşiminin etkisi anlamsızdır (önemsizdir), yani faktörlerin etkisi toplamsaldır.

Eğer tüm grupların n_{ij} büyüklükleri eşit ise, yani veri dengeli ise, “ANOVA 2” matlab fonksiyonu ANOVA nın iki yönlü tasarımını yapar. Eğer grupların büyüklükleri (boyutları) farklı ise, “ANOVA N” yani, genel bir N–yönlü ANOVA yı kullanmak gerekir. Verinin tüm grupları sürekli bir açıklayıcı değişkene sahip olduğundaki durum çoklu regresyonun başka bir özel durumudur. Bu durumda model $i = 1, \ldots, a$ ve $k = 1, \ldots, n_i$ için,

$$Y_{ik} = \alpha + \alpha_i + \left(\beta + \beta_i\right) X_{ik} + \varepsilon_{ik}$$

olacaktır. Burada *a* sayıda gruba ve *i*–yinci grupta n_i sayıda gözleme sahibiz. Bu durumda parametreleri bir tek şekilde belirlemek için

$$\sum_{i=1}^{a} \alpha_i = 0, \sum_{i=1}^{a} \beta_i = 0$$

olduğu kabul edilir.

3.3 Standart Lineer Modelde Hipotez Test Etme

3.3.1 Alışılmış En Küçük Kareler Tahmin Edicisi

K sayıda *X* açıklayıcıları üzerinde (herhangi bir sabit terimi içeren) *N* sayıda gözlem ve bir *y* tepkimesi var olsun. Burada *X* matrisinin $N \times K$ tipinde ve *K* rankına sahip olduğunu kabul edelim. Bir sabit terim tümü birlerden oluşan bir sütundur. Bu durumda model

$$y = X\beta + \varepsilon$$

şeklindedir. Burada

$$E(\varepsilon)=0 \text{ ve } Var(\varepsilon)=E(\varepsilon\varepsilon')=\sigma^2 I$$

dır. Bu durumda β katsayılarının $\hat{\beta}_{OLS}$ OLS tahmin edicisi

$$\hat{\beta}_{OLS}=(X'X)^{-1}X'y \tag{3.23}$$

ile verilir. Sabit terim ile ilgili katsayıya regresyon sabiti denir.

$$\hat{y}=X\hat{\beta}_{OLS}$$

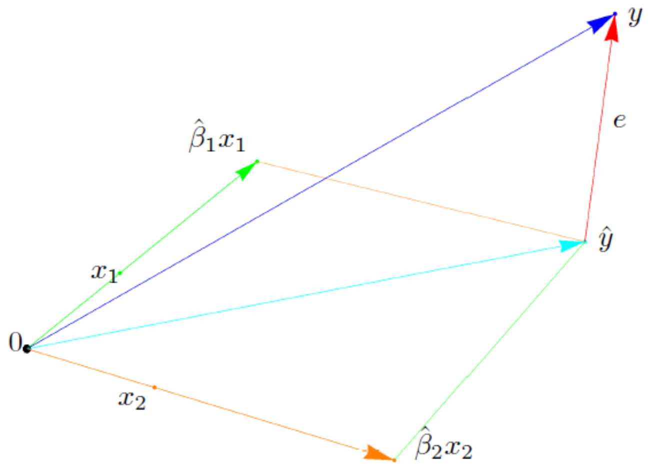
ve

$$e=y-\hat{y}=y-X\hat{\beta}_{OLS}$$

alalım. Bu durumda hataların (tahmin edilen hataların) *e* vektörü

$$X'e=0$$

eşitliğini sağlar. Yani, *e* vektörü açıklayıcıların her bir sütun vektörüne ortogonaldır. Eğer regresyonda bir sabit terim varsa, bu takdirde tahmin edilen hataların toplamı sıfırdır, yani $j'e=0$ olacaktır. Burada $j'=[1,1,\ldots,1]$ dir, bu nedenle tahmin edilen hata ortalaması $\bar{e}$ de sıfırdır. Bazen katsayılarımız hakkında daha karmaşık sıfır hipotezlerini göz önüne almak gerekebilir. Örneğin, ekonomistler çoğu kez tahmin edilen gelir paylaşımının bire eşit olup olmadığı ile ilgilenirler.



Şekil 3.2 OLS nin Geometrisi

β vektörü üzerinde q sayıda lineer kısıtlamanın konulduğu genel lineer hipotezi basit bir şekilde

$$H_0 : a = A\beta$$

olarak yazılabilir. Burada A bir $q \times K$ matristir ve a, $q \times 1$ sütun vektördür. A nın q rankına sahip olduğunu kabul edelim. Örneğin, $\beta_1 + \cdots + \beta_K = 1$ olan hipotez $a = [1]$ ve $A = [1, \ldots, 1]$ i kullanır. Bu şimdi imkansız bir şekilde anlaşılmaz görünebilir, ancak bunu ileride Kısım 3.3.3 te açıklayacağız.

Teorem 3.1. Sıfır hipotezi altında

$$F = \frac{1}{qs^2}\left(a - A\hat{\beta}_{OLS}\right)' \left[A\left(X'X\right)^{-1} A' \right]\left(a - A\hat{\beta}_{OLS}\right)$$

test istatistiği $(q, N-K)$ serbestlik dereceli bir Snedecor F–dağılımına sahiptir. Bu istatistiklere dayanan testlere F–testleri denir. R gibi birçok yazılım paketleri, regresyon için F–istatistiğini hesaplar. Eğer bir sabit terime sahipsek o genellikle ilk gözüken birdir, yani terminolojimizdeki X_1 dir. Regresyon için F–istatistiği sabit olmayan terimler üzerindeki tüm katsayıların sıfır olduğu sıfır hipotezini, yani

$$H_0 : \beta_2 = \beta_3 = \cdots = \beta_K = 0$$

hipotezini test eder. Bu test için F–istatistiği (K–1,N–K) serbestlik derecelerine sahiptir ve sadece açıklayıcıların biri bir sabit terim ise bu testi kullanmayı anlamlı kılar. Bu muhtemelen reddetmeyi istediğimiz bir hipotezdir, bu nedenle küçük bir p–değeri (veya büyük F–istatistiği) iyi bir şeydir. Bunun önemli olmasının nedeni, eğer X in sütunları yaklaşık olarak eş doğrusal ise, $\hat{\beta}_{K\,OLS}$ yi tahmin etmenin zor olduğu, yani onun büyük bir standart hataya sahip olacağı ve bu nedenle, her birinin istatistiksel olarak anlamlı olabilmesi, ancak onların birlikte y yi gerçek olarak belirleyebilmesidir.

3.3.2 Lagrange Çarpanları Teoremi

Burada önce genel lineer kısıtlamanın testinin niçin bir F–testine veya bağımsız χ^2 rasgele değişkenlerinin oranına dayandırıldığını ve bir minimumlaştırma probleminin üzerinde kısıtlamaların nasıl koyulacağını tanımlamamız gerektiğini açıklayacağız.

X; $\mathbb{R}^n$ nin bir alt kümesi olsun ve $f, g_1, \ldots, g_m : X \to \mathbb{R}$ fonksiyonlarının sürekli olduklarını kabul edelim. x^*, $g_i(x) = 0,\ i = 1, \ldots, m$ altında f nin bir iç kısıtlanmış yerel minimumlaştırıcısı olsun. $f, g_1, \ldots, g_m$ nin x^* da diferansiyellenebilir olduklarını ve $\nabla g_1(x^*), \ldots, \nabla g_m(x^*)$ ın lineer bağımsız olduklarını farz edelim. Bu takdirde

$$\nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0 \tag{3.24}$$

olacak şekilde $\lambda_1^*, \ldots, \lambda_m^*$ reel sayıları mevcuttur.

$$L(x, \lambda) = f(x) - \sum_{i=1}^{n} \lambda_i g_i(x)$$

fonksiyonuna Langrange fonksiyonu denir ve (3.24) bir kısıtlanmış ekstremum için birinci mertebe şartlarının kümesidir. f kuadratik olduğunda ve her bir g_i lineer olduğunda, bunun f nin kısıtlanmış minimumlaştırıcısının $L(x, \lambda^*)$ Lagrange fonksiyonunun bir kısıtlanmamış minimumlaştırıcısı olduğu sonucunu verir.

3.3.3 Kısıtlanmış Regresyon

Kısım 3.3.1 de, katsayılar üzerindeki bir lineer kısıtlanmanın nasıl test edildiğini belirtmiştik. Bu kısımda onun nasıl elde edildiği verilecektir.

$$y = X\beta + \varepsilon \tag{3.25}$$

modelinde, A $q \times K$ ve tam ranklı (tam sütun ranklı) olmak üzere, β üzerindeki

$$H_0 : a = A\beta$$

lineer kısıtlanmasını test etmek için ilk olarak hata kareler toplamını minimumlaştırmak suretiyle (3.25) i tahmin ederiz, bu bize alışılagelen tahminleri verir, ancak onları

$$\hat{b}_u = (X'X)^{-1} X'y\,,\ \hat{y}_u = X\hat{b}_u\,,\ e_u = y - \hat{y}_u$$

olarak isimlendirelim. Bundan sonra $a = Ab$ kısıtlamasını koyalım ve $\hat{b}_r, \hat{y}_r$ ve e_r hatalar kısıtlanmış tahminlerini elde etmek için, tahmin edilen hata kareler toplamını (b ye göre) minimumlaştıralım. Kısıtlamalar altında tahmin edilen hataların kareler toplamını minimum yapmak için, y yi X in sütunları tarafından gerilen $\{Xb : b \in \mathbb{R}^K\}$ uzayına iz düşürmek yerine, $\{Xb : b \in \mathbb{R}^K \text{ ve } Ab = a\}$ üzerine iz düşürmeliyiz. Burada y_r noktası yine X in sütunlarının bir lineer kombinasyonudur ve bu nedenle e_u kısıtlanmamış hata vektörü $y_u - y_r$ ye ortogonaldir, bu durumda Pisagor teoremi bize

$$e_r{}'e_r - e_u{}'e_u = (y_u - y_r)'(y_u - y_r)$$

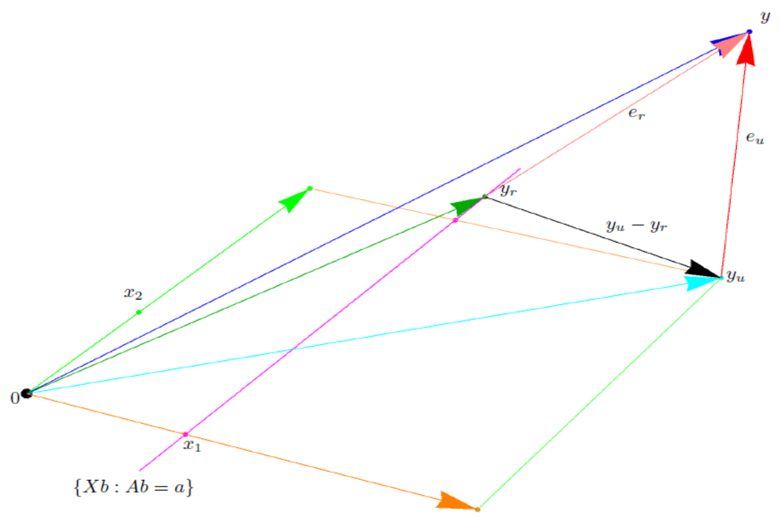
olduğunu söyler Bkz. Şekil 3.3. İlk soru, $\hat{b}_r$ için formülün ne olduğudur. Bunu cevaplandırmak için, λ Lagrange çarpanlarının bir q–vektörü olmak üzere, bu minimumlaştırma için Lagrange'ın

$$(y - Xb)'(y - Xb) - \lambda'(Ab - a)$$

olduğuna dikkat edelim. Lagrange Çarpanları Teoremi kısıtlanmış minimum için birinci mertebe şartının Lagrange'ın b ye göre kısmi türevlerinin vektörünün sıfır olması gerektiğini, yani

$$-2X'y + 2X'X\hat{b}_r - A'\lambda^* = 0 \qquad (3.26)$$

olması gerektiğini ifade eder.



Şekil 3.3 $\beta_1 = 1$ Kısıtlaması Altında Regresyon

y_r, y_u, y; $\overline{y_r y}$ hipotenüslü bir dik üçgen oluşturur. Bu durumda bu ifadeyi $A(X'X)^{-1}$ ile soldan çarparak

$$-2\underbrace{A(X'X)^{-1}X'y}_{=\hat{b}_u}+2\underbrace{A(X'X)^{-1}(X'X)\hat{b}_r}_{=A\hat{b}_r=a}-A(X'X)^{-1}A'\lambda^*=0$$

veya

$$A(X'X)^{-1}A'\lambda^* = 2\left(a - A\hat{b}_u\right) \tag{3.27}$$

elde edilir. Şimdi $A(X'X)^{-1}A'$ nün tam ranklı bir $q \times q$ matris olduğunu ve bu nedenle onun bir terse sahip olduğunu iddia ediyoruz. Bu bize (3.27) bağıntısını λ^* için çözme imkanı verir, bu nedenle

$$\lambda^* = 2\left[A(X'X)^{-1}A'\right]^{-1}\left[a - A\hat{b}_u\right]$$

elde edilir. Öte yandan

$$-X'y + X'X\hat{b}_r - A'\left[A(X'X)^{-1}A'\right]^{-1}\left[a - A\hat{b}_u\right] = 0$$

yi elde etmek için bunu (3.26) eşitliğinde yerine koyalım. Bu durumda bu eşitliği $(X'X)^{-1}$ ile soldan çarparak

$$-\underbrace{\left(X'X\right)^{-1}X'y}_{=\hat{b}_u}+\underbrace{\left(X'X\right)^{-1}X'X\hat{b}_r}_{=\hat{b}_r}-\left(X'X\right)^{-1}A'\left[A\left(X'X\right)^{-1}A'\right]^{-1}\left[a-A\hat{b}_u\right]=0$$

olduğu ve böylece kısıtlanmış tahmin edici için formülün

$$\hat{b}_r=\hat{b}_u+\left(X'X\right)^{-1}A'\left[A\left(X'X\right)^{-1}A'\right]^{-1}\left[a-A\hat{b}_u\right] \tag{3.28}$$

olduğu görülür. Buradan akla gelecek bir diğer soru tahmin edilen hata kareler toplamının ne olduğudur. Haklı olarak

$$e_r'e_r=\left(y-X\hat{b}_r\right)'\left(y-\hat{b}_r\right)$$

dır ve $C=\left(X'X\right)^{-1}A'\left[A\left(X'X\right)^{-1}A'\right]^{-1}\left[a-A\hat{b}_u\right]$ olmak üzere,

$$y-X\hat{b}_r=\underbrace{y-X\hat{b}_u}_{=e_u}-X\left(X'X\right)^{-1}A'\left[A\left(X'X\right)^{-1}A'\right]^{-1}\left[a-A\hat{b}_u\right]=e_u+XC$$

yazmak için (3.28) bağıntısını kullanabiliriz. Bu ise

$$e_r'e_r=\left(e_u+XC\right)'\left(e_u+XC\right)=e_u'e_u+2e_u'XC+\left(XC\right)'\left(XC\right)$$

eşitliğini yazabileceğimizi ifade eder. Ancak, e_u kısıtlanmamış hata tahmini vektörü X in sütunlarına ortogonal olduğundan, $e_u'X=0$ yazılır. Şimdi bir çarpımın transpozesinin tranpozenin ters sırada çarpımı olduğuna ve $\left[A\left(X'X\right)^{-1}A'\right]^{-1}$ ve $\left(X'X\right)^{-1}$ in simetrik olduklarına dikkat ederek

$$\begin{aligned}
&\left(XC\right)'\left(XC\right)\\
&=\left[a-A\hat{b}_u\right]'\left[A\left(X'X\right)^{-1}A'\right]^{-1}A\left(X'X\right)^{-1}X'X\left(X'X\right)^{-1}A'\left[A\left(X'X\right)^{-1}A'\right]^{-1}\left[a-A\hat{b}_u\right]\\
&=\left[a-A\hat{b}_u\right]'\left[A\left(X'X\right)^{-1}A'\right]^{-1}\left[a-A\hat{b}_u\right]
\end{aligned}$$

yazılabilir. Böylece

$$e_r'e_r=e_u'e_u+\left[a-A\hat{b}_u\right]'\left[A\left(X'X\right)^{-1}A'\right]^{-1}\left[a-A\hat{b}_u\right] \tag{3.29}$$

olacaktır. Şimdi $A(X'X)^{-1}A'$ $q\times q$ matrisinin tam ranka sahip olduğunu görmek için, herhangi bir *b* *q*–vektörü için,

$$A(X'X)^{-1}A'b=0$$

olduğunu farz edelim. Bu durumda $b=0$ olduğunu göstermek istiyoruz. *A* tam ranka sahip olduğundan, $A'A$ ve $(A'A)^{-1}$ de tam ranka sahip olacaktır. Bu nedenle

$$(X'X)(A'A)^{-1}A'A(X'X)^{-1}A'b=A'b=0$$

eşitliğini elde etmek için, yukardaki denklemi $(X'X)(A'A)^{-1}A'$ ile soldan çarpalım. Bu durumda *A* tam *q* rankına sahip olduğundan bu, $b=0$ olduğunu gösterir. Bu ise istenen sonuçtur. H_0 hipotezinin bir testi,

$$H_0:\frac{e_r'e_r}{e_u'e_u}=1$$

sıfır hipotezinin

$$H_1:\frac{e_r'e_r}{e_u'e_u}=\frac{e_u'e_u+\left[a-A\hat{b}_u\right]'\left[A(X'X)^{-1}A'\right]^{-1}\left[a-A\hat{b}_u\right]}{e_u'e_u}>1$$

alternatif hipotezine karşı bir testi anlamına gelir. Bu durumda $s^2=e_u'e_u/(N-K)$ nın varyansın yansız tahmini olduğunu hatırlayarak, oranı yeniden

$$1+\frac{\left[a-A\hat{\beta}_{OLS}\right]'\left[A(X'X)^{-1}A'\right]^{-1}\left[a-A\hat{\beta}_{OLS}\right]}{(N-K)s^2}$$

olarak yazabiliriz. Bu nedenle,

$$F=\frac{\left[a-A\hat{b}_u\right]'\left[A(X'X)^{-1}A'\right]^{-1}\left[a-A\hat{b}_u\right]\Big/q}{e_u'e_u\Big/(N-K)}$$

test istatistiğini oluşturabiliriz. Eğer $a=A\hat{b}_u$ ise, yani eğer sıfır hipotezi doğru ise, bu takdirde payın $\chi^2(q)$ olarak dağıldığını gösterebiliriz. Payda ise $\sigma^2\chi^2(n-k)$ olarak dağılır ve paydanın paydan bağımsız olduğu gösterilebilir. Bu nedenle, sıfır hipotezi altında, test istatistiği $(q, n-k)$ serbestlik derecelerine sahip bir *F*–dağılımına sahiptir. Eğer $F \geq F_{1-\alpha,q,n-k}$ ise, sıfır hipotezi reddedilmelidir.

3.3.4 *F*–testine Karşı *T*–testi

Şimdi $H_1 : \beta_j \neq 0$ alternatif hipotezine karşı $H_0 : \beta_j = 0$ sıfır hipotezini test etmek için yöntemleri gördük. Bir yöntem

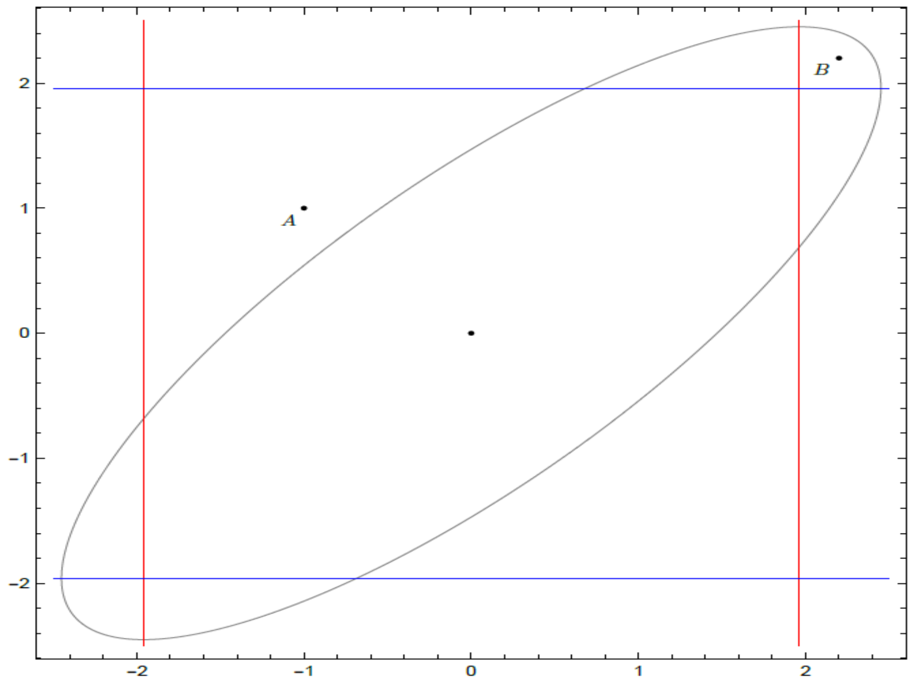
$$t = \frac{\hat{\beta}_{jOLS}}{s\sqrt{(X'X)^{-1}_{jj}}}$$

test istatistiğine dayanan *t*–testini kullanmaktır. Diğeri onu $e^{j'}\beta = 0$ lineer kısıtlaması olarak test etmek ve bu durum için

$$F = \frac{\hat{\beta}_{jOLS}\left[(X'X)^{-1}_{jj}\right]^{-1}\hat{\beta}_{jOLS}}{s^2}$$

olan $\left(A = e^{j'}, a = 0, q = 1\right)$ *F*–istatistiğini kullanmaktır. $F = t^2$ olduğuna, bu nedenle her iki testin (aynı α anlam düzeyi ile) de bizi aynı karara götürdüğüne dikkat edelim.

Öte yandan bir tek $\hat{\beta}_{jOLS}$ için *t*–testi ve *F*–testi uyuşur iken, bir ortak test ve iki ayrı test arasında uyuşma olmayabilir. Örneğin, $H_0 : \left(\hat{\beta}_1, \hat{\beta}_2\right) = (0,0)$ ortak sıfır hipotezini göz önüne alalım. Eğer $\left(\hat{\beta}_1, \hat{\beta}_2\right)$ vektörü Şekil 3.4 te gösterilen elipsin dışında uzanırsa bu $\alpha = 0.05\%$ düzeyinde reddedilecektir. *A* noktası böyle bir noktadır, ancak *B* değildir. Şimdi iki $H_0' : \hat{\beta}_1 = 0$ ve $H_0'' : \hat{\beta}_2 = 0$ hipotezini göz önüne alalım. Eğer $\left(\hat{\beta}_1, \hat{\beta}_2\right)$ düşey doğrular dışında uzanırsa ilki reddedilecek eğer $\left(\hat{\beta}_1, \hat{\beta}_2\right)$ yatay doğrular dışında uzanır ise, ikincisi reddedilecektir. Bu nedenle, *A* ortak testi başarısızlığa uğratır fakat her iki ayrı testi başarıya ulaştırırken *B* ortak testi başarılı fakat ayrı ayrı testlerin her birini başarısız kılar.



Şekil 3.4 İki Ayrı Teste Karşı Bir Ortak (Birleşik) Test

Bu gereksinim her zaman olmaz, ama bazen olabilir. Bu nedenle bir araştırmacı olarak uygun testin ne olduğuna karar vermeliyiz. Yeri gelmişken belirtelim ki, Şekil 3.4 $\left(\hat{\beta}_1, \hat{\beta}_2\right)$ in sıfır ortalamalı ve $\begin{bmatrix} 1.0 & 0.8 \\ 0.8 & 1.0 \end{bmatrix}$ kovaryans matrisli iki değişkenli normal dağılımlı olduğu kabul edilerek çizilmiştir. 0.8 lik korelasyon elipsin örnek için doğru boyut ve şekilde yapılmasında önemlidir.

3.3.5 Çoklu Korelasyon Katsayısı

$$y = Xb + e = X\hat{b}_{OLS} + e = \hat{y} + e$$

standart lineer modeli verildiğinde, $X'e = 0$ iken

$$y'y = \hat{\beta}'_{OLS} X'X \hat{\beta}_{OLS} + e'e + 2\hat{\beta}'_{OLS} \underbrace{X'e}_{=0} = \hat{y}'\hat{y} + e'e \tag{3.30}$$

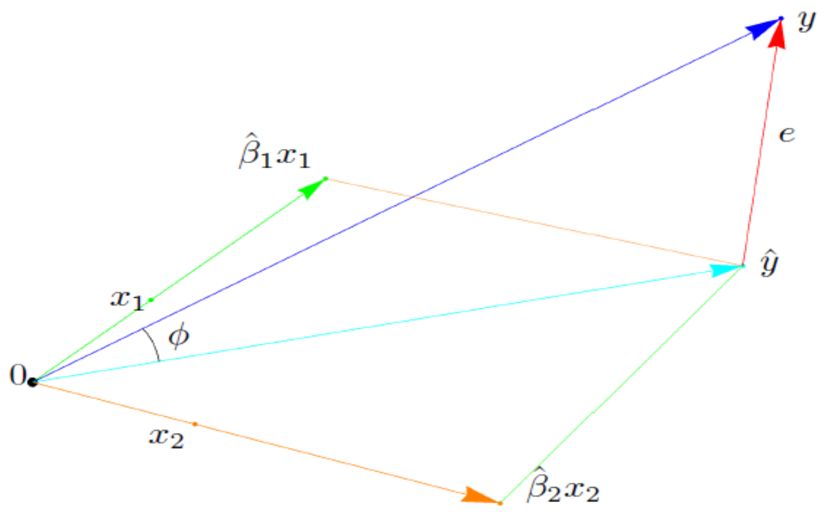
elde edilir. Bu durumda R çoklu korelasyon katsayısı, $y'y$ nin regresyon tarafından “açıklanan kısmının” bir ölçüsüdür. Özellikle,

$$1 - R^2 = \frac{e'e}{y'y}$$

veya

$$R^2 = \frac{\hat{\beta}'_{OLS} X'X \hat{\beta}_{OLS}}{y'y} = \frac{\hat{y}'\hat{y}}{y'y}$$

yazılabilir. Bu durumda (3.30) eşitliğinin önemli sonuçlarından biri, $y'y \geq e'e$ olması ve bu nedenle her zaman $0 \leq R^2 \leq 1$ olması durumudur. R^2 nin geometrik yorumu, onun y ve $\hat{y}$ arasındaki açının kosinüsü olmasıdır. Bu durumda e vektörü $\hat{y} = X\hat{\beta}_{OLS}$ ya ortogonal olduğundan, 0, $\hat{y}$ ve y noktaları $\mathbb{R}^N$ de hipotenüsü $\overline{0y}$ olan bir dik üçgen oluşturur. Bu nedenle R, onlar arasındaki açının kosinüsü olan $\|\hat{y}\|/\|y\|$ oranıdır. Bu durumda bire yakın bir R sıfıra yakın bir açıya karşılık gelir. R^2 niceliğine genellikle sadece "R kare olarak" başvurulur. Yeri gelmişken belirtelim ki, bir regresyon için R değeri hakkında konuşmak yerine R^2 hakkında konuşmak çoğu zaman tercih edilir.



Şekil 3.5 R^2 nin Geometrik Yorumu

Sağ–yan değişkenlerinin artan sayısının sadece hataların karelerinin toplamını azaltabileceğine, bu nedenle onun uyumun ölçüsünü cezalandırmak için istenebildiğine dikkat edelim. Bunu yapmak için bir yöntem ayarlanmış (düzeltilmiş) $\bar{R}^2$ dir. Ayarlanmış R^2,

$$\left(1-\bar{R}^2\right) = \frac{\frac{1}{N-K} e'e}{\frac{1}{N-1} e'e} = \frac{N-1}{N-K}\left(1-R^2\right)$$

veya

$$\bar{R}^2 = \frac{1-K}{N-K} + \frac{N-1}{N-K} R^2$$

ile tanımlanır. Bu durumda ayarlanmış R^2 nin negatif olması mümkündür. $e'e$ yi $(y-\bar{y})'(y-\bar{y})$ yerine niçin $y'y$ ile böldüğümüz merak konusu olabilir. Theil [6, s.164] bunun rotasyona basitlik için olduğunu iddia eder. Fakat eğer regresyon modeli bir sabit terimi içerirse (yani X in sütunlarının biri birlerin bir vektörü ise), bu takdirde ortalama hata (tahmin edilen hata ortalaması) sıfırdır ve bu nedenle, $e'e/(N-K)$; σ^2 nin yansız tahminidir. Ayrıca, regresyon düzlemi, $\bar{x}$ örneklem ortalamalarının vektörü olmak üzere, $\bar{y} = \bar{x}\hat{\beta}_{OLS}$ örneklem ortalaması içinden geçer. Bu nedenle, eğer sabit terim düşürülürse ve tüm değişkenleri onların örneklem ortalamasından sapmalar olarak ifade edilirse, yine β nın aynı tahmini elde edilir. Ancak şimdi çoklu korelasyonun ayarlanmış katsayısı yansız tahmin edilen varyansların bir oranıdır ve bazen ona determinasyon katsayısı denir ve r^2 ile gösterilir. Bu nedenledir ki bu determinasyon katsayısına y lerin örneklem varyansının X ler tarafından açıklanan kısmıdır demeyi anlamlı kılar.

3.3.6 Lineer Modelde Tahmin Aralıkları

Ekseriyetle $X_1,\ldots,X_k$ gözlemlediğimiz değeri için, Y nin değerlerini tahmin etmek isteriz. $x_* = (x_1,\ldots,x_k)$ bağımsız değişkenler için bir vektör olsun. Modelimiz

$$y_* = x_*'\beta + \varepsilon_*$$

olup bu nedenle,

$$\hat{y}_* = x_*'\hat{\beta}_{OLS}$$

tahminini oluşturur. y_* için güven aralığı nedir? Sorusu akla gelebilir. Şimdi

$$\hat{y}_* - y_* = x_*'\hat{\beta}_{OLS} - x_*'\beta - \varepsilon_* = x_*'\left(\hat{\beta}_{OLS} - \beta\right) - \varepsilon_*$$

olacaktır. Bu nedenle,

$$
\begin{aligned}
Var\left(\hat{y}_* - y_*\right) &= Var\left(x_*'\left(\hat{\beta}_{OLS} - \beta\right) - \varepsilon_*\right) \\
&= Var\left(x_*'\left(\hat{\beta}_{OLS} - \beta\right)\right) + Var\left(\varepsilon_*\right) \\
&= \sigma^2\left(x_*'\left(X'X\right)^{-1} x_* + 1\right)
\end{aligned}
$$

elde edilir. Böylece,

$$
\frac{x_*'\hat{\beta}_{OLS} - y_*}{\sqrt{\sigma^2\left(x_*'\left(X'X\right)^{-1} x_* + 1\right)}} \sim N\left(0,1\right) \tag{3.31}
$$

yazılabilir. Bu durumda $s^2 = e'e/\left(N-K\right)$ ve

$$
\frac{\left(N-K\right)s^2}{\sigma^2} \sim \chi^2\left(N-K\right)
$$

olduğunu hatırlayalım. Ayrıca s^2 ve $\hat{\beta}_{OLS}$ bağımsız ε_* örneklemden ε dan bağımsız olduğundan bunun (3.31) oranından bağımsız olduğuna dikkat edelim. Bu nedenle,

$$
\frac{\dfrac{x_*'\hat{\beta}_{OLS} - y_*}{\sqrt{\sigma^2\left(x_*'\left(X'X\right)^{-1} x_* + 1\right)}}}{\dfrac{\left(N-K\right)s^2}{\sigma^2}} = \frac{x_*'\hat{\beta}_{OLS} - y_*}{s\sqrt{\left(x_*'\left(X'X\right)^{-1} x_* + 1\right)}} \sim t\left(N-K\right)
$$

oranı $t(N-K)$ dağılımına sahiptir. Böylece y_* ın bir $1-\alpha$ güven aralığı

$$
\left[\hat{y}_* - t_{\alpha/2,N-K}\, s\sqrt{\left(x_*'\left(X'X\right)^{-1} x_* + 1\right)},\ \hat{y}_* + t_{\alpha/2,N-K}\, s\sqrt{\left(x_*'\left(X'X\right)^{-1} x_* + 1\right)}\right]
$$

olarak ifade edilir.

3.3.7 Ölçüm Hatası

$X_1, \ldots, X_k$ değişkenleri hatalı ölçülürse, bu takdirde önceki sonuçların değiştirilmesi gerekir. $\hat{\beta}_{OLS}$ nin tutarlı ve yansız olduğu sonucu artık doğru değildir. Burada gözlenen X verisinin aslında

$$
X = \tilde{X} + V
$$

olduğunu farz edelim. Burada V ölçüm hatalarının bir matrisidir. Aslında doğru model

$$y = \tilde{X}\beta + \varepsilon$$

veya

$$y = (X - V)\beta + \varepsilon = X\beta + (\varepsilon - V\beta)$$

olduğunda,

$$y = X\beta + u \tag{3.32}$$

modelini tahmin ederiz. (3.32) den elde edilen OLS tahmini

$$\hat{\beta} = \beta + (X'X)^{-1} X'(\varepsilon - V\beta)$$

dır ve buradan beklenen değer

$$E\hat{\beta} = \beta + E(X'X)^{-1} X'V\beta$$

olacaktır. Bu problemi düzeltmenin bir yöntemi, V hataları ve ε ile ilişkisiz olan ve yardımcı değişkenler denen Z rasgele değişkenlerinin bir vektörünü bulmaktır. Bu takdirde bunları $\tilde{X}$ yı tahmin etmek için kullanabiliriz, bundan sonra β yı tahmin etmek için X ler kullanılabilir.

Model denklemlerine göre (modellere bakmak için en anlaşılır yol olan),

$$Y_{ij} = \mu_j + \varepsilon_{ij},\ i = 1,\ldots,n_j;\ j = 1,\ldots,k \tag{3.33}$$

yazılır. Burada k sayıda faktör düzeyi vardır. Y tepkimesinin j düzeyinde n_j tane gözlemine sahibiz ve Y_{ij} tepkimenin j düzeyinde ($j = 1\ldots,k$) i–yinci ölçümüdür. $n = n_1 + \cdots + n_k$ gözlemlerin toplam sayısı olsun. Burada ε_{ij} rasgele değişkenlerinin bağımsız oldukları, sıfır ortalama ve σ^2 ortak varyansına sahip oldukları kabul edilir. Bu nedenle, μ_j sadece tepkimenin j düzeyindeki beklenen değeridir. Analizin amacı $\mu = (\mu_1,\ldots,\mu_k)$ vektörünü tahmin etmek ve onun ile ilgili hipotezleri test etmektir. Bu durumda en yaygın hipotez

H_1: tüm μ_j ler eşit değildir

alternatifine karşı

$$H_0 : \mu_1 = \cdots = \mu_k$$

sıfır hipotezidir. Bu yapının regresyon vasıtasıyla niçin sıfırlandırıldığını görmek için, y gözlemlerinin vektörlerini ve ε_{ij} hatalarını istifleyelim, X_j $j-$düzeyi için bir kukla değişken veya gösterge olsun ve bir X, $n \times k$ matrisini oluşturalım ve böylece (3.33) sistemini

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{n_1 1} \\ y_{12} \\ \vdots \\ y_{n_2 2} \\ \vdots \\ \vdots \\ y_{1k} \\ \vdots \\ y_{n_k k} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{n_1 1} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{n_2 2} \\ \vdots \\ \vdots \\ \varepsilon_{1k} \\ \vdots \\ \varepsilon_{n_k k} \end{bmatrix}$$

olarak tekrar yazalım. Bu model sabit terimsiz bir regresyon modelidir. Eğer bir sabit terim eklenseydi, o diğer sütunların toplamı olacaktı, bu nedenle $X'X$ tekil olacaktı. Bu durumda $X'X$ köşegeni üzerinde n_j ler olan bir köşegen matristir (ve bu nedenle tekil değildir) ve tahmin edilen OLS katsayıları her bir düzey için örneklem ortalamalarıdır. Bu durumda

$$X'X = \begin{bmatrix} n_1 & 0 & \cdots & \cdots & 0 \\ 0 & n_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & n_k \end{bmatrix} \text{ ve } X'y = \begin{bmatrix} \sum_{i=1}^{n_1} y_{i1} \\ \sum_{i=1}^{n_2} y_{i2} \\ \vdots \\ \sum_{i=1}^{n_k} y_{ik} \end{bmatrix}$$

olacaktır ve bu nedenle

$$\left(X'X\right)^{-1}X'y=\begin{bmatrix}\dfrac{\sum_{i=1}^{n_1}y_{i1}}{n_1}\\ \dfrac{\sum_{i=1}^{n_2}y_{i2}}{n_2}\\ \vdots\\ \dfrac{\sum_{i=1}^{n_k}y_{ik}}{n_k}\end{bmatrix}$$

dır. Bu kısımda $(\hat{\mu}_1,\ldots,\hat{\mu}_k)$ katsayı vektörü üzerindeki lineer kısıtlamaları test etmek için bir *F*–istatistiğinin nasıl kullanıldığını gördük. Burada durmamak için bir neden, geleneksel analizde *F*–istatistiğinin daha anlaşılır ve sezgisel olduğudur. Diğer neden ise ANOVA genellikle özel formatta sunulur, bu nedenle onu anlamak önemlidir. Tarihsel açıdan, geleneksel analiz geliştirilmiş olabilir, çünkü o hesapsal olarak OLS den daha az yoğundur. $(X'X)^{-1}$ elde etmek için genel matris tersi *k* sayıda bölme işlemiyle yer değiştirilir.

3.4 Eksik Ranklı Model

Önceki kısımlarda X in $n\times p$ boyutlu ve tam ranklı olduğu, yani $r\left(X\right)=p$ varsayımı dahilinde

$$y=X\beta+\varepsilon$$

lineer modelini göz önüne aldık. Bu varsayım daha kolay bir analiz sağlar, çünkü tam ranklı bir *X* matrisi $X'X$ in tersinir olduğunu ifade eder ve bu nedenle

$$X'Xb=X'y$$

normal denklemi bir tek çözüme sahiptir. Maalesef, tüm lineer modeller bu kategoriye girmez. Eğer öyle olsaydı modeli analiz etmek için başka teknikler geliştirmek zorunda kalırdık.

Eksik ranklı modelin yaygın bir örneği sabit etkilerle tek–yönlü sınıflandırmadır. Bu modelde, örnekler farklı karakteristiklerle *k* bağımsız kitleden gelir. Biz bu kitlelerdeki farklılıkları belirlemek isteriz. Örneğin;

- Bir tıp araştırmacısı üç farklı türde ağrı kesicinin, eklem ağrısını kesmedeki etkisini karşılaştırmak isteyebilir,
- Bir biyolog domates bitkilerini geliştirmek için dört deneysel yaklaşımın etkilerini çalışabilir, veya
- Bir mühendis özel bir coğrafik bölgedeki beş büyük kömür damarlarının içindeki sülfür bileşenini araştırmak isteyebilir.

Genel olarak, kitleler k farklı uygulamanın benzer halk gruplarına uygulanmasının bir sonucu olarak ortaya çıkar. Bu model için karşılık gelen her bir değişkeni hem hangi kitleden alındığını hem de kitledeki örneklem içindeki pozisyonunu belirtmek için iki indekse vereceğiz. Kullandığımız model, $i = 1, 2, \ldots, k$ ve $j = 1, 2, \ldots n_i$ için

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

dir. Burada; k, kitle/etki sayısı ve n_i, $i-$yinci kitledeki örneklem sayısıdır. Bu model tam olarak bir lineer model gibi görünmese de

$$\beta = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_k \end{bmatrix}$$

alarak gayet basitçe

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{21} \\ y_{22} \\ \vdots \\ y_{k,n_k} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_k \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{k,n_k} \end{bmatrix}$$

veya kısaca

$$y = X\beta + \varepsilon$$

formunda yazılabilir. Çünkü X in birinci sütunu, geri kalan sütunların toplamıdır ve sütunlar lineer bağımlıdır, bu yüzden X tam ranklı değildir.

Örnek 3.1. Katran kumundan ve atık sudan organik karbonu çıkarmak için havalı yüzdürme, köpüklü ayrıştırma ve demir klorür katılaşması gibi üç farklı uygulama metodu karşılaştırılmasında çıkarılan karbon miktarları aşağıdaki gibidir:

HY	KA	DKK
34.6	38.8	26.7
35.1	39.0	26.7
35.3	40.1	27.0

Lineer model

$$\underset{y}{\begin{bmatrix} 34.6 \\ 35.1 \\ 35.3 \\ 38.8 \\ 39.0 \\ 40.1 \\ 26.7 \\ 26.7 \\ 27.0 \end{bmatrix}} = \underset{X}{\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}} \underset{\beta}{\begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_k \end{bmatrix}} + \underset{\varepsilon}{\begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{k,n_k} \end{bmatrix}}$$

şeklindendir. Önceden belirtildiği gibi, eksik ranklı modeldeki büyük zorluk bu durumda $X'X$ in tekil olmasıdır. Bu normal denklemlerin bir tek çözüme sahip olmayacağı anlamına gelir. Aslında bunu daha sonra göstereceğiz, şimdi normal denklemler sonsuz sayıda çözüme sahiptir. Bununla birlikte, problem bundan daha derine iner. Parametreleri tahmin edemememiz bir tarafa, parametrelerin kendileri hiçbir sabit değere sahip değildir. Bunu bir örnekte gösterelim.

Örnek 3.2. $k=3$ kitleli bir tek–yönlü sınıflandırma modeline sahip olduğumuzu farz edelim. Her bir kitleden karşılık gelen değişkenler $\mu+\tau_i$ etrafında merkezileşmiştir. Şimdi, bir çalışmadan

$$\mu + \tau_1 = 10$$
$$\mu + \tau_2 = 12$$
$$\mu + \tau_3 = 8$$

bulunduğunu farz edelim. Bu durumda bizim parametrelerimiz $\mu = 10$, $\tau_1 = 0$, $\tau_2 = 2$ ve $\tau_3 = -2$ olabilirdi. Bununla beraber, $\mu = 3$, $\tau_1 = 7$, $\tau_2 = 9$ ve $\tau_3 = 5$ de olabilirdi. Aslında μ yü herhangi bir reel sayı seçebiliriz ve yine de sistemi ifade edebiliriz. Bir bakıma eksik ranklı modeli basit bir anlamla tam ranklı bir modele çevirme olarak ele alabiliriz. Dönüştürülmüş model üzerinde geliştirdiğimiz tüm sistemi kullanabiliriz.

Örnek 3.3. $k = 3$ kitle ile bir–yönlü sınıflandırma modelini düşünelim. Bunun için eksik ranklı model $i = 1, 2, 3$, $j = 1, 2, \ldots, n_i$ için

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

dir. Bununla beraber, her bir kitlenin ortalamasını

$$\mu_i = \mu + \tau_i$$

olarak yazabiliriz. Öyleyse modeli karşılık gelen

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{bmatrix}, \quad \beta = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}$$

matrisleriyle

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

olarak yeniden biçimlendirebiliriz. Şimdi X in sütunlarının lineer bağımsız olduğu aşikardır ve böylece bu bir tam ranklı modeldir. Basit matris hesaplamaları bize

$$X'X = \begin{bmatrix} n_1 & 0 & 0 \\ 0 & n_2 & 0 \\ 0 & 0 & n_3 \end{bmatrix}, \quad (X'X)^{-1} = \begin{bmatrix} \frac{1}{n_1} & 0 & 0 \\ 0 & \frac{1}{n_2} & 0 \\ 0 & 0 & \frac{1}{n_3} \end{bmatrix}$$

ve

$$X'y = \begin{bmatrix} \sum_{i=1}^{n_1} y_{1i} \\ \sum_{i=1}^{n_2} y_{2i} \\ \sum_{i=1}^{n_3} y_{3i} \end{bmatrix}, \quad b = (X'X)^{-1} X'y = \begin{bmatrix} \sum_{i=1}^{n_1} y_{1i} / n_1 \\ \sum_{i=1}^{n_2} y_{2i} / n_2 \\ \sum_{i=1}^{n_3} y_{3i} / n_3 \end{bmatrix}$$

sonuçlarını verir. Bu nedenle, kitlelerin her biri için en küçük kareler tahminleri bu kitleden alınan örneklemin ortalamaları anlamına gelir.

$$\hat{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$$

Bu durumda $t'\beta$ formundaki paramatrelerin lineer fonksiyonları, $t'b$ kullanılarak tahmin edilir. Örneğin $\mu_1 - \mu_2$ fonksiyonu

$$\frac{1}{n_1} \sum_{i=1}^{n_1} y_{1i} - \frac{1}{n_2} \sum_{i=2}^{n_2} y_{2i}$$

ile tahmin edilir. Hataların 0 ortalamalı ve σ^2 varyanslı normal dağılıma sahip olduğu standart varsayımı, tüm kitlelerin ortak bir σ^2 varyansına (fakat farklı ortalamalara) sahip olduğu bağlamında yorumlanır. Bu varyans için standart tahmin edici

$$s^2 = \frac{y'y - y'X(X'X)^{-1} X'y}{n-p} = \frac{y'y - y'Xb}{n-3}$$

biçiminde ifade edilir. O halde

$$s^2 = \frac{1}{n-3}\left[\sum_{i=1}^{3}\sum_{j=1}^{n_i} y_{ij}^2 + \left[\sum_{i=1}^{n_1} y_{1i} \quad \sum_{i=1}^{n_2} y_{2i} \quad \sum_{i=1}^{n_3} y_{3i}\right]\left[\begin{array}{c}\sum_{i=1}^{n_1} y_{1i}/n_1 \\ \sum_{i=1}^{n_2} y_{2i}/n_2 \\ \sum_{i=1}^{n_3} y_{3i}/n_3\end{array}\right]\right]$$

$$= \frac{1}{n-3}\left[\sum_{i=1}^{3}\sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{3}\frac{1}{n_i}\left(\sum_{j=1}^{n_i} y_{ij}\right)^2\right]$$

$$= \frac{1}{n-3}\sum_{i=1}^{3}\left[\sum_{j=1}^{n_i} y_{ij}^2 - \frac{1}{n_i}\left(\sum_{j=1}^{n_i} y_{ij}\right)^2\right].$$

yazılabilir. Bu ise s_i^2 bireysel popülasyon varyans tahmin edicisi olmak üzere

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + (n_3-1)s_3^2}{(n_1-1)+(n_2-1)+(n_3-1)}$$

karma varyans olarak yazılabilir ve öte yandan

$$s_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}\left(y_{ij} - \frac{1}{n_i}\sum_{k=1}^{n_i} y_{ik}\right)^2$$

eşitliğine ulaşılır. Genellikle eksik ranklı bir modeli tam ranklı bir modele yeniden parametrelendirmek mümkündür. Bununla beraber, arzu edilen her zaman bu değildir. Tek yönlü sınıflandırma modeli için, yeniden parametrelendirilmiş kitle ortalamasının güzel bir yorumu vardır. Fakat bu her zaman mümkün değildir.

Örnek 3.4. Faktörlerin her bir kombinasyonundan bir örnekle ve her bir faktörün iki düzeyiyle iki yönlü sınıflandırma modelini (etkileşimsiz) göz önüne alalım.

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, \quad j = 1, 2.$$

Bu modeli daha genel olarak ileride çalışacağız. Bu model için tasarım matrisi

$$X = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

biçimindedir. Birinci sütunun sonrasındaki iki sütunun toplamı olduğu açıktır bundan dolayı X in rankı en çok 4 olur. Bununla beraber, ikinci ve üçüncü sütunların toplamı dördüncü ve beşinci sütunların toplamına eşittir ve böylece $r(X)=3$ tür. Bu ise yorumlanabilirliği daha zor yapan iki parametreyi kaldırmak zorunda olduğumuz anlamına gelir. Neyse ki, modelimizi yeniden parametrelendirmek zorunda değiliz, eksik ranklı modeller için teori geliştirebiliriz. Teorimizin başlangıç noktası (tahmin edilebileceği gibi) daha çok lineer cebirdir. Bu kez koşullu terslerin kavramlarını tanıtacağız.

Tanım 3.1. A $n\times p$ tipinde bir matris olsun. A^c $p\times n$ matrisine A için bir şartlı ters denmesi için gerek ve yeter şart

$$AA^cA=A$$

olmasıdır. Dikkat ettiğmiz ilk şey A nın tekil olmayan ve kare matris olduğudur. Şu halde $A^{-1}=A^c$ dir. Bu yüzden şartlı tersler, alışılmış terslerin kare olmayan ve tekil matrisler için bir genişletilmesidir.

Örnek 3.5.

$$A=\begin{bmatrix}2&4&2\\1&0&-1\\3&1&-2\end{bmatrix},\quad A_1=\begin{bmatrix}0&1&0\\\frac{1}{4}&-\frac{1}{2}&0\\0&0&0\end{bmatrix}$$

matrislerini göz önüne alalım. Bu durumda

$$AA_1A=\begin{bmatrix}2&4&2\\1&0&-1\\3&1&-2\end{bmatrix}\begin{bmatrix}0&1&0\\\frac{1}{4}&-\frac{1}{2}&0\\0&0&0\end{bmatrix}\begin{bmatrix}2&4&2\\1&0&-1\\3&1&-2\end{bmatrix}$$

olacaktır. Bundan dolayı A_1, A için şartlı terstir. Ayrıca

$$A_2=\begin{bmatrix}0&1&0\\0&-3&1\\0&0&0\end{bmatrix}$$

alınırsa

$$AA_2A=\begin{bmatrix}2&4&2\\1&0&-1\\3&1&-2\end{bmatrix}\begin{bmatrix}1&0&-1\\0&1&1\\0&0&0\end{bmatrix}=\begin{bmatrix}2&4&2\\1&0&-1\\3&1&-2\end{bmatrix}=A$$

olarak da gösterilebilir. Buradan A_2 de A için bir şartlı terstir. Bundan yüzden şartlı tersler tek değildir. Bu A nın şartlı tersini konuşma nedenimizdir. Elbette A tekil değildir, o halde şartlı ters eşsiz olarak alışılmış terstir. Bunu yukarıdaki örnekte A nın tekil olduğunu göstermek için kullanabiliriz. Bir kare matrisin alışılmış tersinin olması için, o tekil olmama denen diğer şartları sağlamalıdır. Bununla beraber bu, şartlı terslerin bir durumu değildir.

Teorem 3.2. A, $n\times p$ matris olsun. Bu takdirde A nın şartlı tersi vardır.

İspat. A matrisi r ranklı bir matris olsun. Bu takdirde A üzerinde

$$B=\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

formuna indirgemek için bir dizi elemanter satır ve sütun operasyonları (çarpma, transpoz, ek) gerçekleştirmek mümkündür. Eğer satır ve sütun operasyonlarının matrisleri P ve Q (tekil olmayanlar) olarak ifade edilirse, o zaman

$$PAQ=B$$

yazılabilir. Şimdi

$$B'=\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

$p\times n$ matrisini göz önüne alalım. Burada 0'lar uygun şekilde boyutlandırılmıştır. $BB'B=B$ olduğunu görmek çok zor değildir, bu yüzden B', B nin şartlı tersidir. Öte yandan P ve Q tekil olmadığından, $A=P^{-1}BQ^{-1}$ dir. O halde,

$$\begin{aligned} A\left(QB'P\right)A &= P^{-1}BQ^{-1}QB'PP^{-1}BQ^{-1} \\ &= P^{-1}BB'BQ^{-1} \\ &= P^{-1}BQ^{-1}=A. \end{aligned}$$

yazılabilir. Bu nedenle $QB'P$ matrisi A için bir şartlı terstir ve dolayısıyla A matrisi bir şartlı terse sahiptir.

Bu durumda bir şartlı tersi nasıl buluruz? sorusu akla gelebilir. Yukarıdaki teorem bir yol gösterir fakat bunun daha kolay bir yolu aşağıdaki algoritmadaki gibi verilebilir.

1. A nın tekil olmayan ve $r(A)\times r(A)$ boyutlu bir M minörünü bulunur.
2. M^{-1} ve $\left(M^{-1}\right)'$ matrisleri bulunur.
3. A matrisinde M yerine $\left(M^{-1}\right)'$ yazılıp diğer girdiler sıfırlala yer değiştirilir.
4. Sonuç matrisinin transpozesi alınır.

Örnek 3.6. Yukarıdaki örnekte verilen

$$A=\begin{bmatrix} 2 & 4 & 2 \\ 1 & 0 & -1 \\ 3 & 1 & -2 \end{bmatrix}$$

matrisini göz önüne alalım. Bu durumda $r(A)=2$ olduğu görülebilir. Bu yüzden başlıca

$$M=\begin{bmatrix} 2 & 4 \\ 1 & 0 \end{bmatrix}$$

2×2 minörünü alalım. Buradan

$$\left(M^{-1}\right)'=-\frac{1}{4}\begin{bmatrix} 0 & -4 \\ -1 & 2 \end{bmatrix}'=\begin{bmatrix} 0 & \frac{1}{4} \\ 1 & -\frac{1}{2} \end{bmatrix}$$

ve

$$A^{c}=\begin{bmatrix} 0 & \frac{1}{4} & 0 \\ 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}'=\begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{4} & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

elde edilir. Bu önceki örnekteki A_1 şartlı tersidir, bu yüzden onun işe yaradığı görülmüş olur. Diğer yandan, eğer sol alt 2×2 minörü alırsak devam eden işlem bize A_2 yi verir. Bu yüzden bu işlem şartlı tersten fazlasını üretebilir.

Şartlı ters aşağıdaki özellikleri sağlar. $n\geq p\geq r$ olmak üzere; A matrisi r ranklı $n\times p$ tipinde bir matris olsun. Bu takdirde

- $A^{c}A$ ve AA^{c} idempotenttir,
- $r\left(AA^{c}\right)=r\left(A^{c}A\right)=r$,

- $\left(A^c\right)' = \left(A'\right)^c$,
- $A = A\left(A'A\right)^c\left(A'A\right)$ ve $A' = \left(A'A\right)\left(A'A\right)^c A'$,
- $I - A^c A$ idempotenttir.
- $A\left(A'A\right)^c A'$ tektir, simetriktir ve idempotenttir,
- $r\left(A\left(A'A\right)^c A'\right) = r$,
- $I - A\left(A'A\right)^c A'$ tektir, simetriktir ve idempotenttir,
- $r\left(I - A\left(A'A\right)^c A'\right) = n - r$

dir. Gerçekten örneğin,

$$\begin{aligned}\left[A\left(A'A\right)^c A'\right]' &= A\left[\left(A'A\right)^c\right]'A' = A\left[\left(A'A\right)'\right]^c A' \\ &= A\left(A'A\right)^c A'\end{aligned}$$

$$\begin{aligned}A\left(A'A\right)^c A'A\left(A'A\right)^c A' &= \left[A\left(A'A\right)^c A'A\right]\left(A'A\right)^c A' \\ &= A\left(A'A\right)^c A'\end{aligned}$$

olduğu açıktır. Şimdi normal denklemleri çözme problemine dönelim. Sistemi kurduğumuza göre

$$X'Xb = X'y$$

normal denklemlerini çözmeyi deneyebiliriz. Önce bir çözüme sahip olduklarından emin olmalıyız.

Teorem 3.3. $Ax = g$ sisteminin tutarlı olması için gerek ve yeter şart $\left[A \mid g\right]$ nin rankının A nın rankına eşit olmasıdır.

İspat. $(\Leftarrow)$: $r\left(\left[A \mid g\right]\right) = r\left(A\right)$ olduğunu farz edelim. Yani g yi eklemekle rankın etkilenmediğini varsayalım. Bu durum g nin A nın sütunlarının bir lineer kombinasyonu olduğu anlamına gelmelidir. Bu nedenle

$$x_1a_1 + x_2a_2 + \cdots + x_pa_p = g$$

olacak şekilde hepsi sıfır olmayan $x_1, x_2, \ldots, x_p$ sabitleri vardır. Burada a_i ler A nın *i*–yinci sütunudur. Fakat eğer biz bunu matris notasyonuna çevirirsek

$$x = \begin{bmatrix} x_1 & x_2 & \cdots & x_p \end{bmatrix}'$$

olur ve bu durumda bu tam olarak $Ax = g$ sistemidir. Bu nedenle sistem tutarlıdır.

Teorem 3.4. $y = X\beta + \varepsilon$ bir lineer model olsun. O halde

$$X'Xb = X'y$$

normal denklemleri tutarlıdır.

İspat. $r(X'X) \le r([X'X \mid X'y])$ olduğu açıktır, çünkü bir sütun eklemek lineer bağımsız sütun sayısını azaltamaz. Bununla birlikte önceki rank özelliklerini kullanarak

$$\begin{aligned} r([X'X \mid X'y]) &= r(X'[X \mid y]) \\ &\le r(X') = r(X'X) \end{aligned}$$

olduğu gösterilebilir. Bu nedenle

$$r([X'X \mid X'y]) = r(X'X)$$

dir ve önceki teorem normal denklemlerin tutarlı olduğunu gösterir. Şu halde normal denklemlerin bir çözüme sahip olduğunu biliyoruz, dolayısıyla onu bulmak için şartlı tersleri kullanalım.

Teorem 3.5. $Ax = g$ bir tutarlı sistem olsun. O halde A^c, A nın şartlı tersi olmak üzere $A^c g$ sistem için bir çözümdür.

İspat. $Ax = g$ olduğundan,

$$AA^c g = AA^c Ax = Ax = g$$

dir. Bu nedenle $A^c g$ sistemi çözer. Bu teoremden görürüz ki

$$b = (X'X)^c X'y$$

herhangi şartlı ters için normal denklemleri çözer. Bununla beraber, eksik ranklı modellerde farklı şartlı tersler farklı çözümlere neden olabilir.

Örnek 3.7. Belli bir lineer model için

$$X'X = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad X'y = \begin{bmatrix} 14 \\ 6 \\ 8 \end{bmatrix}$$

alalım. Bu durum potansiyel olarak, her bir sınıftan bir örnekle iki–yönlü sınıflandırma modelinden kaynaklanabilir. Burada

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

dir. O halde normal denklemler

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 14 \\ 6 \\ 8 \end{bmatrix}$$

şeklindedir. $X'X$ in son sütunu ilk iki tanesinin toplamı olduğundan, $X'X$ tam ranklı değildir. Bununla beraber, ilk iki sütun birbirinin katı olmadığından, $r\left(X'X\right) = 2$ dir. $X'X$ in bir şartlı tersini bulmak için, tekil olmayan $\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ minörünü kullanarak algoritmayı uygulayalım. Bu bize

$$\left(X'X\right)^c = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

şartlı tersini verir ve böylece

$$b = \left(X'X\right)^c X'y = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 14 \\ 6 \\ 8 \end{bmatrix} = \begin{bmatrix} 8 \\ -2 \\ 0 \end{bmatrix}$$

elde edilir. Bununla beraber, $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ minörünü kullanmak

$$(X'X)^c = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

verir ki bu da

$$b = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 14 \\ 6 \\ 8 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \\ 8 \end{bmatrix}$$

çözümünü verir. Bu çözümlerin ikisi de normal denklemleri çözer ve aynı ölçüde geçerlidir. Bu eksik ranklı modellerde bir problemdir.

Örnek 3.8. Örnek 3.1 i tekrar göz önüne alalım.

$$X'X = \begin{bmatrix} 9 & 3 & 3 & 3 \\ 3 & 3 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 3 & 0 & 0 & 3 \end{bmatrix}$$

olur ve böylece şartlı ters

$$(X'X)^c = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \end{bmatrix}$$

şeklinde alınabilir. Ayrıca

$$X'y = \begin{bmatrix} 303.3 \\ 105 \\ 117.9 \\ 80.4 \end{bmatrix}$$

matrisini de hesaplayabiliriz. Şartlı tersleri kullanmak bize normal denklemler için

$$b = (X'X)^c X'y = \begin{bmatrix} 0 \\ 35 \\ 39.3 \\ 26.8 \end{bmatrix}$$

gibi bir çözümü verir ama bu tek çözüm değildir. Aslında, eğer model eksik ranklı ise normal denklemler sonsuz sayıda çözüme sahiptir.

Teorem 3.6. $Ax = g$ bir tutarlı sistem olsun. O halde z keyfi bir $p \times 1$ vektör olmak üzere

$$x = A^c g + \left(I - A^c A\right) z$$

sistemi çözer.

İspat. $A^c g$ nin sistemi çözdüğünü biliyoruz ve bu nedenle

$$Ax = A\left[A^c g + \left(I - A^c A\right) z\right]$$

$$= AA^c g + \left(A - AA^c A\right) z$$

$$= g + \left(A - A\right) z = g$$

yazabiliriz. Normal denklemler için bu

$$b = \left(X'X\right)^c X'y + \left[I - \left(X'X\right)^c X'X\right] z$$

formunda herhangi vektör de denklemi sağlar anlamına gelir.

Örnek 3.9. Örnek 3.7 de normal denklemlerin bir çözümü $\begin{bmatrix}8 & -2 & 0\end{bmatrix}'$ idi. İlk bulunan şartlı tersi kullanarak

$$\left(X'X\right)^c X'X = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

elde edilir. Öte yandan keyfi olarak $z = \begin{bmatrix}1 & 1 & 1\end{bmatrix}'$ vektörünü alalım. Bu durumda normal denklemlere diğer bir çözüm

$$b = \begin{bmatrix} 8 \\ -2 \\ 0 \end{bmatrix} + \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \right) \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 1 \\ -1 \end{bmatrix}$$

olarak bulunabilir.

Örnek 3.10. Örnek 3.1 de şartlı invers, bize

$$\left(X'X\right)^c X'X = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

olduğunu verir ve böylece normal denklemlere diğer bir çözüm

$$b = \begin{bmatrix} 0 \\ 35 \\ 39.3 \\ 26.8 \end{bmatrix} + \left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 34 \\ 38.3 \\ 25.8 \end{bmatrix}$$

olarak alınabilir. Yukarıdaki teoremin karşıtı da doğrudur, yani sistemin bütün çözümleri bu formda açıklanabilir.

Teorem 3.7. $Ax = g$ tutarlı bir sistem olsun ve x_0 sistemin herhangi çözümü olsun. Bu durumda $z = x_0$ olmak üzere

$$x_0 = A^c g + \left(I - A^c A\right) z$$

dir.

İspat. x_0 sistemin bir çözümü olduğundan,

$$\begin{aligned} A^c g + \left(I - A^c A\right) z &= A^c g + \left(I - A^c A\right) x_0 \\ &= A^c g + x_0 - A^c A x_0 \\ &= A^c g + x_0 - A^c g = x_0 \end{aligned}$$

yazılabilir. Normal denklemler için bu ise herhangi bir $\left(X'X\right)^c$ şartlı tersi için herhangi bir çözümün

$$b = \left(X'X\right)^c X'y + \left[I - \left(X'X\right)^c X'X\right] z$$

olarak ifade edilebileceği anlamına gelir.

Örnek 3.11. Örnek 3.7 de çözümü orijinal şartlı tersi kullanarak

$$b_1 = \begin{bmatrix} 8 & -2 & 0 \end{bmatrix}'$$

şeklinde bulmuştuk. Fakat ayrıca

$$b_2 = \begin{bmatrix} 0 & 6 & 8 \end{bmatrix}'$$

çözümünü ortaya koyan

$$\left(X'X\right)_2^c = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

şartlı tersini de dikkate aldık. Teoremi kullanarak, ilk çözüm, ikinci çözüm bakımından

$$\begin{aligned} b_1 &= \left(X'X\right)_2^c X'y + \left(I - \left(X'X\right)_2^c X'X\right)z \\ &= \begin{bmatrix} 0 \\ 6 \\ 8 \end{bmatrix} + \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} 8 \\ -2 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 6 \\ 8 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 8 \\ -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 8 \\ -2 \\ 0 \end{bmatrix} \end{aligned}$$

olarak yazılabilir.

Şimdi normal denklemleri nasıl çözeceğimizi biliyoruz, dahası onun için tüm çözümleri nasıl bulacağımızı da biliyoruz. Fakat hangi çözümü istiyoruz? Hangisi en iyisi? Bu soruların yanıtı hepsinin aynı ölçüde geçerli olduğudur. Fakat bu ise parametreleri asla tahmin edemeyeceğimiz anlamına gelir.

Bununla beraber en azından keyfi olmayan bir şey var. O da tepkime değişkeni y nin değeridir. Hangi parametrenin tahmin edileceği fark etmez, y asla değişmeyecektir. Aslında, normal denklemler için hangi çözümü kullandığımız fark etmeksizin parametrelerin daima aynı değerde tahmin edilebilecek lineer kombinasyonları vardır. Bu lineer kombinasyonlara tahmin edilebilir deriz. Tahmin edebileceğimiz üzere, tahmin edilebilir kombinasyonlar bir nevi tepkime değişkeniyle ilişkilendirilebilir.

Tanım 3.2. $y = X\beta + \varepsilon$ bir lineer model olsun. Eğer $E\left[c'y\right] = t'\beta$ olacak şekilde bir c vektörü varsa, bir $t'\beta$ fonksiyonuna tahmin edilebilirdir denir. Ona bakmanın diğer bir yolu bir $t'\beta$ lineer yansız tahmin edici olmasıdır. Şimdi, tahmin edilebilirlik için bazı eşdeğer şartlara bakalım.

Teorem 3.8. ε 0 ortalamalı ve $\sigma^2 I$ varyanslı olmak üzere, $y = X\beta + \varepsilon$ bir lineer model olsun. O halde, $t'\beta$ nın tahmin edilebilir olması için gerek ve yeter şart $X'Xz = t$ lineer sisteminin bir çözümünün mevcut olmasıdır.

İspat. $(\Leftarrow)$ z_0 $X'Xz = t$ nin bir çözümü olsun ve $c = Xz_0$ koyalım. Buradan

$$E(c'y) = E(z_0'X'y) = z_0'X'E(y) = z_0'X'X\beta = t'\beta$$

olup, $t'\beta$ tahmin edilebilirdir.

Örnek 3.12. İki–sınıf örneğimizi göz önüne alalım. Bu durumda

$$X'X = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

elde etmiştik. Öte yandan $\beta_1 - \beta_2$ parametrelerinin kombinasyonunu düşünelim. Bu

$$t = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix}'$$

olmak üzere $t'\beta$ ya karşılık gelir. Şimdi

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

sistemi için bir çözüm arıyoruz. Basit bir düşünce gösterir ki çözüm $z_1 = 0$, $z_2 = 1$, $z_3 = -1$ dir ve bundan dolayı $\beta_1 - \beta_2$ tahmin edilebilirdir.

Teorem 3.9. ε 0 ortalamalı ve $\sigma^2 I$ varyanslı olmak üzere, $y = X\beta + \varepsilon$ bir lineer model olsun. O halde, $t'\beta$ nin tahmin edilebilir olması için gerek ve yeter şart $X'X$ in herhangi şartlı tersi için $t'(X'X)^c X'X = t'$ olmasıdır.

İspat. $(\Leftarrow)$ $t'(X'X)^c X'X = t'$ olduğunu farz edelim. Böylece

$$X'X((X'X)^c)'t = X'X(X'X)^c t = t$$

yazılabilir. Bu ise $(X'X)^c t$ nin $X'Xz = t$ sisteminin bir çözümü olduğu anlamına gelir ve bu nedenle bir önceki teoremden $t'\beta$ nın tahmin edilebilir olduğu görülür.

$(\Rightarrow)$ $t'\beta$ nın tahmin edilebilir olduğunu farz edelim. Bir önceki teoremden, $X'Xz = t$ sisteminin bir çözümü vardır. Şartlı tersi kullanarak bir çözümün $z = (X'X)^c t$ olduğunu biliyoruz. Diğer bir deyişle,

$$X'X(X'X)^c t = t$$

dir ve yukarıdaki gibi transpozlar alarak, $t'(X'X)^c X'X = t'$ olduğu görülür.

Örnek 3.13. Burada yine önceki örneği göz önüne alalım. Bunun için

$$(X'X)^c = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

şartlı tersini alalım ve $t = [0 \quad 1 \quad -1]'$ a karşılık gelen aynı $\beta_1 - \beta_2$ eşitliğini göz önüne alalım. Bu takdirde

$$\begin{aligned} t'\left(X'X\right)^c X'X &= [0 \quad 1 \quad -1]\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \\ &= [0 \quad 1 \quad -1]\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \\ &= [0 \quad 1 \quad -1] = t' \end{aligned}$$

olduğu görülür ve bu nedenle yine $\beta_1 - \beta_2$ nin tahmin edilebilir olduğu görülür. Diğer yandan $t'\beta = \beta_0$ olacak şekilde $t = [1 \quad 0 \quad 0]'$ vektörünü aldığımızı farz edelim. O halde

$$t'\left(X'X\right)^c X'X = [1 \quad 0 \quad 0]\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

$$= [1 \quad 0 \quad 0]\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$= [1 \quad 0 \quad 1] \neq t'$$

elde edilir, böylece β_0 tahmin edilebilir değildir.

Örnek 3.14. Karbon ayırma örneğine (Örnek 3.1) geri dönelim. Çeşitli karbon ayırma uygulamalarının önemli ölçüde farklı anlamlara geldiğini görmeye ilgi duyuyoruz. Bunu test etmek için, $\tau_1 - \tau_2$ ve $\tau_1 - \tau_3$ niceliklerine bakalım. Eğer bunların ikisi de 0 ise, o zaman uygulamalar aynıdır. Bu durumda

$$\left(X'X\right)^c X'X = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

elde edilir ve katsayı vektörleri

$$t_1 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} \text{ ve } t_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}$$

şeklindedir. Buradan

$$t_1'\left(X'X\right)^c X'X = \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix}$$

olacaktır ve bundan dolayı $t_1'\beta = \tau_1 - \tau_2$ tahmin edilebilirdir. Öte yandan

$$t_1'\left(X'X\right)^c X'X = \begin{bmatrix} 0 & 1 & 0 & -1 \end{bmatrix}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & -1 \end{bmatrix}$$

olur ve bundan dolayı $t_2'\beta = \tau_1 - \tau_3$ de tahmin edilebilirdir. Tahmin edilebilirlik tanımını kullanarak hangi şartlı tersi kullanırsak kullanalım tahmin edilebilir bir nicelik için aynı tahmini üreteceğimizi ispat edebiliriz. Bunu ilk olarak bir yardımcı lemma ifade edelim.

Lemma 3.2. ε 0 ortalamalı ve $\sigma^2 I$ varyanslı olmak üzere, $y = X\beta + \varepsilon$ bir lineer model olsun. Herhangi bir $t'\beta$ tahmin edilebilir fonksiyonu için en iyi lineer yansız tahmin edici, z $X'Xz = t$ sisteminin çözümü olmak üzere $z'X'y$ dir.

Teorem 3.10. (Bir Gauss-Markov Teoremi): ε 0 ortalamalı ve $\sigma^2 I$ varyanslı olmak üzere, $y = X\beta + \varepsilon$ bir lineer model olsun. $t'\beta$ nın tahmin edilebilir olduğunu farz edelim. O halde, $X'Xz = t$ sisteminin herhangi çözümü $t'\beta$ için aynı tahmini verir. Dahası, bu tahmin b normal denklemlerin herhangi çözümü olmak üzere $t'b$ şeklindedir. Sonuç olarak bu tahmin bir en iyi lineer yansız tahmin edicidir.

İspat. $X'Xz = t$ sistemi için z_0 ve z_1 olarak iki çözüme sahip olduğumuzu farz edelim. Ayrıca *b*,

$$X'Xb = X'y$$

normal denklemlerinin herhangi çözümü olsun. Bir önceki lemmadan, $t'\beta$ nın en iyi lineer yansız tahmin edicisi

$$z_0'X'y = z_0'X'Xb = \left(X'Xz_0\right)'b = t'b$$

olacaktır. Benzer olarak,

$$z_1'X'y = t'b = z_0'X'y$$

yazılabilir. Bu, en iyi lineer yansız tahmin edicinin tek ve $t'b$ ye eşit olduğunu gösterir.

Örnek 3.15. İki–sınıf örneğine yeniden bakalım. Daha önce $\beta_1 - \beta_2$ nin tahmin edilebilir olduğunu göstermiştik. Ayrıca normal denklemlerin çözümünün

$$b = \begin{bmatrix} 8 & -2 & 0 \end{bmatrix}', \; b_0 = \begin{bmatrix} 0 & 6 & 8 \end{bmatrix}'$$

vektörlerini içerdiğini de biliyoruz. Eğer $\beta_1 - \beta_2$ yi tahmin etmek istersek bu durumda

$$t'b = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 8 \\ -2 \\ 0 \end{bmatrix} = -2$$

yi kullanırız. Bununla beraber, Teorem 3.10 dan

$$t'b_0 = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 6 \\ 8 \end{bmatrix} = -2$$

yi de kullanabiliriz. Bu tahminin bir öncekiyle aynı olması bir rastlantı değildir. Teorem gösterir ki, normal denklemlerin herhangi bir şartlı ters kullanılarak elde edilen herhangi bir çözümü, tam olarak aynı tahmini ortaya koyar. Diğer bir deyişle, tahmin edici tektir.

Örnek 3.16. Örnek 3.1 e bakalım. Daha önce $\tau_1 - \tau_2$ ve $\tau_1 - \tau_3$ ün tahmin edilebilir olduğunu göstermiştik. Onları sırasıyla

$$t_1'b = \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 35 & 39.3 & 26.8 \end{bmatrix}' = -4.3$$

ve

$$t_2'b = \begin{bmatrix} 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 35 & 39.3 & 26.8 \end{bmatrix}' = 8.2$$

vasıtasıyla tahmin edebiliriz. Gauss-Markov teoremi gösterir ki b yi hesaplamak için hangi şartlı tersi kullanırsak kullanalım, bu tahminler hep aynı kalacaktır. Tahmin edilebilirliği tanımladığımıza göre hangi niceliklerin tahmin edilebilir olduğunu ve hangilerinin olmadığını bilmek de isteriz. Kesinlikle tahmin edilebilir olan ilk nicelikler y nin bileşenleridir çünkü neticede bu tahmin edilebilirliği nasıl tanımladığımızdır.

Teorem 3.11. $y = X\beta + \varepsilon$ bir lineer model olsun. Bu takdirde, $X\beta$ nın bileşenleri tahmin edilebilirdir.

İspat. $E[y] = X\beta$ olduğunu biliyoruz. Bu nedenle, tahmin edilebilir fonksiyonları elde etmek için $X\beta$ yı

$$[1 \quad 0 \quad \cdots \quad 0], [0 \quad 1 \quad \cdots \quad 0], \cdots, [0 \quad 0 \quad \cdots \quad 1]$$

vektörlerinin her biriyle çarpabiliriz. Fakat bunlar $X\beta$ nın bileşenleridir, bu yüzden $X\beta$ nın bileşenleri tahmin edilebilirdir.

Örnek 3.17. Örnek 3.1 i göz önüne alalım.

$$X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}', \quad \beta = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}$$

vektörlerine sahibiz. β parametre vektörünü tahmin edemeyeceğimizi daha önce göstermiştik. Bununla beraber, bu modeldeki etkinin gerçek miktarları, üç uygulamadan tepkilerin ortalamasıdır. Bunlar $\mu + \tau_1$, $\mu + \tau_2$ ve $\mu + \tau_3$ şeklindedir. Bu durumda

$$\begin{aligned} \mu + \tau_1 &= [1 \quad 1 \quad 0 \quad 0]\beta \\ \mu + \tau_2 &= [1 \quad 0 \quad 1 \quad 0]\beta \\ \mu + \tau_3 &= [1 \quad 0 \quad 0 \quad 1]\beta \end{aligned}$$

olduğunu ve bunların her birinin $X\beta$ nın bileşenleri olduğunu görebiliriz. Bu nedenle, onlar tahmin edilebilirdir. Bunları b normal denklemlerin herhangi çözümü olmak üzere, β yı b ile yer değiştirerek tahmin edeceğiz. Aslında, herhangi bir k ile bir sınıflandırma modelinde $\mu + \tau_i$ her zaman tahmin edilebilirdir. $X\beta$ nın elemanlarının tahmin edilebilir olduğunu biliyoruz. Bundan başka eğer tahmin edilebilir nicelikleri lineer bir biçimde kombine edersek, sonuç tahmin edilebilir olmalıdır.

Teorem 3.12. $t_1'\beta, t_2'\beta, \ldots, t_k'\beta$ tahmin edilebilir fonksiyonlar olsun ve

$$z = a_1 t_1'\beta + a_2 t_2'\beta + \cdots + a_k t_k'\beta$$

alalım. Bu takdirde *z* tahmin edilebilirdir ve *z* için en iyi lineer yansız tahmin edici

$$a_1t_1'b + a_2t_2'b + \cdots + a_kt_k'b$$

dir.

İspat. Tanımdan

$$z = (a_1t_1 + a_2t_2 + \cdots + a_kt_k)'\beta$$

yazılabilir. Bu durumda tüm fonksiyonlar tahmin edilebilir olduğu için,

$$\begin{aligned}
&(a_1t_1 + a_2t_2 + \cdots + a_kt_k)'(X'X)^c X'X \\
&= a_1t_1'(X'X)^c X'X + a_2t_2'(X'X)^c X'X + \cdots + a_kt_k'(X'X)^c X'X \\
&= (a_1t_1 + a_2t_2 + \cdots + a_kt_k)'
\end{aligned}$$

dir. Bu nedenle

$$z = (a_1t_1 + a_2t_2 + \cdots + a_kt_k)'b$$

tahmin edicisiyle z tahmin edilebilirdir. Birçok çalışmadaki özel ilgi çeken şey, farklı kitleleri birbirine karşı kıyaslamanın yoludur. Bu karşılaştırmalara sayısal bir değer ermek için $\sum_{i=1}^{k} a_i = 0$ olmak üzere $a_1\tau_1 + a_2\tau_2 + \cdots + a_k\tau_k$ lineer kombinasyonları ortaya koyulur. Bu uygulama zıtlıkları, kitleler arasındaki farklılıkların daha iyi bir tanımlamasını elde etmek amacıyla tüm tepkime ortalamalarının etkilerini ortadan kaldırır. Bir tek–yönlü sınıflandırma modelinde, herhangi bir uygulama zıtlığı tahmin edilebilirdir. Bunu, eğer

$$z = a_1\tau_1 + a_2\tau_2 + \cdots + a_k\tau_k$$

bir uygulama zıtlığı ise, bu takdirde

$$\begin{aligned}
z &= \sum_{i=1}^{k} a_i\mu + a_1\tau_1 + a_2\tau_2 + \cdots + a_k\tau_k \\
&= a_1(\mu + \tau_1) + a_2(\mu + \tau_2) + \cdots + a_k(\mu + \tau_k)
\end{aligned}$$

nın, $\mu + \tau_i$ tahmin edilebilir fonksiyonlarının bir lineer kombinasyonları ve bu yüzden kendi kendilerine tahmin edilebilir olduklarına dikkat ederek göstereceğiz. Uygulama zıtlıkları arasındaki özel ilgi çeken şey bazı $i \neq j$ için $\tau_i - \tau_j$ formlarının zıtlıklarıdır. Çünkü

$$\tau_i - \tau_j = (\mu + \tau_i) - (\mu + \tau_j)$$

olup, bu i popülasyonundaki tepkime ortalaması ve j popülasyonundaki tepkime ortalaması arasındaki farklılıktır. Eğer i popülasyonundan örneklem ortalaması için $\bar{y}_i$ yazarsak, bu takdirde bu zıtlıkları, $\bar{y}_i - \bar{y}_j$ örneklem ortalamalarındaki farklılıklara karşılık getirerek tahmin etmeyi beklerdik. Durumun gerçeğinde olan geliştirmiş olduğumuz teoriyi kullanarak gösterebiliriz. Bunu $k = 3$ ve $\tau_1 - \tau_2$ farkı için yapalım. Bu durumda matrislerimiz

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad y = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n_2} \\ y_{31} \\ y_{32} \\ \vdots \\ y_{3n_3} \end{bmatrix}, \quad \beta = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}$$

olduğundan direk çarpım yaparak

$$X'y = \begin{bmatrix} \sum_{i=1}^{3} \sum_{j=1}^{n_j} y_{ij} \\ \sum_{j=1} y_{1j} \\ \sum_{j=1} y_{2j} \\ \sum_{j=1} y_{3j} \end{bmatrix}, \quad X'X = \begin{bmatrix} n & n_1 & n_2 & n_3 \\ n_1 & n_1 & 0 & 0 \\ n_2 & 0 & n_2 & 0 \\ n_3 & 0 & 0 & n_3 \end{bmatrix}$$

elde edilir. Şartlı ters algoritmasını $X'X$ in sağ alt köşesinde uygulayarak

$$(X'X)^c = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{n_1} & 0 & 0 \\ 0 & 0 & \frac{1}{n_2} & 0 \\ 0 & 0 & 0 & \frac{1}{n_3} \end{bmatrix}$$

elde etmek için kullanabiliriz. Bu nedenle normal denklemlerin bir çözümü

$$b=\left(X'X\right)^c X'y=\begin{bmatrix}0 & \bar{y}_1 & \bar{y}_2 & \bar{y}_3\end{bmatrix}'$$

şeklindedir. Öte yandan $\tau_1-\tau_2$ yi $\begin{bmatrix}0 & 1 & -1 & 0\end{bmatrix}\beta$ olarak yazabiliriz ve böylece $\tau_1-\tau_2$ nin en iyi lineer yansız tahmin edicisi

$$\begin{bmatrix}0 & 1 & -1 & 0\end{bmatrix}\begin{bmatrix}0 & \bar{y}_1 & \bar{y}_2 & \bar{y}_3\end{bmatrix}'=\bar{y}_1-\bar{y}_2$$

olur. Eğer diğer herhangi bir şartlı tersi alsaydık, aynı sonucu elde edecektik. Karbon ayırma örneğinde $\tau_1-\tau_2$ ve $\tau_1-\tau_3$ ün tahmin edilebilir olduğunu gösterdik. Bunların ikisi de zıtlıklardır, bu yüzden tereddüt etmeden onlar tahmin edilebilirdir diyebiliriz.

Şimdi eksik ranklı modelde σ^2 yi tahmin etme problemine dönelim. Tam ranklı modelde σ^2 yi *n* örneklem büyüklüğü, *p* parametrelerin sayısı ve SS_{Res} hata kareler toplamı olmak üzere daha önce $s^2=SS_{Res}/n-p$ ile tahmin ettik. Burada hata kareler toplamı

$$SS_{Res}=\left(y-Xb\right)'\left(y-Xb\right)=y'\left[I-X\left(X'X\right)^{-1}X'\right]y$$

dir. Şimdi eksik ranklı modeller için karşılık gelen açıklamayı bulmak istiyoruz, fakat burada açıkça görülüyor ki $\left(X'X\right)^{-1}$ mevcut olmadığından o aynısı olmayacaktır. Hata kareler toplamını halen

$$SS_{Res}=\left(y-Xb\right)'\left(y-Xb\right)$$

olarak tanımlayalım. Burada b normal denklemlerin herhangi bir çözümüdür. Önemli olan şey b değişebiliyorken Xb nin değişmemesidir, çünkü $X\beta$ nın bileşenleri tahmin edilebilirdir. Bu nedenle SS_{Res}, *b* nin seçimine göre değişmezdir. Şimdi, SS_{Res} için eşdeğer açıklama verebiliriz.

Teorem 3.13. $SS_{Res}=y'\left[I-X\left(X'X\right)^c X'\right]y$ dir.

İspat. $b=\left(X'X\right)^c X'y$ olsun ve $X\left(X'X\right)^c X'X=X$ olduğunu hatırlayalım. Bu takdirde

$$
\begin{aligned}
SS_{Res} &= \left(y' - b'X'\right)\left(y - Xb\right) \\
&= y'y - 2y'Xb + b'X'Xb \\
&= y'y - 2y'X\left(X'X\right)^{c}X'y + y'X\left(X'X\right)^{c}X'X\left(X'X\right)^{c}X'y \\
&= y'y - 2y'X\left(X'X\right)^{c}X'y + y'X\left(X'X\right)^{c}X'y \\
&= y'\left[I - X\left(X'X\right)^{c}X'\right]y
\end{aligned}
$$

olarak yazılabilir. Buradan σ^2 için bir tahmin ediciyi bulmak için daha önce geliştirilen kuadratik formlar teorisini kullanarak,

$$
\begin{aligned}
E\left[SS_{Res}\right] = E&\left[y'\left(I - X\left(X'X\right)^{c}X'\right)y\right] \\
&= iz\left(I - X\left(X'X\right)^{c}X'\right)\sigma^2 + \left(X\beta\right)'\left(I - X\left(X'X\right)^{c}X'\right)X\beta \\
&= iz\left(I - X\left(X'X\right)^{c}X'\right)\sigma^2 + \beta'X'X\beta - \beta'X'X\left(X'X\right)^{c}X'X\beta \\
&= iz\left(I - X\left(X'X\right)^{c}X'\right)\sigma^2 + \beta'X'X\beta - \beta'X'X\beta \\
&= iz\left(I - X\left(X'X\right)^{c}X'\right)\sigma^2
\end{aligned}
$$

olduğu görülür. Öte yandan $I - X\left(X'X\right)^{c}X'$ matrisinin simetrik ve idempotent olduğu gösterilebilir ve böylece *X* in rankı $r = r\left(X\right)$ olmak üzere,

$$
E\left[SS_{Res}\right] = r\left(I - X\left(X'X\right)^{c}X'\right)\sigma^2 = \left(n - r\right)\sigma^2
$$

dir. Bu bize aşağıdaki teoremi verir.

Teorem 3.14. *X*, *r* ranklı ve ε, 0 ortalamalı ve $\sigma^2 I$ varyanslı olmak üzere $y = X\beta + \varepsilon$ bir lineer model olsun. Bu takdirde σ^2 için bir yansız tahmin edici $SS_{Res}/n - r$ biçimindedir.

Örnek 3.18. Karbon ayırma örneğine (Örnek 3.1) geri dönelim. Uygun değerler

$$Xb = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 35 \\ 39.3 \\ 26.8 \end{bmatrix} = \begin{bmatrix} 35 \\ 35 \\ 35 \\ 39.3 \\ 39.3 \\ 39.3 \\ 26.8 \\ 26.8 \\ 26.8 \end{bmatrix}$$

dir. Böylece hatalar

$$y - Xb = \begin{bmatrix} 34.6 \\ 35.1 \\ 35.3 \\ 38.8 \\ 39.0 \\ 40.1 \\ 26.7 \\ 26.7 \\ 27.0 \end{bmatrix} - \begin{bmatrix} 35 \\ 35 \\ 35 \\ 39.3 \\ 39.3 \\ 39.3 \\ 26.8 \\ 26.8 \\ 26.8 \end{bmatrix} = \begin{bmatrix} -0.4 \\ 0.1 \\ 0.3 \\ -0.5 \\ -0.3 \\ 0.8 \\ -0.1 \\ -0.1 \\ 0.2 \end{bmatrix}$$

dir. Bu ise $SS_{Res} = (y - Xb)'(y - Xb) = 1.3$ anlamına gelir. Öte yandan X in rankının 3 olduğu kolayca görülebilir ve böylece

$$s^2 = 1.3/(9-3) = 0.217$$

olarak hesaplanır.

Şimdi de eksik ranklı modelde aralık tahmini problemini ele alalım. Tam ranklı modelde, neyi tahmin edebileceğimizi tahmin etmiştik. Sıradaki adım, tahminlerimiz için güven aralıklarını denemek ve bulmaktır. Şimdiye kadar ε hata vektörünün normal dağıldığını kabul etmemiştik. Halbuki, güven aralıklarını bulmak için değişkenlerin dağılımları hakkında biraz fikre ihtiyacımız vardır, böylece o kabulü şimdi yapacağız. Tam rank modelde, bir normal değişkeni bir χ^2 değişkeniyle bölerek oluşturulan bir t–dağılımı niceliği bularak güven aralıklarını oluşturduk. χ^2 değişkeni $n-p$ serbestlik derecesine sahip SS_{Res}/σ^2 şeklindeydi.

σ^2 terimi bilinmiyordu, fakat paydaki sıfırlanmış başka bir σ^2 terimi hesaplayabileceğimiz bir şeyler bırakıyordu. Büyük ölçüde aynı şeyi eksik ranklı model için yapabiliriz.

Teorem 3.15. ε, 0 ortalamalı ve $\sigma^2 I$ varyanslı rasgele bir normal vektör olmak üzere $y = X\beta + \varepsilon$ bir lineer model olsun. Bu takdirde

$$\frac{(n-r)s^2}{\sigma^2} = \frac{SS_{Res}}{\sigma^2}$$

bir χ^2 dağılımına ve $n-r$ serbestlik derecesine sahiptir. Bu teoremin ispatı tıpkı tam rank model durumunda olduğu gibidir, bu yüzden onu tekrar etmeyeceğiz. Bir güven aralığı elde etmek için adımlar tam rank durumundakine çok benzerdir, fakat iki küçük farklılık vardır. İlk olarak, sadece tahmin edilebilir nicelikler için güven aralıklarını bulabiliriz. İkincil olarak, $(X'X)^{-1}$ tersini, $(X'X)^c$ şartlı ters ile yer değiştiririz. Diğer bütün adımlar aynıdır. Bu bize t dağılımını kullanarak, $n-r$ serbestlik dereceli $t'\beta$ tahmin edilebilir nicelikleri için

$$t'b \pm t_{\alpha/2} s \sqrt{t'(X'X)^c t}$$

güven aralığını verir. Bu formül tahmin edilebilir olması koşuluyla özel parametreler için güven aralıklarını bulmak için de kullanılabilir.

Örnek 3.19. Yine karbon ayırma örneğine (Örnek 3.1) dönelim. $\tau_1 - \tau_2$ için bir 95% güven aralığı bulmak istediğimizi farz edelim. $n-r = 9-3 = 6$ serbestlik derecesini kullanarak

$$t = [0 \quad 1 \quad -1 \quad 0]', \; s^2 = 0.217, \; t_{0.025} = 2.45$$

elde edilir. Ayrıca

$$(X'X)^c = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \end{bmatrix}$$

şartlı tersini kullanalım. Bu ise

$$t'b \pm t_{\alpha/2} s \sqrt{t'(X'X)^c t}$$

$$= \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 35 \\ 39.3 \\ 26.8 \end{bmatrix} \pm 2.45 \times \sqrt{0.217} \times \sqrt{\begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}}$$

$$= -4.3 \pm 0.93 = (-5.23, -3.37).$$

güven aralığını verir. Özellikle 95% güven aralığıyla ilk karbon ayırma uygulamasının, ikincisi kadar etkili olmadığı söylenebilir.

Örnek 3.20. Genel bir üç–yönlü sınıflandırma modelinde $\tau_1 - \tau_2$ zıtlığının, ayrı ayrı $\bar{y}_1 - \bar{y}_2$ kitle ortalamalarındaki fark ile tahmin edilebileceği daha önce gösterilmişti. Ayrıca

$$t = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \quad (X'X)^c = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{n_1} & 0 & 0 \\ 0 & 0 & \frac{1}{n_2} & 0 \\ 0 & 0 & 0 & \frac{1}{n_3} \end{bmatrix}$$

olduğu görülür. Bundan dolayı

$$t'(X'X)^c t = \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{n_1} & 0 & 0 \\ 0 & 0 & \frac{1}{n_2} & 0 \\ 0 & 0 & 0 & \frac{1}{n_3} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} = \frac{1}{n_1} + \frac{1}{n_2}$$

elde edilir ve bu durumda güven aralığı

$$\bar{y}_1 - \bar{y}_2 \pm t_{\alpha/2} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

olur. Eğer $X_{n \times p}$ matrisi $n < p$ rankına sahip ise, normal denklemlere tek bir $\hat{\beta}$ çözümü yoktur. Bir $\hat{\beta}$ çözümünü bulmak için üç yöntem ve $\hat{Y}$ ortogonal iz düşümü vardır.

1. Bir $(X'X)^-$ genelleştirilmiş tersi kullanılır.

2. Model, tam ranklı modele indirgenir.
3. Tahmin edilebilirlik kısıtlamaları koyulur.

Yöntem 1. Bir $\left(X'X\right)^{-}$ genelleştirilmiş tersini kullanalım.

Tanım 3.3. Eğer $AA^{-}A = A$ ve $A^{-}AA^{-} = A^{-}$ ise, A^{-} bir yansımalı g–terstir.

Lemma 3.3. $\hat{\beta} = \left(X'X\right)^{-} X'Y$ ifadesi $X'Y = X'X\beta$ normal denklemlerine bir çözümdür.

Lemma 3.4. Y nin $R(X)$ üzerine ortogonal izdüşümü, $\left(X'X\right)^{-}$ herhangi bir genelleştirilmiş tersi için, $P = X\left(X'X\right)^{-} X'$ olmak üzere, PY ile verilir.

Normal denklemleri

$$\begin{aligned}
X'Y &= X'X\hat{\beta} \\
&= \overbrace{X'X\left(X'X\right)^{-} X'X}^{=X'X}\hat{\beta} \\
&= X'X\left(X'X\right)^{-} X'X\hat{\beta} \\
&= X'X\left(X'X\right)^{-} X'Y \quad \text{(1. satırdan)}
\end{aligned}$$

olarak tekrar yazalım. Bu nedenle, $\hat{\beta} = \left(X'X\right)^{-} X'Y$ normal denklemlere bir çözümdür.

Lemma 3.5. Eğer $PX'X = QX'X$ ise, bu takdirde $PX' = QX'$ dür.

İspat. Herhangi bir A matrisi için eğer $AA' = 0$ ise bu takdirde $A = 0$ dır.

$$\begin{aligned}
\left(PX' - QX'\right)\left(PX' - QX'\right)' &= \left(PX' - QX'\right)X\left(P - Q\right)' \\
&= \left(PX'X - QX'X\right)\left(P - Q\right)' = 0\left(P - Q\right)' = 0
\end{aligned}$$

dır ve bu nedenle $PX' - QX' = 0$ dır.

Teorem 3.16. Eğer G ; $X'X$ in bir g–tersi ise, bu takdirde

1. G' de $X'X$ in bir g–tersidir. Gerçekten $(X'XG'X'X)' = X'XGX'X = X'X$ dir. Buradan her iki tarafın tranpozesini alarak istenen sonuç elde edilir.

2. $GX'XG'$; $X'X$ in bir simetrik yansımalı g–tersidir. Burada simetri açıktır. Öte yandan $GX'XG'$ ifadesinin $X'X$ matrisinin bir g–tersi olduğunu göstermek için

$$X'X\left(GX'XG'\right)X'X=\left(X'XGX'X\right)G'X'X=\left(X'X\right)G'X'X$$

elde edilir ki bu, yukarıdaki (1) i kullanarak $X'X$ e eşittir. Bu nedenle $X'X$ in $GX'XG'$ nün bir g–tersi olduğunu göstermek için,

$$\begin{aligned}GX'XG'\left(X'X\right)GX'XG' &= GX'XG'\left(X'XGX'X\right)G'\\ &= GX'XG'\left(X'X\right)G'=G\left(X'XG'X'X\right)G'\end{aligned}$$

elde edilir ki bu, yine yukardaki (1) i kullanarak $GX'XG'$ ye eşittir.

3. (a) $X'XGX'=X'$ ve (b) $XGX'X=X$ dir. Gerçekten $X'XGX'X=IX'X$ dir ve bu nedenle, $X'XGX'=IX'$ Lemma 3.5 ten (a) gösterilmiş olur. (b) yi ispatlamak için, (a) ve yukardaki (1) i birleştirelim. Buradan $X'XG'X'=X'$ dür. Şimdi transpozesini alarak ispat tamamlanabilir.

4. Diğer herhangi bir H g–tersi için, $XGX'=XHX'$ dür. Gerçekten (3) e göre, $XGX'X=X=XHX'X$ dir. Lemma 3.5 i uygulamak sonucu verir.

5. XGX' simetriktir. Gerçekten $\left(XGX'\right)'=XG'X'$ dür. (1) den, G' de bir g–tersidir. Bu nedenle, (4) e göre, $XG'X'=XGX'$ dür.

$X'X$ in bir genelleştirilmiş tersini hesaplama problemini ele alalım. X_1 X den r sayıda lineer bağımsız sütundan oluşmak üzere, $X=(X_1,X_2)$ olsun. Bu takdirde $X'X$ in bir genelleştirilmiş tersi

$$\left(X'X\right)^{-}=\begin{pmatrix}\left(X_1'X_1\right)^{-1} & 0\\ 0 & 0\end{pmatrix}$$

dır. Bu sonuç aşağıdaki lemmanın bir özel durumudur.

Lemma 3.6. $W_{p\times p}$ matrisi r ranka sahip olsun ve A matrisi tam r ranka sahip olmak üzere,

$$W = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

olarak parçalanmış olsun. Bu takdirde W nun bir genelleştirilmiş tersi

$$W^{-} = \begin{pmatrix} A^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

dır.

İspat.

$$WW^{-}W = \begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} A^{-1} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A & B \\ C & CA^{-1}B \end{pmatrix}$$

dir. Eğer $D = CA^{-1}B$ ise, bu W ya eşittir. Bu niçin doğru olmalı? A matrisi r ranka sahip olduğundan (W ile aynı ranklı olduğundan) dolayı, bu durum $(B \quad D)'$ nin herhangi bir sütunun, $(A \quad C)'$ nin sütunlarının lineer kombinasyonu olarak yazılabildiğini söyler. Yani herhangi bir F matrisi için, $(B \quad D)' = (A \quad C)'F$ dir. Ancak bu takdirde $B = AF$ dir ve bu nedenle $F = A^{-1}B$ dir. Öte yandan $D = CF$ olduğundan $D = CA^{-1}B$ dir. Aynı zamanda $R(X)$ üzerine izdüşüm matrisi için, daima

$$P = X_1\left(X_1'X_1\right)^{-1}X_1'$$

yi kullanabildiğimize de dikkat edelim. Böyle bir P matrisinin $R(X_1)$ üzerine bir izdüşüm olduğunu ve aynı zamanda $R(X_1) = R(X)$ olduğunu biliyoruz.

Örnek 3.21. 2 gruplu bir – yönlü ANOVA:

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n_2} \end{pmatrix}$$

dır. Bu durumda X in rankı 2 dir ve

$$X'X = \begin{pmatrix} n & n_1 & n_2 \\ n_1 & n_1 & 0 \\ n_2 & 0 & n_2 \end{pmatrix}$$

dir. X_1 X in ilk iki sütunu olsun. Bu takdirde

$$(X_1'X_1)^{-1} = \begin{pmatrix} n & n_1 \\ n_1 & n_1 \end{pmatrix}^{-1} = \frac{1}{nn_1 - n_1^2}\begin{pmatrix} n_1 & -n_1 \\ -n_1 & n \end{pmatrix} = \begin{pmatrix} \frac{1}{n_2} & \frac{-1}{n_2} \\ \frac{-1}{n_2} & \frac{1}{n_1} + \frac{1}{n_2} \end{pmatrix}$$

dir ve böylece $X'X$ in bir genelleştirilmiş tersi

$$\left(X'X\right)^{-} = \begin{pmatrix} n_2^{-1} & -n_2^{-1} & 0 \\ -n_2^{-1} & n_1^{-1} + n_2^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

dır. Bu durumda normal denklemlerin bu genelleştirilmiş terse karşılık gelen $\hat{\beta}$ çözümü $j = 1, 2, \ldots, n_i$ olmak üzere,

$$\hat{\beta} = \begin{pmatrix} n_2^{-1} & -n_2^{-1} & 0 \\ -n_2^{-1} & n_1^{-1} + n_2^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \sum_j Y_{1j} + \sum_j Y_{2j} \\ \sum_j Y_{1j} \\ \sum_j Y_{2j} \end{pmatrix} = \begin{pmatrix} \overline{Y}_2 \\ \overline{Y}_1 - \overline{Y}_2 \\ 0 \end{pmatrix}$$

olarak ifade edilir. Buradan $\hat{Y}$ yı $X\hat{\beta}$ olarak hesaplarsak,

$$\begin{pmatrix} 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} \overline{Y}_2 \\ \overline{Y}_1 - \overline{Y}_2 \\ 0 \end{pmatrix} = \begin{pmatrix} \overline{Y}_{1.} \\ \vdots \\ \overline{Y}_{1.} \\ \overline{Y}_{2.} \\ \vdots \\ \overline{Y}_{2.} \end{pmatrix}$$

elde edilir.

Lemma 3.7. $rank\left(X\right) = r < p$ olsun ve $\left(X'X\right)^{-}$ matrisi $\left(X'X\right)$ in bir genelleştirilmiş tersi olmak üzere $P = X\left(X'X\right)^{-} X'$ olsun. Bu takdirde,

(i) P ve $I - P$ izdüşüm matrisleridir.

(ii) $rank(I-P) = iz(I-P) = n-r$ dir.

(iii) $PX = X$ dir.

İspat.

(i) Teorem 3.16 nın (5) kısmına göre, P simetrik ve idempotentdir. $G = (X'X)^{-}$ olsun, bu nedenle $P = XGX'$ dür. Bu takdirde Teorem 3.16 nın (3a) kısmına göre

$$P^2 = (XGX')(XGX') = (XG)(X'XGX') = XGX'$$

dür.

Bu nedenle, P simetrik ve idempotentdir ve bundan dolayı, bir izdüşüm matrisidir. Bu takdirde $I-P$ nin de simetrik ve idempotent olduğu kolayca görülür.

(ii) İzdüşüm matrisleri için $rank = iz$ olduğunu biliniyor.

(iii) Teorem 3.16 nın (3b) kısmına göre, $PX = XGX'X = X$ dir.

Yöntem 2. Modeli tam ranklı modele indirgeyelim. X_1 matrisi X in r sayıda lineer bağımsız sütunundan oluşsun ve X_2 matrisi de geriye kalan sütunlardan oluşsun. Bu takdirde X_1 in sütunları X in sütun uzayını gerdiğinden, bir F için, $X_2 = X_1F$ dir.

$$X = (X_1, X_2) = (X_1, X_1F) = X_1(I_{r\times r}, F)$$

yazalım. Bu ise $rank(K_{n\times r}) = r$ ve $rank(L_{r\times p}) = r$ olmak üzere, $X = KL$ çarpımının bir özel durumudur. Bu genel notasyonu kullanarak,

$$E[Y] = X\beta = KL\beta = K\alpha$$

yazılabilir. Bu nedenle, lineer modeli tam ranklı bir modele yeniden parametrelemiş oluruz. K tam ranka sahip olduğundan, α nın en küçük kareler tahmini

$$\hat{\alpha} = (K'K)^{-1}K'Y$$

ve

$$\hat{Y} = K\hat{\alpha} = K(K'K)^{-1}K'Y$$

olacaktır. Bu nedenle, $P = K\left(K'K\right)^{-1} K'$ alınır. Eğer $K = X_1$ ise, bu bize daha önceki $P = X_1\left(X_1'X_1\right)^{-1} X_1'$ ifadesini verir.

Örnek 3.21 e göre *X* in üçüncü sütunu birinci sütun eksi ikinci sütundur. Bu nedenle $F = (1 \quad -1)'$ dir. Buradan

$$X = \left(X_1, X_1 F\right) = \underset{K}{X_1} \underbrace{\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix}}_{L}$$

yazılabilir. $\alpha = L\beta$ olmak üzere $E(Y) = X\beta = KL\beta = K\alpha$ dır. Buradan

$$\hat{\alpha} = \left(K'K\right)^{-1} K'Y$$

$$= \begin{pmatrix} n & n_1 \\ n_1 & n_1 \end{pmatrix}^{-1} \begin{pmatrix} \sum_j Y_{1j} + \sum_j Y_{2j} \\ \sum_j Y_{1j} \end{pmatrix} = \begin{pmatrix} n_2^{-1} & -n_2^{-1} \\ -n_2^{-1} & n_1^{-1} + n_2^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \sum_j Y_{1j} + \sum_j Y_{2j} \\ \sum_j Y_{1j} \end{pmatrix} = \begin{pmatrix} \bar{Y}_2 \\ \bar{Y}_1 - \bar{Y}_2 \end{pmatrix}$$

olup,

$$\hat{Y} = X_1\hat{\alpha} = \left(\bar{Y}_{1.} \quad \cdots \quad \bar{Y}_{1.} \quad \bar{Y}_{2.} \quad \cdots \quad \bar{Y}_{2.}\right)$$

elde edilir.

Yöntem 3. Tahmin edilebilirlik kısıtlamasını koyalım. β yı bir tek şekilde belirlemek, yani, herhangi bir $\hat{Y} \in R\left(X\right)$ için $\hat{\beta}$ nın bir tek olduğunu belirlemek için β üzerine $H_{s\times p}\beta = 0$ biçiminde $s = p - r$ sayıda kısıtlamayı koyalım. Bu takdirde, $X\hat{\beta} = \hat{Y}$ ve $H\hat{\beta} = 0$ denklemlerini sağlayan bir tek $\hat{\beta}$ vardır. Görüntü düzeni için "şapkaları" düşürelim. Bunu

$$\left(Y \quad 0\right)' = \left(X \quad H\right)' \beta = G\beta$$

olarak yeniden yazabiliriz. Hangi şartlar altında amacımızı başarıyla tamamlarız? $\left(Y \quad 0\right)' = G\beta$ için ne zaman bir tek çözüm vardır? Bu sorunun cevabını bir sonraki lemmada vereceğiz.

Lemma 3.8. Bir tek çözümün var olması için gerek ve yeter şart G nin *p* ranka sahip olması ve *H* nin satırlarının *X* in satırlarından lineer bağımsız olmasıdır.

İspat. Her $Y \in R(X)$ için bir β çözümü mevcut olması için gerek ve yeter şart her $(Y \quad 0)' \in R(G)$ ve her $u \in R^{n+s}$ için $G'u = 0 \Rightarrow (Y \quad 0)u = 0$ olmasıdır. Bu durumda $(Y \quad 0)' = GF$ olduğundan, $(GF)' u = F'G'u = F'0 = 0$ olacaktır. Bu ise

$$(X', H')\begin{pmatrix} u_X \\ u_H \end{pmatrix} = 0 \Rightarrow (Y', 0')\begin{pmatrix} u_X \\ u_H \end{pmatrix} = 0$$

eşitliğine eşdeğerdir, yani

$$\begin{aligned} X'u_X + H'u_H = 0 &\Rightarrow \text{her } Y \in R(X) \text{ için, } Y'u_x = 0 \\ &\Rightarrow \text{her } a \text{ için, } (Xa)' u_X = 0 \\ &\Rightarrow X'u_X = 0 \Leftrightarrow X'u_X + H'u_H = 0 \Leftrightarrow X'u_X = 0 \text{ ve } H'u_H = 0 \end{aligned}$$

dır. Bu tamı tamına X in satırlarının H nin satırlarından lineer bağımsız olduğunu söyler. Çözümün bir tek olması için gerek ve yeter şart G nin sütunlarının lineer bağımsız olacağıdır, yani $rank(G) = p$ olacağıdır.

Sonuç 3.1. Bir tek çözümün mevcut olması için gerek ve yeter şart G nin p ranka ve H nin $s = p - r$ ranka sahip olmasıdır. β yı tahmin etmede bu yöntemi kullanmak için $\hat{Y} = X\hat{\beta}$ yı ve $H\hat{\beta} = 0$ yı, yani

$$X'X\hat{\beta} = X'Y \text{ ve } H'H\beta = 0$$

artırılmış normal denklemlerini çözeriz. Bu eşitlikler birlikte

$$(X'X + H'H)\hat{\beta} = (G'G)\hat{\beta} = X'Y$$

eşitliğini verir. Bu durumda çözüm

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

ile verilir ve dolayısıyla $P = X(G'G)^{-1} X'$ olmak üzere,

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

dır. Şimdi Yöntem 3 ü örnek 3.21 e uygulayalım. Bu durumda

$$H\beta \equiv (0,1,1)\begin{pmatrix}\mu \\ \alpha_1 \\ \alpha_2\end{pmatrix} = 0$$

$$G'G = \begin{pmatrix} n & n/2 & n/2 \\ n/2 & n/2+1 & 1 \\ n/2 & 1 & n/2+1 \end{pmatrix}$$

$$(G'G)^{-1} = \begin{pmatrix} \frac{n+4}{4n} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{n+4}{4n} & \frac{n-4}{4n} \\ -\frac{1}{4} & \frac{n-4}{4n} & \frac{n+4}{4n} \end{pmatrix}$$

$$\hat{\beta} = (G'G)^{-1} X'Y = \begin{pmatrix} \bar{Y}.. \\ \frac{1}{2}(\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}) \\ \frac{1}{2}(\bar{Y}_{2\cdot} - \bar{Y}_{1\cdot}) \end{pmatrix}$$

dir. $\hat{\beta}$ nın $\alpha_1 + \alpha_2 = 0$ kısıtlamasını sağladığı açıktır. Yine,

$$\hat{Y} = X\hat{\beta} = (\bar{Y}_{1\cdot} \quad \cdots \quad \bar{Y}_{1\cdot} \quad \bar{Y}_{2\cdot} \quad \cdots \quad \bar{Y}_{2\cdot})'$$

elde edilir.

Yöntem $3\frac{1}{2}$. Farklı olarak uygulanan belirlenebilirlik kısıtlamaları koyalım. Bunun için, beş gruplu bir yönlü ANOVA örneğini göz önüne alalım. Burada $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 = 0$ kısıtlamasını kullandığımıza dikkat edelim. Bu örnekle açıklayacak olduğumuz, bir tam–ranklı modeli formüllemek için, bu kısıtlamayı kullanabiliriz. Bu nedenle bu yöntem, yöntem 2 ve yöntem 3 arasında düşünülmelidir. Beş grubun her birinden bir "tekrarlamayı" beş gözlemi göz önüne alalım. $\mu, \alpha_1, \alpha_2, \alpha_3, \alpha_4$ (ancak α_5 değil) için X tasarım matrisi beş sütuna (altıncı sütun yok) sahip olacaktır. Yani

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 \end{pmatrix}$$

dir. Beş gruptan gözlem için X in satırı $\alpha_5 = -\alpha_1 - \alpha_2 - \alpha_3 - \alpha_4$ ü kullanarak elde edilir.

Eğer $X_{n\times p}$, $r < p$ ranka sahip ise, normal denklemlere bir tek çözüm yoktur. Bir $\hat{\beta}$ çözümünü ve $\hat{Y}$ ortogonal izdüşümünü bulmak için üç yola sahibiz:

1. Modeli tam ranklı bir modele indirgeme.
2. Bir $\left(X'X\right)^{-}$ genelleştirilmiş tersini bulma.
3. Belirlenebilirlik (tahmin edilebilirlik) sınırlamalarını koyma.

Tam ranklı modele indirgeme metodunu aşağıdaki gibi açıklayabiliriz. X_1, X den r sayıda lineer bağımsız sütunlardan oluşsun ve X_2 geriye kalan sütunlardan ibaret olsun. Bu takdirde $X_2 = X_1 F$ dir, çünkü X_2 nin sütunları, X_1 in sütunlarına lineer bağımlıdır. Bu nedenle

$$X = \left(X_1, X_2\right) = \left(X_1, X_1 F\right) = X_1\left(I_{r\times r}, F\right)$$

dir. $rank\left(K_{n\times r}\right) = r$ ve $rank\left(L_{r\times p}\right) = r$ olmak üzere, bu; $X = KL$ nin bir özel durumudur. Bu durumda

$$E\left[Y\right] = X\beta = KL\beta = K\alpha$$

dır. K tam ranka sahip olduğundan, α nın en küçük kareler tahmini $\hat{\alpha} = \left(K'K\right)^{-1} K'Y$ ve ortogonal izdüşüm $\hat{Y} = K\hat{\alpha} = K\left(K'K\right)^{-1} K'Y$ dir. Bu nedenle, $P = K\left(K'K\right)^{-1} K'$, yani $P = X_1\left(X_1'X_1\right)^{-1} X_1'$ dür.

Örnek 3.22. (2 – gruplu bir – yönlü ANOVA**)**

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n_2} \end{pmatrix}$$

X_1 X in ilk iki sütunundan oluşsun. Bu takdirde

$$X = X_1 \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix}$$

ve

$$\alpha = \begin{pmatrix} \mu + \alpha_2 \\ \alpha_1 - \alpha_2 \end{pmatrix}$$

olmak üzere $X\beta = X_1\alpha$ dır. Bu takdirde

$$\hat{\alpha} = \begin{pmatrix} n & n_1 \\ n_1 & n_1 \end{pmatrix}^{-1} \begin{pmatrix} \sum_j Y_{1j} + \sum_j Y_{2j} \\ \sum_j Y_{1j} \end{pmatrix} = \begin{pmatrix} n_2^{-1} & -n_2^{-1} \\ -n_2^{-1} & n_1^{-1} + n_2^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \sum_j Y_{1j} + \sum_j Y_{2j} \\ \sum_j Y_{1j} \end{pmatrix}$$
$$= \begin{pmatrix} \bar{Y}_2 \\ \bar{Y}_1 - \bar{Y}_2 \end{pmatrix}$$

dir ve bu nedenle, $\hat{Y} = X_1\hat{\alpha} = \left(\bar{Y}_{1.}, \ldots, \bar{Y}_{1.}, \bar{Y}_{2.}, \ldots, \bar{Y}_{2.}\right)'$ olacaktır. Öte yandan

$$X'X = \begin{pmatrix} n & n_1 & n_2 \\ n_1 & n_1 & 0 \\ n_2 & 0 & n_2 \end{pmatrix}$$

yazılabilir. Eğer X_1 matrisi X in ilk iki sütunundan oluşur ise, bu takdirde

$$\left(X_1'X_1\right)^{-1} = \begin{pmatrix} n & n_1 \\ n_1 & n_1 \end{pmatrix}^{-1} = \begin{pmatrix} n_2^{-1} & -n_2^{-1} \\ -n_2^{-1} & n_1^{-1} + n_2^{-1} \end{pmatrix}$$

dir ve $X'X$ in genelleştirilmiş tersi

$$\left(X'X\right)^{-}=\begin{pmatrix} n_2^{-1} & -n_2^{-1} & 0 \\ -n_2^{-1} & n_1^{-1}+n_2^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

dır. Böylece normal denklemlere bir çözüm

$$\hat{\beta}=\begin{pmatrix} n_2^{-1} & -n_2^{-1} & 0 \\ -n_2^{-1} & n_1^{-1}+n_2^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \sum_j Y_{1j}+\sum_j Y_{2j} \\ \sum_j Y_{1j} \\ \sum_j Y_{2j} \end{pmatrix}=\begin{pmatrix} \overline{Y}_2 \\ \overline{Y}_1-\overline{Y}_2 \\ 0 \end{pmatrix}$$

dır ve önceden olduğu gibi $\hat{Y}=X\hat{\beta}=\left(\overline{Y}_{1.},\ldots,\overline{Y}_{1.},\overline{Y}_{2.},\ldots,\overline{Y}_{2.}\right)'$ dür.

Bir $\left(X'X\right)^{-}$ genelleştirilmiş tersini bulmak için $X=\left(X_1,X_2\right)$ olsun. Burada X_1 matrisi X in r sayıda lineer bağımsız sütundan oluşmaktadır. Bu takdirde $X'X$ in bir genelleştirilmiş tersi

$$\left(X'X\right)^{-}=\begin{pmatrix} \left(X_1'X_1\right)^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

dır. Dolayısıyla normal denklemlerin bir çözümü $\hat{\beta}=\left(X'X\right)^{-}X'Y$ ve $\hat{Y}=X\hat{\beta}=X\left(X'X\right)^{-}X'Y=PY$ dir. Burada $P=X\left(X'X\right)^{-}X'$ dir. Bunun aynı zamanda $P=X_1\left(X_1'X_1\right)^{-1}X_1'$ yü verdiğine de dikkat edelim. Bu sonuç aşağıdaki teoremin bir özel durumudur.

Teorem 3.17. *A* matrisi *r* ranka sahip olmak üzere, $W_{p\times p}$ *r* ranka sahip olsun ve

$$W=\begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

olarak parçalanmış olsun. Bu takdirde *W* un bir genelleştirilmiş tersi

$$W^{-}=\begin{pmatrix} A^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

olarak ifade edilir.

β yı bir tek şekilde belirleyebilmek için, yani herhangi bir $\theta\in R\left(X\right)$ için,

$$X\beta=\theta \text{ ve } H\beta=0$$

bağıntılarını sağlayan bir tek β var olacak şekilde, β üzerine $s=p-r$ sayıda kısıtlama koyalım. Bu durum

$$\begin{pmatrix}\theta\\0\end{pmatrix}=\begin{pmatrix}X\\H\end{pmatrix}\beta=G\beta$$

olarak yazılabilir. Şimdi ne zaman bir bir tek çözüm vardır? sorusuna cevap arayalım.

Teorem 3.18. Bir tek çözümün var olması için gerek ve yeter şart G nin p ranka sahip olması ve H nin $p-r$ ranka sahip olup, satırlarının X in satırlarından lineer bağımsız olmasıdır.

β yı tahmin etmek için $\hat{Y}=X\hat{\beta}$ ve $H\hat{\beta}=0$ denklemlerini, yani $X'X\hat{\beta}=X'Y$ ve $H'H\hat{\beta}=0$ denklemlerini veya $\left(X'X+H'H\right)\hat{\beta}=\left(G'G\right)\hat{\beta}=X'Y$ denklemini çözeriz. Bu nedenle

$$\hat{\beta}=\left(G'G\right)^{-1}X'Y \text{ ve } \hat{Y}=X\hat{\beta}=PY$$

dir. Burada $P=X\left(G'G\right)^{-1}X'$ dir.

Örnek 3.23. (2 gruplu bir – yönlü ANOVA ya devam) $\alpha_1+\alpha_2=0$, yani

$$H\beta\equiv(0,0,1)\begin{pmatrix}\mu\\\alpha_1\\\alpha_2\end{pmatrix}=0$$

alalım ve $n_1=n_2=m$ olduğunu farz edelim. Bu takdirde

$$\hat{\beta}=\begin{pmatrix}\overline{Y}_{..}\\\frac{1}{2}\left(\overline{Y}_{1.}-\overline{Y}_{2.}\right)\\\frac{1}{2}\left(\overline{Y}_{2.}-\overline{Y}_{1.}\right)\end{pmatrix}$$

nın normal denklemleri sağladığı ve daha açıkçası $\alpha_1+\alpha_2=0$ kısıtlamasını sağladığı gösterilebilir. Bu nedenle önceden olduğu gibi

$$\hat{Y} = X\hat{\beta} = \left(\bar{Y}_{1.},\ldots,\bar{Y}_{1.},\bar{Y}_{2.},\ldots,\bar{Y}_{2.}\right)'$$

elde edilir.

3.4.1 Eksik Ranklı Modelleri Görselleştirme

Doğrudaş bağımsız değişkenlerin problemi meşhurdur. Bir regresyon modelinde doğrudaşlık (bağımsız değişkenler arasında bir lineer bağımlılık) var olduğunda regresyon katsayıları için bir tek çözüm yoktur. Çözümlerin bir tek kümesinden 3 ya da çözümlerin sonsuz bir sayısı var olduğundan bu regresyon katsayılarının belirlenmediğini söyleriz. Rank eksikliği 1 olduğunda bu çözümlerden birini elde etmek için, bir yaygın strateji, regresyon katsayıları üzerine bir kısıtlama koymaktır. Örneğin, regresyon katsayılarının ikisini eşit olarak veya onların birinin başka birinin iki katı kadar büyük olarak yerleştirelim. Bu alışılageldik şekliyle istatistiksel programları kullanarak başarılabilir veya genelleştirilmiş terslerin kullanımıyla veya veriyi yaratıcı yeniden kodlamayla yerine getirilebilir. Bu çalışma doğrudaşlığı dikkate alarak regresyon problemlerinin çözümü için yeni bir yöntem önermez. Bunun yerine lineer bağımlılık probleminin, (doğrudaşlığın) genel bir geometrik görünümü önerilir. Böyle durumlarda regresyon denklemlerini çözmeye en uygun yaklaşımın (kısıtlanmış-regresyon) genelleştirilmiş terslerin geometrik olarak nasıl incelenebildiğini gösterir. Kısıtlanmış çözümler bir veya daha fazla kısıtlamayı birleştiren özel genelleştirilmiş tersleri yaratarak herhangi bir genelleştirilmiş ters hem bir cebirsel hem de bir geometrik anlamda çözümleri sınırlar. Paket programların kullanıcıları kısıtlanmış çözümleri elde etmek için ne bu geometriyi kullanacak ne de bu genelleştirilmiş tersleri kullanan araştırmacı olacak, ancak bu geometri doğrudaşlık problemini anlamada, bu kısıtlanmış çözümlerin nasıl işlediğinde, bir çözümü belirlemek için birçok kısıtlamanın nasıl gerekli olduğunda ve bazı kısıtlamaların belirlenen çözümleri niçin üretmediğinde son derece yararlıdır. Genel geometrik bakış açısı aynı zamanda bir kısıtlanmış çözümün akla yatkınlığını kılan şeyi yargılamada da yardımcı olabilir.

Eksik ranklı matrisler bağımsız değişkenlerin biri veya daha fazlasının modeldeki diğer bağımsız değişkenlerin bir lineer fonksiyonu olduğunda ortaya çıkar. Bu tür bağımlılıklar araştırma esnasında doğal olarak (kendiliğinden) ortaya çıkabilir. İki değişik örnek; yekün (toplam) test puanını matematik kısmı üzerindeki puan artı sözlü kısım üzerindeki puandan oluştuğunda ve biri toplam puanın (TP), yani, GPA üzerinde matematik puanı (MP) ve sözlü puanı (SP) yi incelemek istediğinde. Lineer bağımlılık SP+MP = TP dir. Eğitim durumu (ED), mesleki durum (MD) ve durum tutarsızlığı (DT) yi ayırma DT = MD – ED dir. Bu örnekler çoğaltılabilir. Bu senaryolarda, bağımsız değişkenlerin her biri sonuç değişkeni üzerinde bir etkiye sahip olabilir, ancak bu durumların tümünde bağımsız değişkenler lineer bağımlıdır. Burada amacımız, bağımsız değişkenlerin matrisi eksik ranklı olduğunda (tam ranklı olmadığında) rank eksikliği probleminin basit bir geometrik bakış açısını ve çözümlerin genelleştirilmiş tersleri; kısıtlanmış regresyonu kullanarak nasıl elde edildiğini vermektir.

Kullanılan yöntem belirlidir. Bir, iki, üç boyutlu basit uzaylarla işe başlanır. Bundan sonra da bu yaklaşımı dört veya daha fazla boyutlu durumlara genişletilir. Bu geometriyi anlamak bir, iki, üç boyutlu durumlarda bile biraz çaba gerektirir ve aşikar olarak, dört veya daha fazla boyutların geometrisine taşındığımızda daha fazla çaba gerektirir. Basitleştirmek için bundan böyle okuyanlar için bu en bilinen durum olduğundan alışılmış en küçük kareler (OLS) regresyonu ile ilgili normal denklemlerle ilgileneceğiz. En basit durum olan iki değişkenli durum ile başlayacağız. Her bir bağımsız değişken değerlerinden bağımsız değişkenin ortalamasını ve bağımlı değişken değerlerinden, bağımlı değişken ortalamasını çıkarırız. Bu bizi standart değerlere götürür ve bu durum regresyon sabiti sıfır olduğundan bize bu iki değişken arasında sadece bir regresyon katsayısını göz önüne alma imkanı verir. Bu durumda sadece normal bir denklem vardır.

Bir bağımsız değişken bir bağımlı değişken durumunda regresyon katsayısını bulmak için gereken sadece iki nicelik vardır, bunlar bağımsız değişken için kareler toplamı $\sum x^2$ ve bağımlı ve bağımsız değişkenler için çarpımların toplamı $\sum xy$. Bu iki değişken durumunda, bilinen $b = \sum xy \big/ \sum x^2$ çözümünü veren bir

$$\left(\sum x^2\right) b = \sum xy \tag{3.34}$$

normal denklemi vardır. Matris cebirini kullanarak bu aynı denklemi $X'Xb = X'y$ olarak yazalım. Burada X bağımsız değişken üzerindeki n sayıda gözlemin sayılarının üzerindeki sapmaların bir $n \times 1$ vektörüdür. Matris çarpımlarını gerçekleştirdiğimizde bir tek denkleme, yani, (3.34) denklemine bakarız. Somutluk için, $\sum x^2$ ve $\sum xy$ için değerlerini hesaplayıp onları (3.34) te yerine koyalım ve $\sum x^2 = 4$ ve $\sum xy = 8$ olsun.

Bu takdirde (3.34) denklemini $4b = 8$ olarak yazar ve bu nedenle, $b = 2$ buluruz. Geometrik olarak çözüm uzayı sadece bir boyuta sahiptir ve (3.34) denklemi bize bir doğru üzerinde bir tek nokta için çözüm imkanı verir. (3.34) denklemi, b çözümünün olası değerlerinin bir boyut üzerinde nerede olduğunu belirler.

Bu yöntemi iki bağımsız değişken durumuna genişletelim. Onların ortalamalarını onlardan çıkartmak suretiyle yine değişkenleri merkezileştiririz, böylece değişkenlerin tümü sapma sayı biçimindedir. Onları bir veya iki ile indisleyerek iki bağımsız değişken arasındaki ayrım x_1 ve x_2 olarak ortaya konulur. Bir cebirsel bakış açısından ilgili nicelikler $\sum x_1^2$, $\sum x_2^2$, $\sum x_1x_2$, $\sum x_1y$, $\sum x_2y$ olacaktır. Çoklu regresyon içeren tanıtıcı metinlerden formüller, bu nicelikleri formüllerde yerine koyma ve iki regresyon katsayısı için çözme imkanı verir. Matris cebiri gösterimi $X'Xb = X'y$ yi aynı bırakır, fakat bu durumda X matrisi iki sütun (bağımsız değişkenlerin her biri için 1) ve n satır (gözlemlerin her biri için 1) içerir. b vektörü birinci bağımsız değişken bir (b_1) ve ikinci bağımsız değişken için bir (b_2) olmak üzere iki elemana sahiptir. Kareler ve çapraz çarpımlar toplamlarını kullanarak, matris formu açık bir şekilde aşağıdaki gibi yazılır.

$$\begin{bmatrix} \sum x_1^2 & \sum x_1x_2 \\ \sum x_1x_2 & \sum x_2^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sum x_1y \\ \sum x_2y \end{bmatrix} \tag{3.35}$$

(3.35) deki matris çarpımını yerine getirerek, (3.35) normal denklemi aşağıdaki gibi yazılır

$$\begin{aligned} \left(\sum x_1^2\right)b_1 + \left(\sum x_1x_2\right)b_2 &= \sum x_1y \\ \left(\sum x_1x_2\right)b_1 + \left(\sum x_2^2\right)b_2 &= \sum x_2y \end{aligned} \tag{3.36}$$

elde edilir. Bu normal denklemlerin her biri bir doğru için denklemidir. Bir doğru için denklemin genel formu $Ab_1 + Bb_2 = c$ şeklindedir.

Yine [$\sum x_1^2$, $\sum x_2^2$, $\sum x_1 x_2$, $\sum x_1 y, \sum x_2 y$] kareler ve çapraz çarpımların toplamları için bazı uygun değerler hesaplanır ve bunlar (3.36) da yerine koyularak gerçek veriden çıkarılabilen

$$\begin{aligned} 4b_1 + 2b_2 &= 8 \\ 2b_1 - 3b_2 &= -4 \end{aligned} \tag{3.37}$$

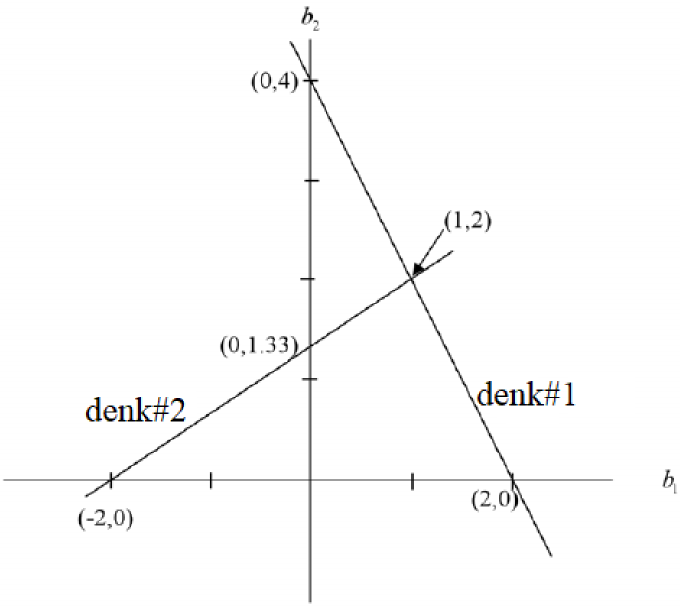
normal denklemler takımı üretilir. Şimdi bu iki denklem sistemini örneğin $1.5b_2 - 2$ yi b_1 için birinci denklemde yerine koyarak çözebiliriz ve $b_2 = 2$ olduğunu buluruz, bu takdirde b_2 yi bilerek b_1 için çözümün 1 e eşit olduğunu kolayca görebiliriz.

Geometrik olarak, çözüm uzayı b_1 için bir ve b_2 için bir olmak üzere iki boyuta sahiptir. (3.37) deki normal denklemler doğrular için denklemlerdir ve eğer bu iki doğru bu iki boyutlu çözüm uzayındaki bir noktada kesişirse, bu nokta bu iki denklem sistemine bir tek çözümü belirleyecektir. Bu, Şekil 3.6 da gösterilir. Yatay eksen b_1 için çözümleri ve düşey eksen b_2 için çözümleri temsil eder. (3.37) deki denklemlere dayanan iki doğru aşağıdaki tarzda inşa edilir. Birinci denklemi kullanarak, eğer $b_2 = 0$ ise bu takdirde $b_1 = 2$ dir, bu nedenle doğru üzerindeki noktalardan biri (2,0) dır. Öte yandan, eğer $b_1 = 0$ ise, bu takdirde $b_2 = 4$ dür ve bu birinci doğru üzerindeki bir ikinci nokta (0,4) dür ve bu iki nokta bize iki boyutlu çözüm uzayında bu birinci doğruyu çizme imkanı verir. İkinci doğru aynı tarzda inşa edilir, $b_2 = 0$ alarak $b_1 = -2$ buluruz, bu nedenle doğru üzerindeki bir nokta $(-2, 0)$ dır. Eğer $b_1 = 0$ ise, bu takdirde $b_2 = 1.33$ dür, bu durumda doğru üzerindeki bir ikinci nokta (0,1.33) dür. Bu bize ikinci doğruyu inşa etme imkanını verir. Bu iki doğru (1,2) noktasında, yani $b_1 = 1$ ve $b_2 = 2$ de kesişir. Bu, iki bağımsız değişkenli normal denklemlere çözümün geometrik görünümüdür. O, çoğu okuyucu için muhtemelen biliniyordur (yine farklı bir bağlamdan) iki denklemin lineer olarak bağlı olduğu, örneğin

$$\begin{aligned} 4b_1 + 2b_2 &= 8 \\ 2b_1 + 1b_2 &= 4 \end{aligned} \tag{3.38}$$

durumunu düşünelim. İkinci denklem birinci denklemin $1/2$ ile çarpılmış şeklidir. Bu denklemlere bir tek çözüm yoktur. b_1 $(b_1 = -.50b_2 + 2)$ ikinci denklemler değerini b_1 için birinci denklemler değerine bedel olarak koyduğumuzda ve b_2 için çözdüğümüzde b_2 herhangi bir değeri alabildiğinden, pek bilgilendirici olmayan bir sonucu, yani $0b_2 = 0$ sonucunu elde ederiz. Buradan b_2 nin belirlenmediğini söyleriz. Eğer ikinci denklemden b_2 nin değerini $(b_2 = -2b_1 + 4)$ birinci denklemde yerine koyarsak, $0b_1 = 0$ olduğunu görürüz. Geometrik olarak birinci denklemi önceki gibi çizeriz ve Şekil 3.6 daki denklem 1 için doğru ile sonuçlandırırız.

İkinci doğruyu çizdiğimizde onun b_1 eksenini (0,4) de, b_2 eksenini (2,0) da kestiğini buluruz. Yani bu iki denklem için doğrular çakışır. Bu denklemlere herhangi çözümler bu doğru üzerinde yer alır. Örneğin, bu denklemlerin her ikisine bir çözüm (2,0) dır, aynı zamanda (0,4) ve (1,2) de çözümdür.



Şekil 3.6 Lineer Bağımlılık Olmamak Üzere Bir İki–Boyutlu Çözüm Uzayında Bir Regresyon Çözümünün Geometrik Görünümü

Burada çözüm 1 ve 2 denklemlerini temsil eden iki doğrunun kesiştiği (1,2) dir. Bu iki denkleme sonsuz sayıda çözüm vardır ve onların hepsi iki boyutlu bir uzaydaki bu doğru üzerinde bulunur. Bu doğru için denklemi yazmak için bir aydınlatıcı yöntem "bir doğru için vektör denklemi" gibidir. Yani doğru üzerindeki noktalardan biri artı bir (k) skaleri çarpı "doğrunun doğrultusu" gibidir, yani

$$\begin{bmatrix} 0 \\ 4 \end{bmatrix} + k \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

biçimindedir. Çözümlerin bir doğrusunun geometrik ifadesi bize sadece b_1 ve b_2 katsayılarının belirlenmediğini söylemez, aynı zamanda b_1 ve b_2 katsayılarının kombinasyonlarının normal denklemleri çözdüğünü de söyler. Bunun nasıl işlediğini açıklamak için daha önce (0,4) ün doğru üzerinde olduğunu ve $k=0$ olduğunda onun bir çözüm olduğunu ve (1,2) nin bu doğru üzerinde olduğunu ve $k=1$ olduğunda onun için bir çözüm olduğunu ve (2.0) ın bu doğru üzerinde olduğunu ve $k=2$ olduğunda onun için bir çözüm olduğunu göstermiş olduğumuza dikkat edelim. k için diğer değerleri seçmek bu doğru üzerindeki diğer değerler; yani iki denklemli bu sisteme diğer çözümlerin herhangi birini üretecektir. Önemle, her ne kadar bu iki denklemin sonsuz çoklukta çözümleri mevcut ise de yegane çözümler bu doğru üzerinde yer alır. Bu noktada sıfır vektörünü ortaya koymak uygundur. Sıfır vektör bir matris ile çarpıldığında sıfırların bir vektörünü ortaya koyan bir vektördür. Şimdi normal denklemler ve $X'X$ üzerine odaklanalım. Bu bağlamda sıfır vektör, sıfır vektörün elemanlarının tümü sıfırlar olmaması şartı altında $X'X$ ile çarpıldığında sıfırların bir vektörünü üreten vektördür. (3.38) için $X'X$ matrisini yazarak $\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$ elde ederiz ve $X'X$ ile çarpıldığında $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ vektörünün $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ vektörünü ürettiğine dikkat edelim. Sıfır vektör $(1,-2)'$ dür ve bunu v olarak gösterelim. Bu, v nin bir skalerle çarpımı kadar tektir. (3.38) deki $X'X$ için sadece bir tane sıfır vektör vardır, çünkü yalnız bir lineer bağımlılık vardır (sadece iki bağımsız değişken ile ilgili birden daha fazla lineer bağımlılık var olamaz). Onlar aynı doğrultuyu paylaştığından çözümlerin doğrusunun sıfır vektöre paralel olduğuna dikkat edelim. Sıfır vektör –2 lik bir eğimle (0,0) noktasından geçen bir doğrudur. Bağımsız değişkenler arasında lineer bağımlılığın sebep olduğu çözümleri ve problemleri geometrik olarak görselleştirmenin nispeten kolay olduğu son durum üç bağımsız değişkenin var olduğu durumdur. Kareler ve çapraz çarpımlarının toplamlarının matrisi matris formunda aşağıdadır.

$$\begin{bmatrix} \sum x_1^2 & \sum x_1x_2 & \sum x_1x_3 \\ \sum x_2x_1 & \sum x_2^2 & \sum x_2x_3 \\ \sum x_3x_1 & \sum x_3x_2 & \sum x_3^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} \sum x_1y \\ \sum x_2y \\ \sum x_3y \end{bmatrix} \tag{3.39}$$

Bu matris formüllemesine dayanan üç normal denklemi,

$$\begin{aligned} \left(\sum x_1^2\right)b_1 + \left(\sum x_1x_2\right)b_2 + \left(\sum x_1x_3\right)b_3 &= \sum x_1y \\ \left(\sum x_2x_1\right)b_1 + \left(\sum x_2^2\right)b_2 + \left(\sum x_2x_3\right)b_3 &= \sum x_2y \\ \left(\sum x_3x_1\right)b_1 + \left(\sum x_3x_2\right)b_2 + \left(\sum x_3^2\right)b_3 &= \sum x_3y \end{aligned} \tag{3.40}$$

olarak yazabiliriz. Bunlar b_1, b_2 ve b_3 için çözüldüğünde en küçük kareler çözümlerini temin eden normal denklemlerdir. Geometrik olarak bu denklemlerin her biri bir $Ab_1 + Bb_2 + Cb_3 = d$ düzlemi için denklemi temsil eder. Burada *A*, *B*, *C* ve *d* reel sayılardır. Yine kareler ve çapraz çarpımlarının bu toplamları için bazı uygun sayıları verebiliriz (uygulamada, şüphesiz bunlar gözlemlerden elde edilir). Bu durum veri için üç normal denklemi üretir. Bu denklemler

$$\begin{aligned} 4b_1 + 4b_2 + 2b_3 &= 8 \\ 4b_1 + 6b_2 + 4b_3 &= 10 \\ 2b_1 + 2b_2 + 4b_3 &= 12 \end{aligned} \tag{3.41}$$

olarak ifade edilebilir. (3.37) deki iki sistemi çözmek için yaptığımız gibi yerine koyma metodunu kullanarak bu denklemleri çözebiliriz veya matris cebirini kullanabiliriz. Bu durumda sistemin çözüm kümesi $b_1 = 2.333$, $b_2 = -1.667$ ve $b_3 = 2.667$ olacaktır. Bu çözüm bu veri için tek en küçük kareler çözümüdür. Şimdi çözüm uzayının b_1 için bir, b_2 için bir ve b_3 için bir olmak üzere üç boyuta sahip olması hariç, geometrik yorumumuzu önceki gibi oluşturabiliriz. Üç denklemin her biri bir düzlemi gösterir. Düzlemlerden birini inşa etmek için düzlemin b_1 ekseninin kestiği yeri; yani b_2 ve b_1 in her iki ikisi sıfıra eşit olduğunda b_1 in ne olduğunu belirleyebiliriz. Cevap $b_1 = 2$, yani bu düzlem üzerindeki (2,0,0) noktasıdır. Benzer şekilde birinci satırdaki denklem ile temsil edilen düzlem b_2 eksinini 2 de keser, bu nedenle düzlem üzerindeki ikinci bir nokta (0,2,0) noktasıdır. Son olarak düzlem b_3 eksenini 4 de keser, bu nedenle düzlem üzerindeki başka bir nokta (0,0,4) noktasıdır.

Bu üç nokta bu üç boyutlu uzayda birinci denklem vasıtasıyla temsil edilen düzlemi belirler.

Aynı şekilde üç ekseni kestiği (2.50,0,0), (0,1.667,0) ve (0,0,2.50) yerlerini bularak ikinci satır denklemi için düzlemi ve (6,0,0), (0,6,0), (0,0,3) yerlerini bularak üçüncü satır düzlemini belirleyebiliriz. Bu düzlemlerin ikisi lineer bağımlı olmadığından, onlar birbirlerini keser ve kesişim (ara kesit) bir doğruyu belirler. Denklemlere çözüm bu doğru üzerinde olmalıdır. (3.41) deki üçüncü denklem ilk iki düzleme lineer bağlı değildir bu nedenle bu doğruyu bir noktada kesecek ve bu nokta; bu üç denklem sistemi için bir tek çözümü belirleyecektir. Bu, $\left(2.333, -1.667, 2.667\right)$ kesişim noktası lineer cebir yöntemini kullanarak elde edilen çözümün aynısı olacaktır. Dikkatli bir geometrici düzlemlerin kesişimlerini kullanarak bu çözümü ortaya koyabilecektir. Şüphesiz biz geometrik bakış açısıyla temin edilen görünüm, yani sezgi ile ilgileniyoruz ve böyle geometrik yorumlamaları bu sonuçları hesaplamak için bir araç olarak tavsiye etmemeliyiz. Şimdilik sadece üç boyutlu uzayda bir doğru boyunca kesişen iki düzlemi (üç boyutlu uzay bir oda olarak hayal edilebilir) ve doğruyu kesen başka bir düzlemi görselleştirmemiz gerekiyor. Bu kesişim noktası bir üç boyutlu uzayda şahsına münhasır koordinatları temin eder ve bu nedenle parametre tahminleri için bir tek çözümü verir.

Aşağıda da üçüncü satır denkleminin ilk satır denkleminin yarısı ile ikinci satır denkleminin yarısının toplamına eşit olduğu bir lineer bağımlılığı gösterelim.

$$\begin{aligned} 4b_1 + 4b_2 + 2b_3 &= 8 \\ 4b_1 + 6b_2 + 4b_3 &= 10 \\ 4b_1 + 5b_2 + 3b_3 &= 9 \end{aligned} \tag{3.42}$$

Bu denklem sisteminin bir tek çözümü yoktur. Eğer bu üç denklemin ikisi için düzlemleri oluştursaydık, bu durumda bu denklemlerin herhangi ikisi lineer bağımlı bir küme oluşturmadığından, onlar bir doğru içinde kesişeceklerdi. Bu doğru geriye kalan düzlem üzerinde olacaktır, bu nedenle bu doğru üzerindeki herhangi bir çözüm bu denklem sisteminin bir çözümü olacaktır. Bir tek çözümün var olması için, geriye kalan düzlem diğer iki düzlemin kesişimiyle oluşturulan doğruyu bir noktada kesmiş olmalıdır. Üç veya daha fazla boyutlu bir uzaya bir doğru alışılageldik şekliyle, bir doğru için vektör denklemini kullanarak tamamlanır. Bu denklem bize doğru üzerindeki noktaların tümünün boyutların her biri üzerindeki koordinatlara bağlı olduğu yeri verir. (3.42) deki ilk iki denklem için onların kesişiminin (ara kesitinin) doğrusu doğru için

$$(1 \quad 1 \quad 0)' + k(1 \quad -2 \quad 2)'$$

vektör denklemiyle tanımlanabilir. İkinci iki düzlemin ara kesiti birinci ve üçüncü düzlemlerin ara kesiti olan aynı doğruyla tanımlanabilir. Belirtildiği gibi, geriye kalan düzlem (ara kesitte içerilmeyen, yani hesaba katılmayan düzlem) bize bir tek çözümü bulmaya yardım etmez, çünkü bu düzlem çözümlerin doğrusu ile bir tek noktada kesişmez. Doğru üzerindeki noktaların tümü geriye kalan düzlem üzerinde yatar ve bu nedenle doğru ve düzlem için bir tek kesişim (ara kesit) noktası yoktur. Birinci ve üçüncü denklemi veya ikinci ve üçüncü denklemi kullanabildiğimizden ve bu arakesitlerle oluşturulan doğru aynı olması gerektiğinden, bu durum anlamlı olur. Yani düzlemlerin hepsi onların yüzeyleri üzerinde uzanan bu doğruya sahiptir. Eğer geriye kalan denklemin düzlemi diğer iki denkleme lineer olarak bağlı olmasaydı onun düzlemi ilk iki düzlem tarafından saptanan doğruyu kesecekti. Kısıtlanmış regresyonda, geriye kalan düzlemi doğrultuya değiştirmeye zorlarız ve böylece kısıtlama altında bir tek çözüm buluruz.

$$\begin{bmatrix} 4 & 4 & 2 \\ 4 & 6 & 4 \\ 4 & 5 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

olduğundan (3.42) için sıfır vektör $(1,-2,2)'$ dür. Bu sıfır vektör iki kesişen düzlemle saptanan doğruya paraleldir. Kesişim doğrusu “çözümler doğrusu” olarak adlandırılır çünkü bu doğru üzerindeki herhangi bir nokta rank eksikliği 1 olan denklemlerin kümesini çözer. Şüphesiz bu çözümler lineer bağımsız denklemler için elde edilen tek çözümler değildir. Üç normal denklem ile lineer bağımlılığa göre bir veya birden fazla olabilirlik vardır. Bağımsız değişkenlerin matrisi 2 kadar eksik ranklı olabilir, yani bu üç denklemin ikisinin bir kümesi lineer bağımsız olmayabilir. Başka bir deyişle iki lineer bağımsız sıfır vektör var olabilir. Bu durum 2 kadar eksik ranklı olan bir $X'X$ matrisini üretmek için tasarlayarak seçtiğimiz veri için (3.43) teki normal denklemler ile vuku bulur.

$$\begin{aligned} 4b_1 + 4b_2 + 2b_3 &= 8 \\ 2b_1 + 2b_2 + 1b_3 &= 4 \\ 3b_1 + 3b_2 + 1.5b_3 &= 6 \end{aligned} \tag{3.43}$$

Bu düzlemlerin hiç birisi kesişmez, onların her üçü bir diğeri ile çakışır. Onlar aynı iki boyutlu alt uzaya düşer. Örneğin, düzlemlerin her üçü b_1 eksenini (2,0,0) da keser, çünkü $b_2 = 0$ ve $b_3 = 0$ olduğunda denklemlerin her üçü için $b_1 = 2$ dir. Benzer şekilde, her üç denklem için düzlem b_2 eksenini (0,2,0) da ve b_3 eksenini (0,0,4) de keser. Bu üç düzlemin çakıştığı açıktır. Bu denklemlere çözümler bu "çözümler düzlemi" üzerinde herhangi bir yerde bulunabilir. Bu potansiyel çözümleri çözümlerden herhangi biri (bu düzlem üzerindeki noktalar) artı bir k skaleri çarpı bu düzlemin doğrultularının biri artı bir s skaleri çarpı bu düzlemin diğer bir doğrultusu olarak yazabiliriz. Örneğin,

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} + k \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} + s \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} \tag{3.44}$$

şeklinde yazabiliriz. Eğer $k = 2$ ve $s = -1$ seçilirse bu düzlem üzerindeki nokta $(1, 2, -2)$ olacaktır. Bu nokta bu düzlem üzerindeki herhangi bir noktanın yaptığı gibi (3.43) için bir çözüm olarak çalışır. Beklenildiği gibi k ve s ile çarpılan iki vektör (3.43) için sıfır vektörlerdir, yani

$$\begin{bmatrix} 4 & 4 & 2 \\ 2 & 2 & 1 \\ 3 & 3 & 1.5 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{ve} \quad \begin{bmatrix} 4 & 4 & 2 \\ 2 & 2 & 1 \\ 3 & 3 & 1.5 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{3.45}$$

dır. Bu iki sıfır vektör birbirine lineer bağımlı değildir ve sıfır vektörleri üreten diğer herhangi sıfır vektörler de bu iki sıfır vektöre lineer bağlıdır. Bu durumda sıfır uzayı (0,0,0) orijininden geçen

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + k \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} + s \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} \tag{3.46}$$

ile tanımlananan bir düzlemdir. Denklemlere çözümler, çözümlerin bir düzleminde bulunur ve bu düzlem, bir düzlem olan sıfır uzayına paraleldir. Toparlamak gerekirse bu kısımda

$$X'Xb = X'y$$

normal denklemleri irdelenecektir. Bu denklemlere herhangi bir çözüm lineer bağımlılıklar ve bu nedenle sonsuz sayıda çözümlerin var olduğu durumlarda bile, bir en küçük kareler çözümünü garanti eder. Çözümlerin herhangi biri bir en küçük kareler çözümü ortaya koyar. Bir lineer bağımlılık problemi bir çözümü bulamadığımız problem olmayıp bir tek çözümün var olmadığı problemdir.

Lineer bağımlı olmayan iki değişken durumunda normal denklemler doğrular için iki denklemden oluşur ve bu doğrular iki–boyutlu çözüm uzayında kesişir ve denklemlere bir tek çözümü verir. Üç bağımsız değişken ile, üç normal denklem vardır ve her biri bir düzlem için denklemdir. Bu üç düzlem denklemlere bir tek çözümü veren üç–boyutlu çözüm uzayında bir tek noktada kesişir. Bu sezgisel iki ve üç boyutlu durumların ötesine geçerek bir genelleme oldukça basittir, fakat terminoloji ve görselleştirmeler daha zordur. Dört bağımsız değişken ile dört normal denklem vardır. Her biri üç–boyutlu bir hiperdüzlemi gösterir. Eğer lineer bağımlılıklar yoksa bu dört üç–boyutlu düzlem dört–boyutlu çözüm uzayındaki bir noktada kesişir ve bir tek çözümü ortaya koyar.

Lineer bağımlı iki değişken durumunda iki normal denklemi temsil eden iki doğrunun çakıştığını; onların kesişmediklerini ve bu çakışan doğrular, yani, "çözümler doğrusu" üzerindeki herhangi bir çözümün normal denklemleri çözdüğü görülür. Üç normal denklemin düzlemleri gösterdiği, üç bağımsız değişken durumunda, eğer bağımsız değişkenlerin matrisi 1 kadar eksik ranklı ise, (iki lineer bağımsız denklemin bir kümesi var ise) bu takdirde düzlemlerin ikisi üç – boyutlu çözüm uzayındaki bir doğru içinde kesişir. Bununla beraber, geriye kalan düzlem bu doğruyu bir tek noktada kesmez ve çözümlerin doğrusu düzlem üzerinde yatar. Eğer matrisin rankı 2 kadar eksik ise iki lineer bağımsız sıfır vektör vardır ve her üç düzlem çakışır. Bu çözümler düzlemi üzerindeki herhangi bir nokta normal denklemleri çözer. Bir dört–boyutlu uzayda, bağımsız değişkenlerin matrisi 1 kadar eksik ranklı olduğunda üç–boyutlu hiperdüzlemlerden üçü bir doğru (çözümlerin doğrusu) içinde kesişir fakat geriye kalan üç–boyutlu düzlem çözümlerin doğrusunu bir tek noktada kesmez. Bizim bağlamımızda, (v) sıfır vektörü hepsi sıfırlardan oluşmayan bir vektördür ve bu vektör için $X'Xv = 0$ olduğu ifade edilebilir. Öte yandan, bir matris 1 eksik ranklı olduğunda böyle bir vektör vardır, matris 2 kadar eksik ranklı olduğunda böyle iki lineer bağımsız vektör vardır ve bu iki lineer bağımsız vektör bir düzlem olan bir sıfır uzayını tanımlar. Bu sıfır uzayı çözümlerin düzlemine paraleldir. Genelleştirme vasıtasıyla matris q kadar eksik ranklı olduğunda böyle q sayıda vektör vardır ve bu q sayıda lineer bağımsız sıfır vektör bir q-boyutlu hiperdüzlem olan sıfır uzayını oluşturur. Bu q boyutlu hiperdüzlem çözümlerin q boyutlu hiperdüzlemine paraleldir. Bu tip genelleştirmeler bizim sonuçlarımız için tabanı ortaya koyar. Kendall bu sonuçların bazıları için daha teknik bir tabanı ortaya koymuş, fakat tam sütun rank durumuna odaklanmıştır.

Önce tanımlanmış modeller, yani eksik ranklı olmayan modeleleri göz önüne alalım. İki bağımsız değişkenli bir regresyon modeli tanımlanmış olduğunda, iki boyutlu bir uzayda iki bağımsız normal bir denklem (doğrular için) vardır ve iki doğru denklemlere bir tek çözümü sağlayan bir tek noktada kesişir. Üç bağımsız normal denklemli düzlemler için bir üç – boyutlu uzayda düzlemlerin ikisi bir doğru içinde kesişir ve geriye kalan düzlem bu doğruyu denklemlere bir tek çözümü sağlayan bir tek noktada keser. Dört bağımsız denkleme (üç – boyutlu hiperdüzlemler için) sahip bir dört – boyutlu uzayda hiperdüzlemlerden ikisi bir düzlemde kesişir, bir üçüncü hiperdüzlem bu düzlemi bir doğru içinde keser ve dördüncü hiperdüzlem bu doğruyu denklemlere bir tek çözümü sağlayan bir tek noktada keser. m bağımsız denkleme (her bir denklem bir (m – 1) boyutlu hiperdüzlemi temsil eder) sahip bir m boyutlu uzayda m hiperdüzlem denklemlere bir tek çözümü veren bir tek noktada kesişir.

Bağımsız değişkenlerin matrisi tam sütun ranklı (rank eksikliği yok) olduğunda bağımsız değişkenlerin her biri için bir tek çözümleri bulma; güvenilir (regular) bir regresyon programını veya matris cebirini kullanarak daha basittir. Eğer seçersek, regresyon katsayılarının bir veya daha fazlasını sınırlayabiliriz, hiperdüzlemlerin bir veya daha fazlasını değiştirerek modelin uyumunu büyük olasılıkla indirgemeliyiz, bu nedenle onların kesişimi belirlenen çözümden ziyade farklı bir noktadır. Bu durum bir en küçük kareler çözümü olmayan bir çözümü yaratır. Sınırlamanın modelin uyumunun anlamlı olarak indirgenip indirgenmediğini görmek için biri bunu yapabilir. Bununla beraber, buradaki odaklanmamız sınırlama ile tam belirlenmiş olan modellere bir çözümü vermek için kısıtlamaların kullanıldığı eksik ranklı modellerin geometrisi üzerindedir.

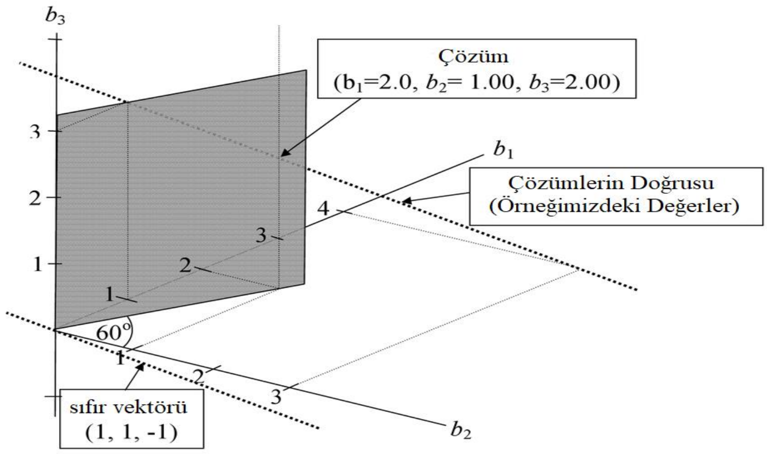
Şimdi bir eksik ranklı modelleri göz önüne alalım. Tam sütun ranktan bir eksik rank durumu yukarıda atıfta bulunulan deneysel örneklerin her birinde açıklanan durumdur.

Bir iki ranklı üç bağımsız değişkenli durumda çözümlerin üzerine düşmesi gereken doğruyu (normal denklemlerin ikisi bir doğru üzerinde kesişir) belirleyebiliriz. Buna çözümlerin doğrusu adını veririz; fakat geri kalan düzlem (denklem) bu doğruyu kesmez (çözümlerin doğrusu bu düzlem üzerinde uzanır). Bu durumda çözümün üzerine düşmesi gereken doğruyu belirleyebiliriz, ancak bu doğru üzerindeki noktayı değil. Bu ikileme kısıtlanmış regresyon çözümü düzlemin doğrultusunu çözümlerin üzerine düşmesi gereken doğruyu kesecek şekilde ayarlamaktır. Bunu yapmak için bir yol bir özel kısıtlamaya dayanan bir genelleştirilmiş tersi kullanmaktır (3.37). Bu durum denklemlerin sistemine (o kısıtlama altında) bir çözümü verir. Biri koyduğu kısıtlama hakkında endişe duymaksızın herhangi bir uygun genelleştirilmiş tersi kullanabilir, fakat o kesin surette bir kısıtlama koyar.

İrdelemeyi daha somut yapmak için, matris rankının iki olduğu üç denklemli bir örnek verebiliriz. Bu değişkenler üzerindeki gözlemlerin değerlerinin her birinden onların ortalamalarını çıkararak, bu analizdeki değişkenlerin tümünü merkezileştiririz. Bunun nedeni bir üç–boyutlu uzayda üç bağımsız değişkenli çözümleri görselleştirmemize izin vermesidir. Bunun yerine, regresyon sabiti için X matrisinde 1 lerin bir sütununu hesaba katabilirdik ve örneğimizde sadece iki bağımsız değişkeni kullanabilirdik. Bu örnek için aşağıdaki $(X'X)b = X'y$ normal denklemlerini kullanalım.

$$\begin{bmatrix} 56 & 42 & 98 \\ 42 & 44 & 86 \\ 98 & 86 & 184 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 350 \\ 300 \\ 650 \end{bmatrix} \tag{3.47}$$

$X'X$ matrisindeki lineer bağımlılık açıktır. Sıfır vektör ($X'X$ ile çarpıldığında sıfır vektörü üreten vektör) (1,1,-1) dir. Sıfır vektörün bu gösterimi bir skalerle çarpımda da tektir. Şekil 3.7 eksenlerin bilinmeyen regresyon katsayılarını temsil ettiği bir üç – boyutlu uzayda bu problemi gösterir. b_1, b_2 ve b_3 için eksenlerle yaratılan üç boyutlu uzayda sıfır vektörünü (1,1,-1) ve (0,0,0) noktalarından uzayarak geçen bir doğru olarak gösterebiliriz. Bu doğru en soldaki koyu nokta noktalı doğruyla gösterilir. En sağdaki koyu nokta noktalı doğru çözümlerin doğrusu olarak etiketlenir ve kısıtlanmış regresyona çözümleri üzerinde taşıması gereken doğrudur. (3.47) deki veri için çözümlerin doğrusu $b_1 - b_2$ düzlemini (4,3,0) da keser, çünkü $b_3 = 0$, $b_1 = 4$ ve $b_2 = 3$ her üç düzlem için doğru çözümü verir. Benzer şekilde çözümlerin doğrusu $b_1 - b_3$ düzlemini (1,0,3) de keser. Bu noktaların herhangi birini bir (b_c) çözümü gibi seçerek ve k çarpı (v) sıfır vektörünü ona ekleyerek, bir doğru $(b = b_c + kv)$ için vektör denklemini kullanarak bu doğruyu tanımlayabiliriz: $(4,3,0)' + k(1,1,-1)'$. Bu ise çözümlerin doğrusu ve sıfır vektörünün paralel olduğunu, onların aynı doğrultuyu paylaştığını garanti eder. Çözümler doğrusu aynı zamanda (3.47) deki normal denklemler vasıtasıyla tanımlanan düzlemlerden ikisinin arakesitini de temsil eder. Geriye kalan normal denklem (düzlem) çözümlerin doğrusunu kesmez çünkü çözümler doğrusu onun üzerinde uzanır. Mesele çözümler doğrusu üzerindeki hangi çözümleri seçeceğimizdir. Kısıtlanmış regresyonu açık bir şekilde veya herhangi bir genelleştirilmiş tersi kapalı bir şekilde kullanarak onu seçebiliriz. Geriye kalan düzlemin doğrultusunu sınırlayarak bir çözüm elde edilir. Genel olarak sınırlanan çözüm düzlemi denklemlere bir çözümü veren bir noktada çözümlerin doğrusunu keser. Eğer düzlem çözümlerin doğrusunun doğrultusunda olarak sınırlanırsa bu durumda doğruyu kesmeyeceğinden, terimi genel olarak kullanırız. Örneğin, bu veri için $b_1 = b_2$ kısıtlamasını koymak düzlemin konumlandırmasını değiştirmeyecek ve bir çözüm üretmeyecektir. Diğer taraftan eğer $b_1 = -b_3$ sınırlamasını konulursa diğer birçok kısıtlamalar olacağı gibi $b_1 = -b_2$, $b_1 = b_3$ veya $5b_1 = b_2$ bir çözüm üretecektir. Şekil 3.7 de $5b_1 = b_2$ kısıtlaması altında kısıtlanmış çözüm düzlemini gösterelim. Bu düzlem $b_2 - b_1$ eksenine göre 0.5 lik bir eğime (b_1 üzerindeki 1 lik bir artış b_2 üzerinde 0.5 lik bir artışla ilgili) sahiptir. Kısıtlanmış düzlem gölgelenmiştir ve çözümlerin doğrusunu (2,1,2) de keser. Bu ise $5b_1 = b_2$ kısıtlaması altında çözümdür.



Şekil 3.7 Bir Lineer Bağımlılık İle Üç–Boyutlu Uzayda Kısıtlanmış Regresyonun Geometrik Görünümü

Burada kısıtlama $0.5b_1 = b_2$ dir. Bu durumda sıfır vektörü orijini keser ve çözümler doğrusu sıfır vektörüne paraleldir. Kısıtlanmış çözüm düzlemi çözümler doğrusunu (2.0, 1.0, 2.0) da keser. Bir dikkatli geometrici bu üç boyutlu durumda bu kısıtlanmış regresyona grafiksel olarak bir çözüm bulabilir. Bu durumu ve daha yüksek boyutu içeren diğer durumlar için, Mazumdar ve Ark. (3.37) özel kısıtlamalara karşılık gelen genelleştirilmiş tersleri nasıl kullanabildiğimizi gösterir. Örneğin, Moore–Penrose terse karşılık gelen çözüm (1.67, 0.67, 2.33) tür. Moore–Penrose sıfır vektöre ortogonal olan kısıtlanmış çözüme karşılık gelir. $(1.67, 0.67, 2.33)(1,1,-1)' = 0$ ve bir kısıtlanmış regresyon programında $b_1 = b_3 - b_2$ kısıtlamasını kullanarak veya Moore–Penrose tersi kullanarak, Mazumdar ve Ark. (3.37) nin sistemini kullanarak yerine getirilebilir. Grafiksel olarak ilerleyerek, kısıtlanmış düzlem $(1,1,-1)$ sıfır vektörüne ortogonal olmalıdır ve çözümler doğrusunu (1.67, .67, 2.33) de kesmelidir.

Şüphesiz rank eksikliğinin bir ve dört bağımsız değişkeninin var olduğu durum için bir şekil çizmek daha zordur. Bu durumda dört üç–boyutlu düzlemi gösteren dört denklem vardır. Çözümlerin doğrusu bu hiperdüzlemlerin üçünün arakesiti ile belirlenir ve normal denklemlere çözümlerin birini bulduğumuzda çözümlerin doğrusunu $b = b_c + kv$ olarak yazabiliriz. Çözümler doğrusu sıfır vektörüne paraleldir. Maalesef çözümler doğrusu geriye kalan hiperdüzlemi kesmez. Bir örnek olarak, aşağıdaki $\left(X'X\right)b = X'y$

$$\begin{bmatrix} 50 & 40 & 20 & 55 \\ 40 & 90 & 10 & 70 \\ 20 & 10 & 80 & 55 \\ 55 & 70 & 55 & 90 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} -25 \\ -180 \\ 185 \\ -10 \end{bmatrix} \tag{3.48}$$

normal denklemlerini kullanalım. Bu durumda (3.48) için sıfır vektör $\left(1,1,1,-2\right)$ dir. Sınırlama $b_1 = -b_3$ olduğunda, çözüm vektörü $\left(-1,-4,1,3\right)'$ olacaktır. Bu çözüm vektörü (3.48) de kesin olarak çalışır; örneğin birinci satır için, $-1\times 50 - 4\times 40 + 1\times 20 + 3\times 55 = -25$ dir. Benzer şekilde (3.48) in geriye kalan satırları için de sağlanır. Bu nedenle, çözümler doğrusu $\left(-1,-4,1,3\right)' + k\left(1,1,1,-2\right)'$ olarak yazılabilir. Bu tam olarak çözümler doğrusunu belirtir ve bir ihtimal bir dört–boyutlu uzaydaki bir doğru olarak hayal edilebilir. Geriye kalan üç–boyutlu hiperdüzlemi hayal etmek daha zordur. Çünkü bu durum denklemlere bir çözüm elde etmek için dört–boyutlu uzayda sınırlanan bu düzlemin konumlandırılmasıdır. Lineer bağımlılık ile normal denklemlerin biri ile temsil edilen bu üç–boyutlu hiperdüzlem çözümler doğrusunu (diğer üç hiperdüzlem tarafından belirlenen) kesmez. Bu hiperdüzlemi çözümler doğrusunu bir tek noktada kesmeye zorlamak için bir kısıtlama kullanmalıyız. Bu durumda, eğer $b_1 = -b_3$ kısıtlamasını kullanırsak, kısıtlanan hiperdüzlem $b_1 - b_3$ üzerinde –1 lik bir eğime sahiptir. b_1 üzerinde 1 lik bir artış, b_2 üzerinde –1 lik bir azalış ile ilgilidir (burada, hiperdüzlemin dört–boyutlu çözüm uzayından (0,0,0,0) noktasından karşıdan karşıya (çaprazlama) geçmesi gerektiğine dikkat edelim). Konumlandırmadaki bu değişim bu hiperdüzlemi çözümler doğrusunu bir tek noktada kesmeye sınırlar. Yine, çözümler doğrusunu kesmeyen bir hiperdüzlemi veren bir sınırlama koyabiliriz. Bu durumda $b_1 = b_3$ alabiliriz ve hiperdüzlem doğruyu kesmeyecektir. Bununla beraber, genel olarak hemen hemen tüm kısıtlamalar için hiperdüzlem, çözümlerin doğrusu ile kesişecektir. Buradan, m boyuta genelleştirme basittir. Bu durumda m sayıda

$(m-1)-$boyutlu hiperdüzlemi gösteren *m* denklem vardır. Çözümlerin doğrusu (*m*–boyutlu uzayda bir doğru) bu hiperdüzlemin $m-1$ tanesinin arakesiti ile belirlenir. Çözümler doğrusu için onun vektör denklemi *b*, b_c ve v her biri *m*–elemana (bileşene) sahip olmak üzere, $b = b_c + kv$ dir. Geriye kalan hiperdüzlem çözümlerin doğrusunu kesmez. Geriye kalan hiperdüzlem üzerine koyacağımız tek kısıtlama, genel olarak *m*–boyutlu uzayda onu yeniden konumlandırır ve *m* sayıda denklemin sistemine bir çözümü veren çözümler doğrusunu bir tek noktada kesen kısıtlanmış hiperdüzlemi ortaya koyar.

Şimdi de iki eksik ranklı modeller durumunu göz önüne alalım. *X* matrisi tam sütun ranktan iki eksik ranklı olduğunda bir üç–boyutlu uzaydaki çözümü görselleştirmek yine mümkündür. Bunu yapmak için yeni bir $\left(X'X\right)b = X'y$,

$$\begin{bmatrix} 144 & 72 & 216 \\ 72 & 36 & 108 \\ 216 & 108 & 324 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 792 \\ 396 \\ 1188 \end{bmatrix} \tag{3.49}$$

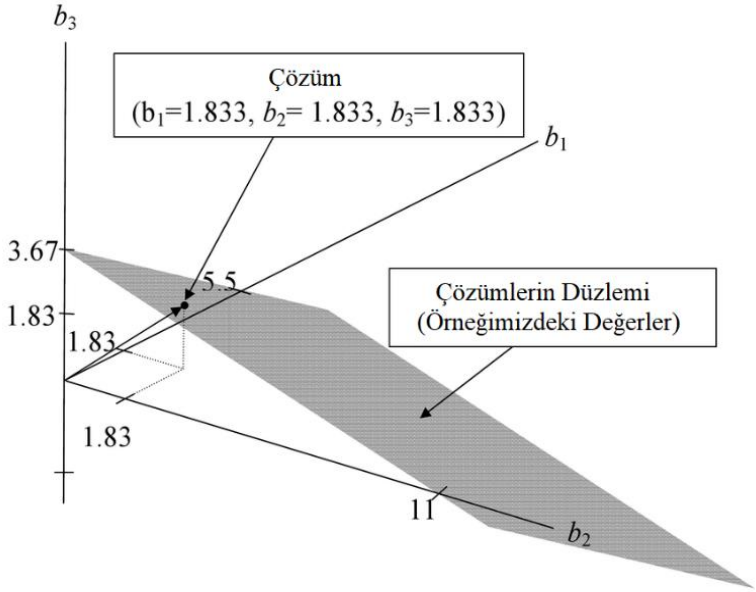
normal denklemler kümesini uygulamaya koyalım. $X'X$ matrisindeki lineer bağımlılıklar açıktır. Bu durumda iki lineer bağımsız sıfır vektör vardır. Bunlar $(1,-2,0)$ ve $(1,1,-1)$ dir. Bu vektörler sıfır uzayını tanımlar ki bu durumda sıfır uzayı bir düzlem (iki–boyutlu bir hiperdüzlem: bir düzlem) dir. Olası kısıtlanmış sonsuz çözümlerin biri için bir sefere mahsus çözeriz. Bir düzlem için vektör denklemini kullanarak çözümlerin düzlemini yazmak basit bir meseledir. Çözüm $b = b_c + kv_1 + sv_2$ olarak tanımlanan düzlem üzerinde bulunmalıdır. Burada b_c herhangi özel kısıtlanmış çözümlerdir, *b* olası çözümlerin tümünü gösterir. *k* ve *s* skalerlerdir ve v_1 ve v_2 iki bağımsız sıfır vektördür. (Bu iki lineer bağımsız sıfır vektörü göstermek için başka yöntemler vardır, ancak diğer tüm yöntemler bu iki sıfır vektöre lineer bağlıdır). Bu durumda, üç normal denklem vasıtasıyla belirlenen her üç düzlemin biri diğeriyle çalışır ve çözümlerin düzlemini oluşturur. Bu 2 kadar rank eksikliği durumunda bir çözümü belirlemek için iki kısıtlama gerekir.

Kısıtlamalardan birine iki düzlemden birinin konumlandırılmasını değiştirmek olarak bakabiliriz, bu nedenle o genel olarak birinci kısıtlama altında bir doğruyu üreten diğer iki düzlemin biri ile kesişir. İkinci kısıtlama ise genel olarak çözümlerin düzlemi ile kesişecek şekilde bu doğruyu konumlandırır. Terminolojinin bir kadar eksik ranktan iki kadar eksik ranka hareket ettiğimiz gibi olduğuna dikkat edelim. Önceki kısımdaki çözümler doğrusu şimdi çözümlerin düzlemidir. Çözümlerin düzlemi sıfır uzayına paralel alt uzaydır (iki–boyutlu) ki onun üzerinde çözüm yatmalıdır. Bu durumda (3.49) daki veri için normal denklemlere bir çözüm $\left(5.5, 0, 0\right)$ şeklinde olup çözümler düzlemi $\left(5.5, 0, 0\right)' + k\left(1, -2, 0\right)' + s\left(1, 1 - 1\right)'$ olarak tanımlanabilir. Lineer kısıtlamaları kullanan çözümlerin tümü bu düzlemin üzerine düşecektir. Dolayısıyla sorun da burasıdır. Bu durumda kısıtlanmış regresyon kullanıldığında cevap çözüm üzerine koyduğumuz kısıtlamalara bağlıdır.

Şekil 3.8 de karmaşıklıktan kaçınmak için, sıfır uzayını (çözümlerin düzlemine paralel olan ve (0,0,0) dan geçen bir düzlem) gösterilmedi. Çözümlerin düzlemi Şekil 3.8 de gösterilir ve (5.5, 0, 0), (0,11,0) ve (0, 0, 3.67) noktalarından geçer. Bu noktaların hepsi çözümler düzlemi üzerinde yer alır ki bu, bu düzlem için vektör denklemi kullanarak doğrulanabilir. Bu durumda $X'X$ tam sütun ranktan iki kadar eksik ranklı olduğundan, çözüm üzerine iki kısıtlama koymalıyız. Şekil 3.8 de $b_1 = b_2$ ve $b_2 = b_3$ kısıtlamalarını kullanırız. Birlikte onlar çözümü eksenlerin her biri ile eşit açılı olan (45 derecelik açıları oluşturan) bir doğru üzerinde olmaya sınırlar. Bu iki kısıtlamayı kullanarak çözüm (1.833, 1.833, 1.833) olarak bulunur ki bu (0,0,0) dan okun çözümlerin düzlemini kestiği yer olarak Şekil 3.8 de gösterilir. Bu çözümün (3.49) daki veri için çalıştığını göstermek kolaydır. Bu bir en küçük kareler çözümünü verecektir, ancak bir doğruyu çözümlerin düzlemini kesmeye zorlayan iki kısıtlamanın farklı kombinasyonlarına dayanan diğer çözümlerin sonsuz bir sayısı da bir en küçük kareler çözümünü verir.

2 kadar eksik ranklı $X'X$ ile bir dört–boyutlu uzayda iki lineer bağımsız sıfır vektör vardır ve bu durumda sıfır uzayı bir düzlemdir. Normal denklemlerin her biri üç boyutlu bir hiperdüzlemi temsil eder ve onlardan ikisi sıfır uzayına paralel olan çözümler düzlemini belirlemek için kesişir. Geriye kalan iki üç–boyutlu hiperdüzlem çözümler düzlemini oluşturmak için kesiştirilen iki hiperdüzleme lineer bağımlıdır. İki geri kalan düzlemin üzerine bir kısıtlama koymak genel olarak onun diğer geri kalan hiperdüzlemle kesişimine götürecek ve bir düzlemi (bir iki–boyutlu

hiperdüzlemi) belirleyecektir. Bu düzlem çözümlerin düzlemini kesmez. İkinci kısıtlama bu düzlemin genel olarak çözümlerin düzlemini bir tek noktada kesecek olan doğrultusunu belirleyecektir.



Şekil 3.8 İki Lineer Bağımlılık İle Üç–Boyutta Kısıtlanmış Regresyonun Geometrik Görünümü

Burada, kısıtlamalar $b_1 = b_2$ ve $b_2 = b_3$ tür. Çözümler uzayı (çözümler bu uzay üzerine düşmeli) sıfır uzayına (gösterilmedi) paraleldir. Kısıtlanan çözüm doğrusu çözümler düzlemini (1.833, 1.833, 1.833) noktasında keser. Bir dört–boyutlu uzayda iki düzlem genel olarak bir noktada kesişir. Neyse ki, hem sıfır uzayı hem de düzlemler olan çözüm uzayı onlar bir–dört boyutlu uzay içine yerleştirilmiş olsa bile akla uygun bir şekilde sezgiseldir.

Boyutların sayısını artırdığımız zaman, çözümler aynı geometrik kalıba uyar. m sayıda normal denklemin her biri $(m-1)-$boyutlu hiperdüzlemi temsil eder. Sadece iki lineer bağımsız sıfır vektör var olduğu sürece, çözümlerin bir $b = b_c + kv_1 + sv_2$ düzlemi var olacaktır. Çözümlerin bu düzlemi hiperdüzlemlerin $m-2$ tanesinin (düzlemlerin ikisinden gayri hepsinin) kesişimi ile belirlenir. Geriye kalan iki düzlem diğer her biri diğeri ile kesiştirilen $m-2$ tane hiperdüzleme lineer bağlıdır. Bu $(m-1)-$boyutlu hiperdüzlemleri her biri diğeri ile kesişecek şekilde sınırlamamız gerekir. Kesişim bir $(m-2)-$boyutlu hiperdüzlemi ortaya koyar ve ikinci kısıtlama bu hiperdüzlemin doğrultusunu sınırlamak için kullanılır. Genel olarak bu kısıtlanmış $(m-2)-$boyutlu hiperdüzlem ve çözümlerin iki–boyutlu düzlemi m–boyutlu çözüm uzayındaki bir noktada kesişecek ve böylece konulan kısıtlamalar altında denklem sistemin bir tek çözümünü verecektir.

Genel durumda, eğer bağımsız değişkenlerin bir m sütunlu matrisine sahip isek, m sayıda normal denklem (her biri için bir satır) vardır. Bu durumda her bir denklem $(m-1)-$boyutlu bir hiperdüzlemi temsil eder. Eğer m sütunlu matris d kadar eksik ranklı ise, bu takdirde sıfır uzayı d–boyutludur ve çözümlerin hiperdüzlemi d–boyutludur. Bu durumda çözümün hiperdüzlemi

$$b = b_c + kv_1 + sv_2 + \cdots + qv_d$$

vasıtasıyla temsil edilebilir. Çözümlerin bu d–boyutlu hiperdüzlemi, hiperdüzlemlerin $m-d$ tanesinin kesişimi ile belirlenir. Denklemlerin sistemini çözmek için, geriye kalan d tane hiperdüzlemler arasındaki kesişimi üretmek için bu kısıtlamalardan $d-1$ tanesini kullanırız. Bu kesişimler $(m-d)-$boyutlu hiperdüzlemi ortaya koyar ve son kısıtlama bu $(m-d)-$boyutlu hiperdüzlemi konumlandırır. Bu iki hiperdüzlem (çözümlerin d–boyutlu hiperdüzlemi ve kısıtlanan $(m-d)-$boyutlu hiperdüzlem), genel olarak, m–boyutlu çözüm uzayında bir tek noktada kesişir. Bu nedenle, onlar koyulan kısıtlamalar altında denklemlerin sistemine bir tek çözüm verir.

Bağımsız değişkenlerin matrisi eksik sütun ranklı olduğunda, bir normal denklem sistemine bir çözüm bulmayı özel kısıtlamalar koyarak ele aldık. Vurgumuz, her bir satır bir $(m-1)-$boyutlu hiperdüzlemi göstermek üzere, normal denklemlerin satırlarında dayandırıldı. Sıfır uzayı ile aynı boyuta sahip ve ona paralel olan çözümlerin hiperdüzlemini görselleştirmeye yardım edecek sıfır vektörlerini kullandık. Çözümlerin d–boyutlu hiperdüzlemi normal denklemlerin satırlarının her biriyle temsil edilen $(m-1)-$boyutlu hiperdüzlemlerin $m-d$ tanesinin arakesiti (kesişimi) vasıtasıyla yaratılır. Her ne kadar normal denklemlere çözümlerin sonsuz bir sayısı var ise de onların bu uzayda yattığını biliyoruz. Geriye kalan d tane

$(m-1)-$ boyutlu hiperdüzlemin konumlandırılmasını uygun bir şekilde kısıtlayarak, normal denklemlere kısıtlamalar altında tek olan bir çözümü bulabiliriz.

Hesaplayarak kısıtlamaya dayanan bir genelleştirilmiş ters yardımıyla bu kısıtlanmış çözümleri bulabiliriz (3.37). Özel bir kısıtlama ile bir genelleştirilmiş tersi kasıtlı olarak üretmediğimizde bile, herhangi bir genelleştirilmiş tersin bir kısıtlanmış çözümü ürettiğine dikkat etmek önemlidir. Bu anlamda, rank eksikliği olan bu denklemleri çözmek için kullanılan genelleştirilmiş terslerin geometrisi kısıtlanmış regresyonda kullanıldığının aynısıdır. İrdelememiz; normal denklemlerin satırları ve onların arakesitleri üzerine odaklanarak satır bakış açısından kısıtlanmış regresyonun geometrisi üzerine odaklanmıştır.

Bazı bakımlardan, bu bakış açısı boyutların sayısı büyük olduğunda sütun bakış açısından daha zor olabilir, fakat bu satır bakış açısını alarak kazandırılacak geometrik sezgiler vardır.

Özellikle bir ve iki kadar rank eksikliği durumlarında çözümler doğrusunu ve çözümler düzlemini düşünmek sezgiseldir. Satır geometrisi, büyük bir genişletme için, satır işlemleri arasındaki kısıtlanmamış kesişimlerin çözümler hakkında bildiğimiz, onun bu uzaya, yani sıfır uzayına paralel uzaya düşmesi gerektiğini ortaya koyar. Onların tümünün her birinin diğeriyle kesiştiği (eğer birden daha fazla var ise) böyle bir yöntemde kısıtlamaları geriye kalan hiperdüzlemlerin düzenlenmesi olarak düşünmek için yararlıdır. Bu kısıtlanmış arakesitlerden (kesişimlerden) (geriye kalan birden daha fazla hiperdüzlem var olduğunda) yaratılan hiperdüzlem çözümlerin hiperdüzlemiyle kesişecek şekilde konumlandırılır. Bu kesişim uygulanan kısıtlamalar altında normal denklemlere bir çözümü üretir. Bunlar genelleştirilmiş tersler kısıtlanmış regresyon çalışmasının iç yüzünün nasıl olduğunu anlatmaya yardımcıdır.

Analizde ortaya koyulanın ne olduğunun iç yüzünü anlamayı kazanmamıza yardım etmek için bu geometri özel bir probleme nasıl uygulanabilir? Bir en küçük kareler çözümünü üretmek için ekseriye kısıtlanmış regresyon ve genelleştirilmiş tersleri kullanarak çözülen bir rank eksikliğinin bir örneği olarak Yaş – Periyot – Grup modelini kullanabiliriz. Geometri bu bir kadar rank eksikliği durumunda problemin ne olduğunu ortaya koyar. Bağımsız değişkenlerin biri hariç tümünün bir kümesi lineer bağımsızdır. Normal denklemlerin biri hariç hepsi için, normal denklemlerin kesişimi bir doğruyu, yani çözümler doğrusunu oluşturur. Geriye kalan normal denklem bir hiperdüzlemle temsil edilebilir. Ancak bu hiperdüzlem çözümler doğrusunu bir noktada kesmez. Kısıtlanmış regresyon ve genelleştirilmiş tersler bu hiperdüzlemin konumlandırılmasını değiştirir. Bu nedenledir ki o, çözümler doğrusunu keser ve çözümler doğrusu üzerindeki çözümlerden birini ortaya koyar.

Bu çözüm bir en küçük kareler çözümüdür. Bazen kısıtlanmış regresyonda koyduğumuz bir kısıtlama bir çözümü ortaya koyma bakımından çalışmaz. Kısıtlama lineer bağlı hiperdüzlemin doğrultusunu değiştirmediğinden ve bu nedenle çözümler doğrusunu kesmediğinden bu durum vuku bulur. Bu çözümün kısıtlamaya bağlı olduğunu hatırlamak önemlidir ve böyle bir kısıtlamayı kullanan herhangi birine teorisini ve asıl mülahazalarını bu taban üzerinde yapmasını öneririz. Çözümler doğrusu veriden bildiğimiz şeydir. Bu doğruyu veriden belirleyebiliriz. Bununla beraber, çoğu zaman onu göz önüne almayız. Bu doğru Yaş – Periyot – Grup modeli (3.35) için diğer teşhis edilen karakteristikleri ve bu nedenle tahmin edilebilen fonksiyonlar denilen karakteristikleri elde etmek için bu doğru kullanılabilir.

Hiperdüzlemlerin arakesitini belirlemek için altı yardımcı nokta söz konusudur. Bu kısımda m sayıda bağımsız değişken ve m sayıda normal denklemle ilgilenilir. Sunumun kolaylığı için değişkenleri sapma puanı biçiminde olduklarını varsayalım. m–boyutlu uzayda diğer hiperdüzlemler ile kesişme ile oluşturulan hiperdüzlemleri saptamada yardımcı olabilen altı noktayı aşağıdaki gibi verebiliriz. İlk beş nokta hiperdüzlemlerin tümünün doğal olarak kesiştiği lineer bağımlılıkların olmadığı durumlarla ilgilidir.

1. Bir m–boyutlu çözüm uzayında m sayıda hiperdüzlem (normal denklemleri temsil eden) vardır.
2. Normal denklemlerin her biri $(m-1)-$ boyutlu bir hiperdüzlemi temsil eder.
3. İki lineer bağımsız $(m-1)-$ boyutlu hiperdüzlemi gösteren hiperdüzlemler kesiştiğinde arakesit $(m-2)-$ boyutlu bir hiperdüzlemdir. Normal denklemlerin her birinin bir düzleme, yani bir (3 – 1) – boyutlu hiperdüzlemi gösterdiği ve iki düzlemin arakesitinin bir doğru, yani bir (3 – 2) – boyutlu hiperdüzlem olduğu uzayı düşününüz.

4. Bir $(m-p)-$boyutlu hiperdüzlemin bir $(m-s)-$boyutlu hiperdüzlem ile arakesiti bir $(m-(p+s))-$boyutlu bir hiperdüzlemdir. Bir dört–boyutlu uzayda iki düzlemin, yani $(4-2)-$boyutlu hiperdüzlemlerin arakesiti bir noktadır, yani bir $(m-(2+2))$ veya $(4-4)-$boyutlu hiperdüzlemdir.
5. Bir *m*–boyutlu çözüm uzayında $p+s=m$ boyutlu hiperdüzlemler (eğer onlar kesişirler ise) bir noktada kesişir. Örneğin, $p+s=m$ olduğunda $(m-s)-$ boyutlu bir hiperdüzlem, genel olarak $(m-p)-$boyutlu bir hiperdüzlemi keser. Lineer bağımlı denklemler ile ilgili problemimiz onlar kesiştirmek için kısıtlanmadıkça hiperdüzlemlerin tümünün bir diğeriyle kesişmediğidir.
6. q kadar lineer eksiklik durumunda, hiperdüzlemlerin $m-q$ tanesi $m-(m-q)$–boyutlu bir hiperdüzlemi yaratarak kesişir. Yani, kesişimlerle yaratılan hiperdüzlem *q*–boyutludur ve sıfır uzayı ile aynı boyuta sahiptir. Eksik rank durumunda, geriye kalan q tane hiperdüzlemi her biri diğerini kesmek ve çözüm uzayını kesmek için sınırlarız. Onların her biri diğeri ile bir $(m-q)-$boyutlu hiperdüzlemde kesişir. Yukarıdaki bilgileri kullanarak, bu iki düzlemin her biri diğeriyle kesişmek için sınırlandırıldığında sonuç $\left[m-((m-q)+q)\right]-$boyutlu bir hiperdüzlemdir. Yani, bir 0 – boyutlu hiperdüzlem ya da bir noktadır. Bu; kısıtlanmış regresyonun geometrik olarak nasıl çalıştığını görmek için bir yöntemdir.

3.5 Genel Lineer Modellerde Genelleştirilmiş Tersin Kullanılması

Bir matrisin genelleştirilmiş tersi olarak bilinen matematiksel kavram ilk olarak 1920 de E.H Moore tarafından ortaya konulmuştur. Moore 1935 de genelleştirilmiş tersin özelliklerinin ilk basılmış sistematik incelemesini vermiştir.

Keyfi bir A matrisinin Moore–Penrose genelleştirilmiş tersi aşağıdaki denklemlerin $G=A^{+}$ çözümü olarak Penrose tarafından tanımlanmıştır.

$$AGA=A \tag{3.50}$$

$$GAG=G \tag{3.51}$$

$$(AG)' = AG \tag{3.52}$$

$$(GA)' = GA \tag{3.53}$$

Bu denklemlere eşdeğer olan iki denklemle keyfi bir matrisin genelleştirilmiş tersini tanımlamak uygundur (Ackley, 1967). Bu denklemler (3.52) denklemini (3.51) denkleminde ve (3.53) denklemini (3.50) denkleminde yerine koyarak oluşturulabilir. Bu denklemler sırasıyla

$$GG'A' = G \tag{3.54}$$

ve

$$AA'G' = A \tag{3.55}$$

dır.

Genelleştirilmiş tersin, bu çalışmanın istatistiksel görünümlerini keşfetmede yararlı olarak birkaç özelliği vardır. Bu özellikler, uygun olduğunda, özelliğin ayrıntılı bir keşfinin bulunabildiği kaynak referanslar ile aşağıda ifade edilir.

Özellik (1). Bu dört denklem ile belirlenen genelleştirilmiş ters tektir [Penrose (3)].

Özellik (2). $A^{++} = A$ dır [Penrose (3)].

Özellik (3). $A^{+'} = A'^{+}$ dır [Penrose (3)].

Özellik (4). $A^{+} = A'A^{+'}A^{+}$ dır [Penrose (3)].

Özellik (5). $A^{+}A$, AA^{+}, $I - A^{+}A$ ve $I - AA^{+}$ nın her biri idempotentdir.

Özellik (6). $rank\left(A^{+}\right) = rank\left(A\right) = rank\left(A^{+}A\right) = rank\left(AA^{+}\right) = iz\left(A^{+}A\right)$ dır. [Penrose (3)].

Özellik (7). $(AB)^{+} = B^{+}A^{+}$ olması için gerek ve yeter şart $A^{+}A$ ve BB' nün değişmeli ve $A'A$ ve BB^{+} nın değişmeli olmasıdır [Greville (4)].

Özellik (8). $AXB = C$ nin X için bir çözüme sahip olması için gerek ve yeter şart $AA^{+}CB^{+}B = C$ olmasıdır. Bu durumda Z keyfi olmak üzere, genel çözüm $X = A^{+}CB^{+} + Z - A^{+}AZBB^{+}$ şeklindedir [Penrose (3)].

Özellik (9). Eğer A matrisi p ranklı $n \times p$ tipinde bir matris ise, bu takdirde $(A'A)^{+} = (A'A)^{-1}$ dir.

İspat: $p \times p$, p ranklı bir $p \times p$ matris olduğundan, $(A'A)$ bir terse (normal terse, yani bildiğimiz terse) sahiptir. Bu durumda $(A'A)^{-1}$ in $(A'A)$ nın genelleştirilmiş tersi için verilen denklemleri sağladığı doğrulanabilir. Genelleştirilmiş ters tek olduğundan, $(A'A)^{+} = (A'A)^{-1}$ dir.

Özellik (10). Eğer A matrisi p ranklı $n \times p$ tipinde bir matris ise, bu takdirde $A^{+}A = I$ dır.

İspat: Eğer A matrisi p ranklı $n \times p$ tipinde matris ise, bu takdirde matris teorisinden, A nın bir sol terse sahip olduğunu biliyoruz.

Tanımlanan denklemlerde yerine koyarak, A^{+} nın sol ters olup olmadığı gösterilebilir. p ranklı bir $p \times n$ matrisi için, sağ ters benzer şekilde, $AA^{+} = I$ olduğu sonucunu ortaya koyar.

Özellik (11). Eğer bir A matrisi tam ranklı ise ve $A = (A_1 \vdots A_2)$ olacak şekilde parçalanmış ise, bu takdirde

$$A^{+} = \begin{bmatrix} A_1^{+} - A_1^{+}A_2\left[\left(I - A_1A_1^{+}\right)A_2\right]^{+} \\ \left[\left(I - A_1A_1^{+}\right)A_2\right]^{+} \end{bmatrix}$$

dır [Cline (5)]. Özellikle eğer, $A_1'A_2 = 0$ ise, bu takdirde, $A_1^{+}A_2 = \left(A_1^{+}A_1^{+\prime}A_1'\right)A_2 = 0$ olduğundan,

$$A^{+} = \begin{bmatrix} A_1^{+} \\ A_2^{+} \end{bmatrix}$$

olduğu kolayca görülür.

3.5.1 Tahmin

$E(e) = 0$ ve $E(ee') = \sigma^2 I$ olmak üzere,

$$Y = X\beta + e$$

genel lineer istatistiksel model için bazı varsayımları yapma, bilinmeyen parametrelerin vektörünün ve parametrelerin fonksiyonlarının tahmini için bize bir taban sunar. Bu niceliklerin tahmini gözlemler, yani *Y* ve bilinen sabitlerin matrisi, yani *X* arasında mevcut olan ilişkilerin tahminine eşdeğerdir. Daha evvel gösterildiği üzere, $(Y - X\beta)'(Y - X\beta)$ yı minimum yapmak için en küçük karelerin kullanımı

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

normal denklemlerini verir.

X in rankını dikkate almaksızın, (4) ve (8) özellikleri bize, β nın en küçük kareler tahmin edicisi için,

$$\hat{\beta} = (X'X)^{+} X'Y + \left[I - (X'X)^{+}(X'X)\right]Z = X^{+}Y + \left(I - X^{+}X\right)Z$$

genel çözümünü verir. Burada, *Z* bir keyfi $p \times 1$ vektörüdür.

Eğer *X* tam ranklı ise, bu takdirde (9) ve (10) özellikleri yardımıyla

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

olduğu görülür. Bununla beraber, genel çözüm $rank(X) = r \leq p$ için de sağlandığından tam rank durumunu ayrı tutmayacağız.

En küçük kareler tahmini σ^2 nin bir tahminini doğrudan doğruya ortaya koymaz. Bununla beraber, σ^2 nin tahminini $\hat{\beta}$ ya dayandırarak, σ^2 nin bir yansız tahmini

$$\hat{\sigma}^2 = \frac{\left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right)}{n - r} = \frac{Y'\left(I - XX^{+}\right)Y}{n - r}$$

dir. Öncelikle ilgilenilen $\hat{\beta}$ tahmin edicisinin ortalama vektörü ve varyans–kovaryans matrisidir. *Y* nin $X\beta$ ortalama vektörüne ve $\sigma^2 I$ varyans–kovaryans matrisine sahip olduğunu hatırlayarak,

$$E\left(\hat{\beta}\right) = E\left[X^{+}Y + \left(I - X^{+}X\right)Z\right] = X^{+}X\beta + \left[I - X^{+}X\right]Z$$

ve

$$Var\left(\hat{\beta}\right)=E\left[\left(X^{+}Y-X^{+}X\beta\right)\left(X^{+}Y-X^{+}X\beta\right)'\right]=E\left[X^{+}ee'X^{+\prime}\right]=\sigma^{2}\left(X'X\right)^{+}$$

dır. Genel olarak $\hat{\beta}$ nın tek olmadığı ve bir yansız tahmin edici olmadığı belirtilir. Bununla beraber, eğer β vektörünü keyfi Z vektörü olarak seçersek $\hat{\beta}$ tahmin edicisi yansızdır. Eğer hataların vektörünün dağılımını $e \sim MVN\left(0,\sigma^{2}I\right)$ olarak belirlersek, en çok olabilirlik tahmini yönteminin kesinlikle β nın en küçük kareler tahmini ile aynı sonuçlara götürdüğü gösterilebilir.

V keyfi bir pozitif tanımlı matris olmak üzere, $E\left(e\right)=0$ ve $E\left(ee'\right)=\sigma^{2}V$ olduğunu kabul ederek, tahmin bir dereceye kadar genelleştirilebilir. Bu daha genel durumda, en küçük kareler tahmininden sistemleştirilen normal denklemler

$$\hat{\beta}=\left(X'V^{-1}X\right)^{+}X'V^{-1}Y+\left[I-\left(X'V^{-1}X\right)^{+}\left(X'V^{-1}X\right)\right]Z$$

çözümlerine sahip olan,

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

olacaktır. Bu durumda

$$E\left(\hat{\beta}\right)=\left(X'V^{-1}X\right)^{+}\left(X'V^{-1}X\right)\beta+\left[I-\left(X'V^{-1}X\right)^{+}\left(X'V^{-1}X\right)\right]Z$$

$$\begin{aligned}Var\left(\hat{\beta}\right)&=E\left[\left(\hat{\beta}-E\left(\hat{\beta}\right)\right)\left(\hat{\beta}-E\left(\hat{\beta}\right)\right)'\right]\\&=E\left[\left(X'V^{-1}X\right)^{+}X'V^{-1}ee'V^{-1}X\left(X'V^{-1}X\right)^{+}\right]\\&=\left(X'V^{-1}X\right)^{+}\left(X'V^{-1}X\right)\left(X'V^{-1}X\right)^{+}\sigma^{2}=\sigma^{2}\left(X'V^{-1}X\right)^{+}\end{aligned}$$

dır. Yukarıda irdelenen dağılımsal durumların her birinde, eğer $rank\left(X\right)=r<p$ ise, β nın tahmin edicisi tek değildir. Eksik rank durumunda, β nın elemanlarının tüm değerleri ile ilgili direkt çıkarım yapılamaz. Bununla beraber, tahmin edilebilir fonksiyonlar denilen herhangi bir çözümde değişmez nicelikler vardır. $\hat{\beta}$ nın

elemanlarının, $f'\hat{\beta}$ ile gösterilen, bir lineer kombinasyonunu göz önüne alalım. $f'\hat{\beta}$ nın keyfi Z ye göre değişmez olması için gerek ve yeter şart

$$f'\left[I - X^{+}X\right] = 0$$

olmasıdır. Eğer X tam ranklı ise, bu takdirde bu eşitlik her bir f' için sağlanır. Eksik rank durumunda $f' = b'X$ biçimindeki tüm lineer kombinasyonlar bu özelliğe sahiptir, çünkü

$$f'\left[I - X^{+}X\right] = b'\left[X - XX^{+}X\right] = 0$$

dır. Ayrıca, $\left[I - X^{+}X\right]$ in sıfırlığı r olduğundan, $\left[I - X^{+}X\right]$ in sıfır uzayı için herhangi bir taban $f'\left[I - X^{+}X\right] = 0$ denkleminin r sayıda lineer bağımsız çözümlerinin bir kümesidir. X in rankı r olduğundan ve $b'X\left[I - X^{+}X\right] = 0$ olduğundan, böyle f' vektörleri $f' = b'X$ olarak yazılabilir. Bir lineer tahmin edilebilir fonksiyonu bu değişmezlik özelliğine dayanarak tanımlayabiliriz, yani parametrelerin bir $f'\beta$ lineer fonksiyonunun tahmin edilebilir olması için gerek ve yeter şart $f'\left[I - X^{+}X\right] = 0$ olmasıdır.

Graybill(1) bir lineer tahmin edilebilir fonksiyonu, bilinmeyen parametrelerin bir fonksiyonu olarak tanımlar ki onun için,

$$E\left(b'Y\right) = b'X\beta = f'\beta$$

olacak şekilde bir b' vektörü vardır.

Aşikar olarak iki tanım eşdeğerdir. Graybill ayrıca, lineer olarak tahmin edilebilir bir fonksiyonunun

$$f' = r'X'X$$

denkleminin r için bir çözümünün bulunduğu parametrelerin bir lineer kombinasyonuna eşdeğer olduğunu kanıtlar. Bize en faydalı olacak olan, tahmin edilebilir fonksiyonun bu biçimidir.

Graybill'in genelleştirilmiş ters formüllemesinde uygulanabilirliği ile geliştirdiği sonuçlardan üçü aşağıda listenmiştir.

(1) Tahmin edilebilir bir $f'\beta$ fonksiyonunun en iyi lineer yansız tahmin edicisi

$$f'\beta = f'\hat{\beta}$$

dır.

(2) Eğer fonksiyonların her biri lineer olarak tahmin edilebilir ve $f_1', f_2', \ldots, f_m'$ ler lineer olarak bağımsız vektörler ise, bu takdirde $f_1'\beta, f_2'\beta, \ldots, f_m'\beta$ fonksiyonlarına lineer olarak tahmin edilebilir fonksiyonlar denir.

(3) Lineer olarak tahmin edilebilir fonksiyonların bir kümesi, en çok r tane fonksiyonu içerir, burada $r = rank(X)$ dir. Bir lineer olarak tahmin edilebilir $F'\beta$ matris fonksiyonu, $1 \le m \le r$ olmak üzere, m tane lineer olarak tahmin edilebilir fonksiyonunun bir kümesidir. $f_i'\beta$ lar bir lineer olarak tahmin edilebilir fonksiyonlar olmak üzere, bir matris fonksiyonu

$$F'\beta = \begin{bmatrix} f_1' \\ f_2' \\ \vdots \\ f_m' \end{bmatrix} (\beta) = (R'X'X)\beta$$

olarak gösterilebilir. Sık sık yararlanılacak olan bir sonuç, bir tahmin edilebilir matris fonksiyonunun tahmin edicisinin, keyfi Z matrisine göre değişmediğidir ki bu $F'\hat{\beta}$ nın her bir bileşeninin bu özelliğe sahip olduğu gerçeğinden direkt olarak görülür.

$e \sim MVN(0, \sigma^2 I)$ olduğu dağılımsal durumu göz önüne alarak bir tahmin edilebilir matris fonksiyonunun tahmin edicisi daima tektir ve bu nedenle tahmin edilebilir fonksiyonlarla ilgili çıkarsamalar yapmada faydalıdır. Bu durumda tahmin edicimizin ortalama vektörü ve varyans–kovaryans matrisi sırasıyla,

$$E(F'\hat{\beta}) = E(F'X^+Y) = R'X'(XX^+X)\beta = F'\beta$$

ve

$$Var\left(F'\hat{\beta}\right) = E\left[\left(F'X^{+}Y - F'\beta\right)\left(F'X^{+}Y - F'\beta\right)'\right] = E\left(F'X^{+}ee'X^{+\prime}F\right) = \sigma^2 F'F$$

dır.

3.5.2 Hipotezlerin Testleri

Şimdi, $F'\beta$ lineer bağımsız tahmin edilebilir fonksiyonların bir $m \times 1$ vektörü olmak üzere, $F'\beta = C$ hipotezinin test edilmesini ele alacağız. Eğer $F'\beta$ lineer olarak bağımsız tahmin edilebilir fonksiyonların bir kümesi ve C bilinen sabitlerin bir vektörü ise, $F'\beta = C$ ye bir lineer olarak tahmin edilebilir hipotez denir. Bu kavramı açıklamak için, β nın üç bileşeninin eşitliğini test etmeyi istediğimizi, yani

$$H_0 : \beta_1 = \beta_2 = \beta_3$$

hipotezini test etmek için istediğinizi farz edelim. Eğer,

$$f_1'\beta = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \end{pmatrix}\beta$$

ve

$$f_2'\beta = \begin{pmatrix} 1 & 1 & -2 & 0 & \cdots & 0 \end{pmatrix}\beta$$

lineer olarak tahmin edilebilir ise, bu takdirde

$$F'\beta = \left(f_1', f_2'\right)'\beta = \begin{bmatrix} \beta_1 - \beta_2 \\ \beta_1 + \beta_2 - 2\beta_3 \end{bmatrix} = 0$$

olması için gerek ve yeter şart $\beta_1 = \beta_2 = \beta_3$ olmasıdır. Hipotezin lineer tahmin edilebilir fonksiyonlarının bağımsız olmasını isteriz. Bunların herhangi birinin bağımsız olmaması gereksiz olacaktır. Örneğin yukarıdaki açıklamada

$$f_3'\beta = \begin{pmatrix} 0 & 1 & -1 & 0 & \cdots & 0 \end{pmatrix}\beta = 0$$

β_2 ve β_3 ün eşitliğini test edecektir ki bu daha önce $f_1'\beta$ ve $f_2'\beta$ vasıtasıyla test edilmektedir. Tam rank durumunda, lineer olarak tahmin edilebilir fonksiyonunun ve tahmin edilebilir hipotezin biçiminin bir sonucu, β nın elemanlarının ayrı ayrı ve birlikte olarak tahmin edilebilir olduğudur. Bu durumda $\beta = C$ hipotezi tahmin edilebilirdir. Ayrıca, β nın elemanlarının herhangi bir alt kümesinin sabitlerin belirlenmiş bir kümesine eşitliği $p \times p$ birim matrisinin uygun satırlarını F' matrisi olarak seçmek suretiyle test edilebilir.

Olabilirlik oran testi kriterinin geliştirilmesinde, sadece $e \sim MVN\left(0, \sigma^2 I\right)$ olduğu normal teori ile ilgileneceğiz. Bu durumda, uygun olabilirlik fonksiyonu

$$f\left(e; \beta, \sigma^2\right) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left[\frac{-\left(Y - X\beta\right)'\left(Y - X\beta\right)}{2\sigma^2}\right]$$

dir. Bu durumda $H_1 : F'\beta \neq C$ alternatif (karşıt) hipotezine karşı $H_0 : F'\beta = C$ hipotezinin bir testini geliştirmek için,

$$\theta = \frac{L\left(\hat{\omega}\right)}{L\left(\hat{\Omega}\right)}$$

olabilirlik oranını kullanacağız. Burada $L\left(\hat{\Omega}\right)$, kısıtlanmamış olan, $p+1$ boyutlu Ω uzayında, $\left(\beta, \sigma^2\right)$ de içerilen parametrelere sahip olabilirlik fonksiyonunun maksimum değeridir. Benzer şekilde $L\left(\hat{\omega}\right)$ ise H_0 ile kısıtlanan parametre uzayı olabilirlik fonksiyonunun maksimum değeridir. ω uzayı $p - m + 1$ boyutludur, çünkü β nın eleman arasındaki m sayıda bağımsız ilişki için değerler hipotez tarafından belirlenir. ω uzayında parametrelerin maksimumlaştırılan değerleri $\tilde{\beta}$ ve $\tilde{\sigma}^2$ ile belirtilir, halbuki Ω uzayında karşılık gelen değerleri $\hat{\beta}$ ve $\hat{\sigma}^2$ ile göstereceğiz.

Kısıtlanmış parametre uzayında, $f\left(e; \beta, \sigma^2\right)$ yi $F'\beta = C$ kısıtlaması altında σ^2 ve β ya göre maksimumlaştırmayı isteriz.

$$\phi\left(e; \beta, \sigma^2\right) = \log_e\left[f\left(e; \beta, \sigma^2\right)\right]$$

olsun.

$$\frac{\partial\phi}{\partial\sigma^2} = \frac{\left(Y - X\beta\right)'\left(Y - X\beta\right)}{2\tilde{\sigma}^4} - \frac{n}{2\tilde{\sigma}^2} = 0$$

dır. Verilen herhangi bir β için σ^2 nin maksimumlaştıran değeri

$$\tilde{\sigma}^2 = \frac{(Y - X\beta)'(Y - X\beta)}{n}$$

dir. Olabilirlik fonksiyonunu inceleyerek, olabilirlik fonksiyonunun

$$\sigma^2 = \tilde{\sigma}^2 = (Y - X\beta)'(Y - X\beta)/n$$

olmak üzere, β ya göre maksimum değerinin $(Y - X\beta)'(Y - X\beta)$ minimum olduğunda vuku bulduğu dikkate değerdir. Bununla beraber, $F'\beta = C$ ilişkisiyle β yı sınırlamalıyız. Bu takdirde hipotez kısıtlaması içinde olabilirlik fonksiyonunu maksimumlaştıran β nın değerini bulma problemi

$$F'\beta = C$$

kısıtlaması altında

$$\min (Y - X\beta)'(Y - X\beta)$$

kuadratik programı formüllenebilir. Amaç fonksiyonunun Lagrange fonksiyonunun türevleri $2\lambda'$ Lagrange çarpanlarının uygun vektörü olmak üzere, bir minimumlaştırma için gerek ve yeter şartlar olan

$$\frac{\partial \psi}{\partial \beta} = -2Y'X + 2\beta' X'X - 2\lambda' F' = 0 \tag{3.56}$$

$$\frac{\partial \psi}{\partial \lambda} = F'\beta - C = 0 \tag{3.57}$$

dır. İlk olrak (3.56) denklemini $AX^{+\prime}$ ile önden çarparak

$$-AY + F'\beta - AA'\lambda = 0$$

elde ederiz. (3.57) denkleminden $F'\beta$ için C yi yerine koyarak, λ için çözüm

$$\lambda = (AA')^{+} C - A^{+\prime} Y$$

dır. λ nın bu değeri ile (3.56) denkleminin $\tilde{\beta}$ için çözümü Q keyfi bir $p \times 1$ vektör olmak üzere,

$$\tilde{\beta} = (X'X)^{+} X'A'\left(A'^{+}A^{+}C - A'^{+}Y\right) + X^{+}X'^{+}X'Y + \left(I - X^{+}X\right)Q$$
$$= X^{+}A^{+}C - X^{+}A^{+}AY + X^{+}Y + \left(I - X^{+}X\right)Q \qquad (3.58)$$

olur. Bu durumda

$$-2X'Y + 2X'X\left[X^{+}A^{+}C - X^{+}A^{+}AY + X^{+}Y + \left(I - X^{+}X\right)Q\right]$$
$$-2X'A'\left[(AA')^{+}C - A^{+\prime}Y\right] = 0$$

ve

$$-X'Y + X'A^{+}C - X'A^{+}AY + X'Y - X'A^{+}C + X'A'A^{+\prime}Y = 0$$

olduğundan, (3.56) denklemi $\tilde{\beta}$ ve λ ile sağlanır. Aynı zamanda,

$$F'\beta = R'X'X\left[X^{+}A^{+}C - X^{+}A^{+}AY + X^{+}Y + \left(I - X^{+}X\right)Q\right]$$
$$= AA^{+}C - AA^{+}AY + AY$$

dır. Öte yandan A matrisi m ranklı bir $m \times n$ matris olduğundan

$$AA^{+} = I$$

olur ve (3.57) denklemi sağlanır. $\tilde{\beta}$ genel terimini muhafaza eder iken olabilirlik fonksiyonuna geri dönüp $\tilde{\sigma}^2$ yi yerine koyarak,

$$L(\hat{\omega}) = \max_{\substack{\sigma^2 \in \omega \\ F'\beta = C}} f\left(e;\beta,\sigma^2\right) = \frac{n^{n/2}\exp[-n/2]}{(2\pi)^{n/2}\left[\left(Y - X\tilde{\beta}\right)'\left(Y - X\tilde{\beta}\right)\right]^{n/2}}$$

elde edilir. Kısıtlanmamış parametre uzayında σ^2 ve β nın maksimuma çıkaran değerlerini bulmak

$$\hat{\beta} = X^{+}Y + \left(I - X^{+}X\right)Z$$
$$\hat{\sigma}^2 = \frac{\left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right)}{n} \qquad (3.59)$$

sonuçlarını ortaya koyar. Burada Z keyfi bir $p \times 1$ vektör ve $\hat{\sigma}^2$; σ^2 nin bir yanlı tahmin edicisidir. Olabilirlik fonksiyonunda bu sonuçları yerine yazmak

$$L\left(\hat{\Omega}\right) = \max_{\substack{\sigma^2 \in \Omega \\ \beta \in \Omega}} f\left(e;\beta,\sigma^2\right) = \frac{n^{n/2}\exp\left[-n/2\right]}{\left(2\pi\right)^{n/2}\left[\left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right)\right]^{n/2}}$$

maksimum değerini verir. Bu takdirde $\tilde{\beta}$ ve $\hat{\beta}$ sırasıyla, (3.58) ve (3.59) denklemlerinde sunulduğu gibi olmak üzere, olabilirlik oranı

$$\theta = \frac{L\left(\hat{\omega}\right)}{L\left(\hat{\Omega}\right)} = \left[\frac{\left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right)}{\left(Y - X\tilde{\beta}\right)'\left(Y - X\tilde{\beta}\right)}\right]^{n/2}$$

dir. Test kriterimizi şekillendirmek için, ilk olarak olabilirlik oranının kuadratik formlarını ki-kare rasgele değişkenleri biçimine getiririz. Bu takdirde kuadratik formların bağımsızlığı gösterilecek ve merkezi olmayan bir *F*–istatistiği saptanacaktır. Herhangi bir $\tilde{\beta}$ için,

$$\left(Y - X\tilde{\beta}\right)'\left(Y - X\tilde{\beta}\right) = \left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right) + \left(\hat{\beta} - \tilde{\beta}\right)'\left(X'X\right)\left(\hat{\beta} - \tilde{\beta}\right)$$

dır. Bunun sonucu olarak, olabilirlik oranı

$$\theta = \left(1 + \frac{\left(\hat{\beta} - \tilde{\beta}\right)'\left(X'X\right)\left(\hat{\beta} - \tilde{\beta}\right)}{\left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right)}\right)^{-n/2} \tag{3.60}$$

olarak yazılabilir. Şimdi paydadaki oranı inceleyelim. $\hat{\beta}$ ve $\tilde{\beta}$ için çözümleri kullanarak,

$$\begin{aligned}
&\left(\hat{\beta} - \tilde{\beta}\right)' X'X\left(\hat{\beta} - \tilde{\beta}\right) \\
&= \left(XX^{+}Y - XX^{+}A^{+}C + XX^{+}A^{+}AY - XX^{+}Y\right)'\left(XX^{+}Y - XX^{+}A^{+}C + XX^{+}A^{+}AY - XX^{+}Y\right) \\
&= \left(XX^{+}A^{+}AY - XX^{+}A^{+}C\right)'\left(XX^{+}A^{+}AY - XX^{+}A^{+}C\right) \\
&= \left(AY - C\right)'\left(A^{+\prime}XX^{+}A^{+}\right)\left(AY - C\right)
\end{aligned}$$

olduğu görülür. Merkezi olmayan bir ki-kare rasgele değişkeninin tanımına göre, eğer $Z \sim MVN\left(\mu, V\right)$ ise, bu takdirde $Z'BZ \sim \chi^2\left(q, \lambda\right)$ olması için gerek ve yeter şart BV nin idempotent olmasıdır. Burada q, BV nin rankıdır ve $\lambda = \mu' B\mu / 2$ dir. Eğer $\left[\left(A^{+\prime} XX^{+} A^{+}\right) / \sigma^2\right]\left[Var\left(AY - C\right)\right]$ matrisi idempotent ise $\left(\hat{\beta} - \tilde{\beta}\right)' X'X\left(\hat{\beta} - \tilde{\beta}\right) / \sigma^2$ merkezi olmayan bir ki-kare dağılımına sahiptir. $Y \sim MVN\left(X\beta, \sigma^2 I\right)$ olduğundan, $\left(AY - C\right) \sim MVN\left(AX\beta - C, \sigma^2 AA'\right)$ ve

$$BV = \frac{A^{+\prime} XX^{+} A^{+}}{\sigma^2} \sigma^2 AA'$$

dir. Bu durumda $A^{+\prime} XX^{+} A^{+} AA'$ matrisinin idempotent olup olmadığını belirlemeliyiz. $R'X'X = AX$ olmak üzere,

$$A^{+\prime} XX^{+}\left(A^{+} AA'\right) = A^{+\prime} XX^{+}\left(A'\right) = A^{+\prime} XX^{+}\left(XR\right)$$

olduğu açıktır. Bu nedenle,

$$A^{+\prime} XX^{+} A^{+} AA' = A^{+\prime} XR = A^{+\prime} A'$$

yazılabilir. A', m ranklı bir $p \times m$ matris olduğundan bir $A^{+\prime}$ sol tersine sahiptir ve

$$BV = A^{+\prime} XX^{+} A^{+} AA' = I$$

eşitliği sağlanır. Bu nedenle, BV idempotentdir ve m ranklıdır. Böylece, $\left(AY - C\right)'\left(A^{+\prime} XX^{+} A^{+}\right)\left(AY - C\right) / \sigma^2$ nin m serbestlik dereceli ve

$$\lambda = \left(F'\beta - C\right)'\left(A^{+\prime} XX^{+} A^{+}\right)\left(F'\beta - C\right) / 2\sigma^2$$

merkezi olmama parametreli merkezi olmayan bir ki-kare dağılımına sahip olduğu sonucuna varırız.

$\left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right)$ kuadratik formu $Y'\left(I - XX^{+}\right)Y$ ye eşittir. Bu durumda $\left(I - XX^{+}\right)$ matrisi $n - r$ ranklı bir idempotent matris olduğundan, $\left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right)$, $n - r$ serbestlik dereceli merkezi bir ki-kare dağılımına sahiptir. Sonuç olarak,

$$\gamma \sim \chi^2\left(m, (F'\beta - C)' \left(A^{+\prime} X X^{+} A^{+}\right)(F'\beta - C) \Big/ 2\sigma^2\right)$$

ve $\xi \sim \chi^2(n-r)$ değişkeni olmak üzere, olabilirlik oranımız

$$\theta = \left[\frac{1}{1+\dfrac{\gamma}{\xi}}\right]^{n/2} \tag{3.61}$$

dir.

Genel olarak, iki $T'DT$ ve $T'ET$ kuadratik formunun bağımsız olması için gerek ve yeter şart $DVE = 0$ olmasıdır. Burada $T \sim MVN(\mu, V)$ dir.

$$\gamma = \left(\hat{\beta} - \tilde{\beta}\right)' X'X \left(\hat{\beta} - \tilde{\beta}\right) \Big/ \sigma^2 = \left(Y - X\tilde{\beta}\right)' XX^{+} \left(Y - X\tilde{\beta}\right) \Big/ \sigma^2$$

ve

$$\xi = \left(Y - X\hat{\beta}\right)' \left(Y - X\hat{\beta}\right) \Big/ \sigma^2 = \left(Y - X\tilde{\beta}\right)' \left(I - XX^{+}\right)\left(Y - X\tilde{\beta}\right) \Big/ \sigma^2$$

özdeşlikleri γ ve ξ nin gösterimi ve bağımsızlığı için uygun biçimlerdir. Bu durumda

$$\left(Y - X\tilde{\beta}\right) = \left(I - XX^{+} + XX^{+}A^{+}A\right)Y - XX^{+}A^{+}C$$

ve Y nin varyansı $\sigma^2 I$ olduğundan,

$$Var\left(Y - X\tilde{\beta}\right) = \sigma^2\left(I - XX^{+} + XX^{+}A^{+}A\right)\left(I - XX^{+} + XX^{+}A^{+}A\right)' = \sigma^2\left(I - XX^{+} + A^{+}A\right)$$

dır. Bu takdirde DVE ye karşılık gelen nicelik, kolayca doğrulanabilen

$$\left(\frac{X'X}{\sigma^2}\right)\left(I - XX^{+} + A^{+}A\right)\left(\frac{I - XX^{+}}{\sigma^2}\right) = 0$$

dır. Merkezi olmayan ki-kare değişkeninin serbestlik derecesine bölünen her bir değişkene sahip olabilirlik oranımızın paydasındaki merkezi ki-kare değişkenine oranı m ve $n-r$ serbestlik dereceli ve

$$\lambda = \left(F'\beta - C\right)'\left(A^{+'}XX^{+}A^{+}\right)\left(F'\beta - C\right)$$

merkezi olmama parametreli merkezi olmayan bir F–dağılımına sahiptir. $\lambda = 0$ olması için bir gerek ve yeter şart hipotezin doğru olması, yani $F'\beta - C = 0$ olmasıdır. Bu durumda, serbestlik derecelerine bölünen kuadratik formların oranı

$$\frac{\gamma(n-r)}{\xi(m)} = \frac{\left(AY - C\right)'\left(A'^{+}XX^{+}A^{+}\right)\left(AY - C\right)}{\left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right)} \sim F_{m,n-r}$$

F–dağılımına sahiptir. Hipotezin reddedilmesi olabilirlik oranının küçük değerlere sahip olması ve $\left[\gamma(n-r)\right] / \left[\xi(m)\right]$ ye göre monoton azalıyor olması ile tutarlıdır, $F_{m,n-r}(\alpha)$ α anlam düzeyine karşılık gelen üst anlam noktası olmak üzere, bu test için kritik bölge

$$\frac{\gamma(n-r)}{\xi(m)} > F_{m,n-r}(\alpha)$$

dır. Olabilirlik oranı kriterimizin doğrulaması olarak, onu tam rank durumunda iki tip test için çoğunlukla kullanılan kriterlerle karşılaştıracağız. $\beta = \beta^*$ için tam hipotez testi kriterinin alışılmış biçimi, örneğin Graybill (1)

$$\frac{n-p}{p}\frac{\left(Y - X\beta^*\right)'\left(X\left(X'X\right)^{-1}X'\right)\left(Y - X\beta^*\right)}{\left(Y - XX^{+}Y\right)'\left(Y - XX^{+}Y\right)} > F_{p,n-p}(\alpha)$$

dır. Paydaların aynı olduğu görülür. Bundan başka

$$\begin{aligned}\left(Y - X\beta^*\right)'\left(X\left(X'X\right)^{-1}X'\right)\left(Y - X\beta^*\right) &= \left(Y - X\beta^*\right)'\left(XX^{+}\right)'\left(XX^{+}\right)\left(Y - X\beta^*\right) \\ &= \left(X^{+}Y - \beta^*\right)'\left(X'X\right)\left(X^{+}Y - \beta^*\right)\end{aligned}$$

dır. Bu nedenle test kriterleri tam rank hipotez testinde eşdeğerdirler. Diğer $p-s$ sayıda eleman belirlenmemiş iken, β nın elemanlarından s tanesinin bilinen sabitlere eşit olduğunu test etmek için, Graybill(1)

$$\frac{n-p}{s}\frac{\left(Y-X_1\gamma_1^*\right)'\left(X\left(X'X\right)^{-1}X'-X_2\left(X_2'X_2\right)^{-1}X_2'\right)\left(Y-X_1\gamma_1^*\right)}{\left(Y-X_1\gamma_1^*\right)'\left(I-X\left(X'X\right)^{-1}X'\right)\left(Y-X_1\gamma_1^*\right)}>F_{s,n-p}\left(\alpha\right)$$

alt hipotez kriterini kullanır. Bu tam rank test amacıyla, genel lineer modelin elemanları (unsurları)

$$X=\left(X_1,X_2\right),\ \beta'=\left(\gamma_1,\gamma_2\right)$$

olacak şekilde parçalanmıştır ve γ_1^* bilinen sabitlerin bir $s\times 1$ vektörü olmak üzere, hipotez $\gamma_1=\gamma_1^*$ dır. Genelleştirilmiş ters formüllemesine göre bu hipotez $C=\gamma_1^*$ koyarak ve *I*, *s* boyutlu birim matris ve 0, $s\times\left(p-s\right)$ boyutlu sıfır matris olmak üzere, $F'=\left[I\ \vdots\ 0\right]$ seçerek test edilebilir. Bir basitleştirme olarak, sadece X_1 in X_2 ye ortogonal olduğu, yani $X_1'X_2=0$ olduğu tam ranklı durumunu göz önüne alacağız. $F'=AX=\left[I\ \vdots\ 0\right]$ olduğundan,

$$AX_1=I \text{ ve } AX_2=0$$

yazmalıyız. Kısım 3.6 daki (11) özelliğine göre, $A=X_1^+$ olduğu açıktır. Bu takdirde hipotezimiz $F'\beta=\gamma_1^*$ dır ve test kriteri

$$\frac{n-p}{s}\frac{\left(X_1^+Y-\gamma_1^*\right)'\left(X_1'XX^+X_1\right)\left(X_1^+Y-\gamma_1^*\right)}{\left(Y-X\hat{\beta}\right)'\left(Y-X\hat{\beta}\right)}>F_{s,n-p}\left(\alpha\right)$$

dır. Bu durumda yine paydaların aynı olduğu gözükür.

$$\begin{aligned}
&\left(Y-X_1\gamma_1^*\right)'\left(X\left(X'X\right)^{-1}X'-X_2\left(X_2'X_2\right)^{-1}X_2'\right)\left(Y-X_1\gamma_1^*\right)\\
&=\left(Y-X_1\gamma_1^*\right)'\left(XX^+-X_2X_2^+\right)\left(Y-X_1\gamma_1^*\right)\\
&=\left(Y-X_1\gamma_1^*\right)'\left(X_1X_1^+\right)\left(Y-X_1\gamma_1^*\right)\\
&=\left(Y-X_1\gamma_1^*\right)'X_1'^+\left(X_1'XX^+X_1\right)X_1^+\left(Y-X_1\gamma_1^*\right)\\
&=\left(X_1^+Y-\gamma_1^*\right)'\left(X_1'XX^+X_1\right)\left(X_1^+Y-\gamma_1^*\right)
\end{aligned}$$

olduğundan paylar da aynıdır. Bu takdirde test kriterimizin alt hipotez testinin bu özel durumunda çoğunlukla kullanılan kritere denkliği açıktır.

Örnek 3.24. $Y_{ij} = \mu + \tau_i + e_{ij}$, $i = 1,2,3$, $j = 1,2$ olmak üzere, tam olarak rasgeleleştirilmiş tasarım modelini göz önüne alalım.

$$e \sim MVN\left(0, \sigma^2 I\right)$$

olmak üzere,

$$Y = X\beta + e$$

istatistiksel modelimiz için gerekli olan normallik varsayımlarının savunulabilir olduğunu farz edeceğiz. İstatistiksel modelimizin gerekli unsurları

$$Y = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 2 \\ 8 \\ 10 \end{bmatrix}, \qquad \beta = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

dır. X in rankı $r = 3$ tür. Genelleştirilmiş tersi tanımlayan denklemlerin çözümüyle

$$X^+ = 1/8 \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 3 & 3 & -1 & -1 & -1 & -1 \\ -1 & -1 & 3 & 3 & -1 & -1 \\ -1 & -1 & -1 & -1 & 3 & 3 \end{bmatrix}$$

ve

$$\left(X'X\right)^+ = 1/64 \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 22 & -10 & -10 \\ 2 & -10 & 22 & -10 \\ 2 & -10 & -10 & 22 \end{bmatrix}$$

dır. Bu tam ranklı olmayan durumun bir örneği olduğu için tahmin tek değildir. Tahmin edici

$$\hat{\beta} = X^{+}Y + \left[I - X^{+}X\right]Z = 1/8\begin{bmatrix} 26 \\ -10 \\ -10 \\ 46 \end{bmatrix} + 1/8\begin{bmatrix} 2 & -2 & -2 & -2 \\ -2 & 2 & 2 & 2 \\ -2 & 2 & 2 & 2 \\ -2 & 2 & 2 & 2 \end{bmatrix} Z$$

genel çözümüne sahiptir. $\hat{\beta}$ nın ortalama vektörü

$$E\left(\hat{\beta}\right) = X^{+}X\beta + \left[I - X^{+}X\right]Z$$

$$= 1/8\left[\begin{bmatrix} 6\mu + 2\tau_1 + 2\tau_2 + 2\tau_3 \\ 2\mu + 6\tau_1 - 2\tau_2 - 2\tau_3 \\ 2\mu - 2\tau_1 + 6\tau_2 - 2\tau_3 \\ 2\mu - 2\tau_1 - 2\tau_2 + 6\tau_3 \end{bmatrix} + \begin{bmatrix} 2Z_1 - 2Z_2 - 2Z_3 - 2Z_4 \\ -2Z_1 + 2Z_2 + 2Z_3 + 2Z_4 \\ -2Z_1 + 2Z_2 + 2Z_3 + 2Z_4 \\ -2Z_1 + 2Z_2 + 2Z_3 + 2Z_4 \end{bmatrix}\right]$$

ve $\hat{\beta}$ nın varyans –kovaryans matrisi

$$Var\left(\hat{\beta}\right) = \sigma^2\left(X'X\right)^{+} = \sigma^2/32\begin{bmatrix} 3 & 1 & 1 & 1 \\ 1 & 11 & -5 & -5 \\ 1 & -5 & 11 & -5 \\ 1 & -5 & -5 & 11 \end{bmatrix}$$

dır. Bundan başka, σ^2 nin yansız tahmini

$$\hat{\sigma}^2 = \frac{\left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right)}{n-r} = 4/3$$

olur. Tam ranklı olmayan durumda β nın tahmin edicisi kendiliğinden β nın her bir bileşeni hakkında bize herhangi bir şeyi söylemez. Bununla beraber,

$$R' = \begin{bmatrix} 0 & 1/2 & -1/2 & 0 \\ 0 & 1/2 & 1/2 & -1 \end{bmatrix}',$$

$$A = R'X' = \begin{bmatrix} 1/2 & 1/2 & -1/2 & -1/2 & 0 & 0 \\ 1/2 & 1/2 & 1/2 & 1/2 & -1 & -1 \end{bmatrix}'$$

ve

$$F' = AX = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & -2 \end{bmatrix}$$

olmak üzere, parametrelerin bir fonksiyonunun tahminini göz önüne alalım. Bu durumda $F'\beta$ tahmin edilebilir matris fonksiyonunun tahmin edicisi

$$E\left(F'\hat{\beta}\right)=F'\beta=\begin{pmatrix}\tau_1-\tau_2\\ \tau_1+\tau_2-2\tau_3\end{pmatrix}$$

ortalamalı ve

$$Var\left(F'\beta\right)=\sigma^2F'F=\begin{pmatrix}2 & 0\\ 0 & 6\end{pmatrix}$$

varyans–kovaryans matrisli

$$F'\hat{\beta}=FX^{+}Y=\begin{pmatrix}0\\ 14\end{pmatrix}$$

tür. Çoğunlukla tüm τ_i lerin yani işlem etkilerinin eşitliği test edilmek istenir. Bu durumda karşılık gelen tahmin edilebilir hipotez

$$F'\beta=\begin{pmatrix}0 & 1 & -1 & 0\\ 0 & 1 & 1 & -2\end{pmatrix}\beta=\begin{pmatrix}\tau_1-\tau_2\\ \tau_1+\tau_2-2\tau_3\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}$$

dır. Bu tahmin edilebilir hipotez için, A' nün genelleştirilmiş tersi

$$A^{+\prime}=\begin{pmatrix}1/2 & 1/2 & -1/2 & -1/2 & 0 & 0\\ 1/6 & 1/6 & 1/6 & 1/6 & -1/3 & -1/3\end{pmatrix}$$

olarak yazılır. Test istatistiği

$$\left(\frac{n-r}{m}\right)\frac{\gamma}{\xi}=\left(\frac{n-r}{m}\right)\frac{\left(AY-C\right)'\left(A^{+\prime}XX^{+}A^{+}\right)\left(AY-C\right)}{\left(Y-X\hat{\beta}\right)'\left(Y-X\hat{\beta}\right)}=24.5$$

tir. Öte yandan bu test için kritik bölge (ret bölgesi)

$$\left(\frac{n-r}{m}\right)\frac{\gamma}{\xi}>F_{2,3}\left(.05\right)=9.55$$

tir. Bu nedenle, $H_0:\tau_1=\tau_2=\tau_3$ hipotezini reddetmeliyiz.

$E(\varepsilon)=0$ ve $Var(\varepsilon)=\sigma^2 I$, fakat X bir tam sütun ranka sahip olmamak, yani $rank(X)=k<p\leq n$ ve $X'X$ tekil olmak üzere

$$y = X\beta + \varepsilon$$

modelini göz önüne alalım. Bu modelde β daki p tane parametre tek değildir. Bu durumda en küçük kareler yaklaşımı bizi

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

normal denklemlerine götürür.

Teorem 3.19. Eğer *X*, $k<p\leq n$ ranklı $n\times p$ matris ise, bu takdirde $X'X\hat{\beta}=X'y$ denklemler sistemi tutarlıdır.

İspat. Sistemin tutarlı olması için gerek ve yeter şart

$$X'X\left(X'X\right)^{-}X'y = X'y$$

olmasıdır. Bu durumda $X'X\left(X'X\right)^{-}X' = X'$ olduğundan sistem tutarlıdır. Böylece ispat tamamlanır.

Normal denklemler tutarlı olduğundan, bir çözüm $\hat{\beta}=\left(X'X\right)^{-}X'y$ dir. Bir takım genel sonuçlar aşağıda verilir.

(1) $\hat{\beta}$, y nin bir lineer fonksiyonudur.

(2) $E\left(\hat{\beta}\right)=\left(X'X\right)^{-}X'E(y)=\left(X'X\right)^{-}X'X\beta$ dır ki bu, $\left(X'X\right)^{-}$ ye bağlıdır ve yanlıdır.

(3) β tahmin edilebilir değildir. Bir lineer Ay fonksiyonunun β yı tahmin ettiğini, yani

$$\beta = E(Ay) = E(AX\beta + A\varepsilon) = AX\beta$$

olduğunu farz edelim. Bu her β için sağlanması gerektiğinden $AX=I_p$ olmalıdır. Ancak, $rank(AX)\leq rank(X)<p$ dir ve bu nedenle $AX\neq I_p$ dir ve böyle bir A mevcut değildir.

Teorem 3.20. $E(y) = X\beta$ ve X matrisi $k < p \leq n$ ranklı $n \times p$ olmak üzere, $y = X\beta + \varepsilon$ modelinde $\lambda'\beta$ lineer parametrik fonksiyonunun tahmin edilebilir olması için gerek ve yeter şart aşağıdaki şartların herhangi birinin sağlanmasıdır.

(i) λ' X in satırlarının bir lineer kombinasyonudur, yani $a'X = \lambda'$ olacak şekilde bir a vektörü mevcuttur.

(ii) λ' $X'X$ in satırlarının bir lineer kombinasyonudur veya λ $X'X$ in sütunlarının bir lineer kombinasyonudur, yani $r'X'X = \lambda'$ veya $X'Xr = \lambda$ olacak şekilde bir r vektörü mevcuttur.

(iii) $(X'X)^{-}$ $X'X$ in herhangi bir simetrik genelleştirilmiş tersi olmak üzere, λ' $X'X(X'X)^{-}\lambda = \lambda$ veya $\lambda'(X'X)^{-}X'X = \lambda'$ olacak şekildedir.

İspat. Sadece "yeter kısmını" ispatlayacağız.

(i) Eğer, $\lambda' = a'X$ olacak şekilde bir a vektörü mevcut ise bu takdirde

$$E(a'y) = a'E(y) = a'X\beta = \lambda'\beta$$

dır.

(ii) Eğer $X'Xr = \lambda$ ya çözüm mevcut ise, bu takdirde $a = Xr$ yi tanımlayarak

$$E(a'y) = E(r'X'y) = r'X'E(y) = r'X'X\beta = \lambda'\beta$$

elde ederiz.

(iii) Eğer $X'X(X'X)^{-}\lambda = \lambda$ ise, bu takdirde $(X'X)\lambda$ (ii) kısmındaki $X'Xr = \lambda$ ya bir çözümdür. Tersine olarak eğer $\lambda'\beta$ tahmin edilebilir ise, bu takdirde $X'Xr = \lambda$ $r = (X'X)^{-}\lambda$ şeklinde bir çözüm vektörüne sahiptir. Bunu $X'Xr = \lambda$ da yerine koymak $X'X(X'X)^{-}\lambda = \lambda$ olduğunu ortaya koyar. Bu da ispatı tamamlar.

Böylece parametrelerin hangi fonksiyonlarının tahmin edilebilir olduklarını görmek için X in veya $X'X$ in satırlarının lineer kombinasyonlarını inceleyebiliriz.

Örnek 3.25.

$$X=\begin{pmatrix}1&1&0&1&0\\1&1&0&0&1\\1&0&1&1&0\\1&0&1&0&1\end{pmatrix},\quad \beta=\begin{pmatrix}\mu\\\alpha_1\\\alpha_2\\\beta_1\\\beta_2\end{pmatrix}$$

yi göz önüne alalım.

Örneğin, (3. Satır) – (1. Satır), yani $\lambda'=(0,-1,1,0,0)$ olduğundan $\lambda'\beta=-\alpha_1+\alpha_2$ tahmin edilebilirdir. Üç lineer bağımsız satırı elde etmek için, X in satırlarının $a'X$ lineer kombinasyonlarını alalım. X in birbirini izleyen her bir satırından birinci satırı çıkarıp sonra da matrisin dördüncü satırından ikinci ve üçüncü satırı çıkarılırsa

$$\begin{pmatrix}1&1&0&1&0\\0&0&0&-1&1\\0&-1&1&0&0\\0&0&0&0&0\end{pmatrix}$$

matrisi elde edilir. Böylece, $\lambda_1'\beta=\mu+\alpha_1+\beta_1$, $\lambda_2'\beta=\beta_2-\beta_1$ ve $\lambda_3'\beta=\alpha_2-\alpha_1$ şeklinde üç lineer bağımsız tahmin edilebilir fonksiyon elde edilir.

Teorem 3.20 (i) ve (ii) den a' ve r' sırasıyla $\lambda'=a'X$ ve $\lambda'=r'X'X$ eşitliklerini sağlamak üzere, $\lambda'\beta$ için $a'y$ ve $r'X'y$ tahmin edicilerine sahip olalım. $\lambda'\beta$ nın bir üçüncü tahmin edicisi, $\hat{\beta}$ $X'X\hat{\beta}=X'y$ ye bir çözüm olmak üzere, $\lambda'\hat{\beta}$ dır.

Teorem 3.21. $E(y)=X\beta$ ve X $k<p\le n$ ranka sahip $n\times p$ matris olmak üzere, $\lambda'\beta$ $y=X\beta+\varepsilon$ modelinde β nın bir tahmin edilebilir fonksiyonu olsun. $\hat{\beta}$ $X'X\hat{\beta}=X'y$ normal denklemlerinin herhangi bir çözümü olsun ve r de $X'Xr=\lambda$ nın herhangi bir çözümü olsun. Bu takdirde $\lambda'\hat{\beta}$ ve $r'X'y$ aşağıdaki özelliklere sahiptir:

(i) $E(\lambda'\hat{\beta})=E(r'X'y)=\lambda'\beta$.

(ii) $\lambda'\hat{\beta}$ herhangi bir $\hat{\beta}$ veya herhangi bir r için $r'X'y$ ye eşittir.

(iii) $\lambda'\hat{\beta}$ ve $r'X'y$ $\hat{\beta}$ ve r nin seçimine göre değişmez .

İspat.

(i)

$$E(\lambda'\hat{\beta}) = \lambda' E(\hat{\beta}) = \lambda'(X'X)^{-} X'X\beta$$

olduğundan Teorem 3.20 (iii) den, $\lambda'\left(X'X\right)^{-} X'X = \lambda'$ yazılabilir ve bu nedenle

$$E\left(\lambda'\hat{\beta}\right) = \lambda'\beta$$

olur. Teorem 3.20 (ii) ye göre

$$E(r'X'y) = r'X'E(y) = r'X'X\beta = \lambda'\beta$$

dır.

(ii) Teorem 3.20 (ii) ye göre eğer $\lambda'\beta$ tahmin edilebilir ise, herhangi bir r için $\lambda' = r'X'X$ dir. Bu durumda $X'X\hat{\beta} = X'y$ normal denklemlerinin r' ile çarpımı

$$r'X'X\hat{\beta} = r'X'y$$

yi verir. Öte yandan $r'X'X = \lambda'$ olduğundan

$$\lambda'\hat{\beta} = r'X'y$$

elde edilir.

(iii) $r'X'y$ nin r nin seçimine göre değişmez olduğunu göstermek için, r_1 ve r_2, $X'Xr_1 = X'Xr_2 = \lambda$ olacak şekilde verilmiş olsun. Bu takdirde

$$r_1'X'X\hat{\beta} = r_1'X'y \text{ ve } r_2'X'X\hat{\beta} = r_2'X'y$$

dir. Öte yandan $r_1'X'X = r_2'X'X$ olduğundan, $r_1'X'y = r_2'X'y$ elde edilir. Her birinin $\lambda'\hat{\beta}$ ya eşit olduğu açıktır. Böylece ispat tamamlanır.

Teorem 3.22. X $k < p \leq n$ ranklı $n \times p$ matris ve $kov\left(y\right) = \sigma^2 I$ olmak üzere, $\lambda'\beta$ $y = X\beta + \varepsilon$ modelinde bir tahmin edilebilir fonksiyon olsun. r de $X'Xr = \lambda$ nın herhangi bir çözümü olsun ve $\hat{\beta}$; $X'X\hat{\beta} = X'y$ nin herhangi bir çözümü olsun. Bu takdirde $\lambda'\hat{\beta}$ veya $r'X'y$ nin varyansı aşağıdaki özelliklere sahiptir.

(i) $\mathrm{var}\left(r'X'y\right)=\sigma^2 r'X'Xr=\sigma^2 r'\lambda$.

(ii) $\mathrm{var}\left(\lambda'\hat{\beta}\right)=\sigma^2\lambda'\left(X'X\right)^{-}\lambda$.

(iii) $\mathrm{var}\left(\lambda'\hat{\beta}\right)$ tektir, yani *r* veya $\left(X'X\right)^{-}$ nin seçimine göre değişmezdir.

İspat. (i) $\mathrm{var}\left(r'X'y\right)=r'X'kov\left(y\right)Xy=\sigma^2 r'X'Xr=\sigma^2 r'\lambda$ dır.

(ii) $\mathrm{var}\left(\lambda'\hat{\beta}\right)=\lambda'kov\left(\hat{\beta}\right)\lambda=\sigma^2\lambda'\left(X'X\right)^{-}X'X\left(X'X\right)^{-}\lambda$ dır. Bu durumda $\lambda'\left(X'X\right)^{-}X'X=\lambda'$ olduğundan (ii) nin ispatı tamamlanır.

(iii) $\lambda'(X'X)^{-}\lambda=r'(X'X)(X'X)^{-}(X'X)r=r'(X'X)r=r'\lambda$ olduğundan, sadece $\lambda'\left(X'X\right)^{-}\lambda$ nın $\left(X'X\right)^{-}$ nin seçimine göre değişmez olduğunu göstermek gerekir. G_1 ve G_2; $X'X$ in iki genelleştirilmiş tersi olsun . Bu takdirde

$$XG_1X'=XG_2X'$$

dür. Eşitliğin her iki yanı $a'X=\lambda'$ olacak şekilde *a* ile çarpılırsa

$$a'XG_1X'a=a'XG_2X'a$$

veya

$$\lambda_1'G_1\lambda_2=\lambda_1'G_2\lambda_2$$

elde edilir. Böylece ispat tamamlanır.

Teorem 3.23. Eğer $\lambda'\beta$; X $k<p\le n$ ranklı $n\times p$ matris olmak üzere $y=X\beta+\varepsilon$ modelinde bir tahmin edilebilir fonksiyon ise, bu takdirde $\lambda'\hat{\beta}$ ve $r'X'y$ tahmin edicileri BLUE dir.

İspat. Genelliği yitirmeksizin $a'y=r'X'y+c'y$ yani, $a'=r'X'+c'$, r' $\lambda'=r'X'X$ nin bir çözümü olmak üzere, $\lambda'\beta$; $a'y$ ile gösterilen bir lineer tahmin edici olsun.

Yansızlık için

$$\lambda'\beta=E\left(a'y\right)=a'X\beta=r'X'X\beta+c'X\beta=\left(r'X'X+c'X\right)\beta$$

elde etmeliyiz. Bu, her β için gerçeklenmelidir ve bu nedenle

$$\lambda' = r'X'X + c'X$$

elde edilir. $\lambda' = r'X'X$ olduğundan, $c'X = 0'$,

$$\begin{aligned}\text{var}(a'y) &= a'kov(y)a = \sigma^2 a'a \\ &= \sigma^2 (r'X' + c')(rX + c) \\ &= \sigma^2 (r'X'Xr + r'X'c + c'Xr + c'c) \\ &= \sigma^2 (r'X'Xr + c'c)\end{aligned}$$

olduğu görülür. Bu nedenle, $\text{var}(a'y)$ yi minimumlaştırmak için, $c = 0$ yazarız ki bu durumda $r'Xy$ BLUE dir. Böylece ispat tamamlanır.

Şimdi σ^2 nin tahmin edicisini inceleyelim. Bu durumda σ^2 nin bir tahmin edicisi

$$s^2 = \frac{SSE}{n-k} \tag{3.62}$$

dır. Burada $SSE = (y - X\hat{\beta})'(y - X\hat{\beta})$ dır. Bu durumda $\hat{\beta}$, $X'X\hat{\beta} = X'y$ normal denklemlerine bir çözümdür ve $k = rank(X)$ dir.

Teorem 3.24. $E(y) = X\beta$ ve $kov(y) = \sigma^2 I$ olmak üzere $y = X\beta + \varepsilon$ eksik ranklı modeli için (3.62) bağıntısında tanımlanan s^2 için aşağıdaki özellikler sağlanır.

(i) $E(s^2) = \sigma^2$

(ii) s^2, $\hat{\beta}$ nın seçimiyle veya $(X'X)^-$ genelleştirilmiş tersinin seçimiyle değişmez.

İspat. (i) $SSE = y'\left[I - X(X'X)^- X'\right]y$ olduğundan,

$$\begin{aligned}E(SSE) &= iz\left\{\left[I - X(X'X)^- X'\right](\sigma^2 I)\right\} + \beta' X'\left[I - X(X'X)^- X'\right]X\beta \\ &= iz\left\{\left[I - X(X'X)^- X'\right](\sigma^2 I)\right\} \\ &= \sigma^2\left\{iz(I) - iz\left[X(X'X)^- X'\right]\right\} \\ &= (n-k)\sigma^2\end{aligned}$$

elde edilir, burada $k = rank(X'X) = rank(X)$ dir.

(ii) $X\left(X'X\right)^{-}X'$; $\left(X'X\right)^{-}$ nin seçimine göre değişmez.

3.5.3 Normal Model

$y = X\beta + \varepsilon$ eksik ranklı modeli için,

$$y \sim N_n\left(X\beta, \sigma^2 I\right) \text{ veya } \varepsilon \sim N_n\left(0, \sigma^2 I\right)$$

olduğunu kabul edelim. Bu durumda normallik varsayımı altında en çok olabilirlik tahmin edicilerini elde edebiliriz.

Teorem 3.25. X, $k < p \le n$ ranklı $n \times p$ matris olmak üzere, eğer $y \sim N_n\left(X\beta, \sigma^2 I\right)$ ise, bu takdirde β ve σ^2 için en çok olabilirlik tahmin edicileri

$$\hat{\beta} = \left(X'X\right)^{-} X'y$$
$$\hat{\sigma}^2 = \frac{1}{n}\left(y - X\hat{\beta}\right)'\left(y - X\hat{\beta}\right)$$

dır.

Teorem 3.26. X $k < p \le n$ ranklı $n \times p$ matris olmak üzere, eğer $y \sim N_n\left(X\beta, \sigma^2 I\right)$ ise bu takdirde $\hat{\beta}$ ve s^2 (yan için düzeltilmiş) en çok olabilirlik tahmin edicileri aşağıdaki özelliklere sahiptir.

(i) $\hat{\beta} \sim N_p\left(\left(X'X\right)^{-} X'X\beta, \sigma^2\left(X'X\right)^{-} X'X\left(X'X\right)^{-}\right)$ dir.

(ii) $\left(n-k\right)s^2/\sigma^2 \sim \chi^2\left(n-k\right)$ dır.

(iii) $\hat{\beta}$ ve s^2 bağımsızdır.

Öte yandan bir hipotez tahmin edilebilir fonksiyonlara göre ifade edilemedikçe, onun test edilemediği gösterilebilir. Bu durum bizi aşağıdaki tanıma götürür.

Tanım 3.4. Eğer bir hipotez tahmin edilebilir fonksiyonlara göre ifade edilebilir ise, bu takdirde ona test edilebilir denecektir. Genel olarak, test edilebilir bir hipotez,

$$C = \left(c'_{(1)}, c'_{(2)}, \ldots, c'_{(m)}\right)', \; c_{(i)}\beta = t_i, \; i = 1, 2, \ldots, m$$

olmak üzere,

$$H_0 : C\beta = t$$

olarak yazılabilir. Burada C nin tam satır ranka sahip olduğu, yani, $rank(C) = m$ olduğu ve $c'_{(i)}\beta$ nın her i için tahmin edilebilir olduğu kabul edilir.

Teorem 3.27. X $k < p \le n$ ranklı $n \times p$ matris olmak üzere, eğer $y \sim N_n\left(X\beta, \sigma^2 I\right)$ ise ve C, $C\beta$ m sayıda lineer bağımsız tahmin edilebilir fonksiyonların bir kümesi olacak şekilde, $m \le k$ ranklı $m \times p$ matris ve $\hat{\beta} = \left(X'X\right)^{-} X'y$ ise, bu takdirde

(i) $C\left(X'X\right)^{-} C'$ tekil değildir ve $\left(X'X\right)^{-}$ nin seçimine göre değişmez.

(ii) $C\hat{\beta} \sim N_m\left(C\beta, \sigma^2 C\left(X'X\right)^{-} C'\right)$ dir.

(iii) $\lambda = \left(C\beta\right)'\left[C\left(X'X\right)^{-} C'\right]^{-1}\left(C\beta\right)\big/ 2\sigma^2$ olmak üzere;

$$SSH/\sigma^2 = \left(C\hat{\beta}\right)'\left[C\left(X'X\right)^{-} C'\right]^{-1}\left(C\hat{\beta}\right)\big/\sigma^2 \;\sim\; \chi^2\left(m, \lambda\right)$$

dır.

(iv) $SSE/\sigma^2 = y'\left[I - X\left(X'X\right)^{-} X'\right]y\big/\sigma^2 \sim \chi^2\left(n-k\right)$ dır.

(v) SSH ve SSE bağımsızdır.

İspat. (i)

$$C\beta = (c_{(1)}\beta \quad \cdots \quad c_{(m)}\beta)'$$

m sayıda lineer bağımsız tahmin edilebilir fonksiyonun bir kümesi olduğundan, Teorem 3.20 (iii) ye göre, $i = 1, 2, \ldots, m$ için, $c_{(i)}\left(X'X\right)^{-} X'X = c_{(i)}$ elde edilir. Bu nedenle,

$$C(X'X)^{-} X'X = C \tag{3.63}$$

yazılabilir. Öte yandan $rank(AB) \le rank(A)$ olduğundan,

$$rank(C) \le rank(C(X'X)^{-} X') \le rank(C)$$

elde edilir. Yani,

$$rank(C(X'X)^{-}X') = m$$

dir. Bu durumda $rank(A) = rank(AA')$ olduğundan,

$$\begin{aligned} rank(C) &= rank\left(C(X'X)^{-}X'\right) \\ &= rank\left(C(X'X)^{-}X'X(X'X)^{-}C'\right) \qquad (3.64) \\ &= rank\left(C(X'X)^{-}C'\right) \end{aligned}$$

elde edilir. Son eşitlikte, $C(X'X)^{-}X'X = C$ eşitliğini kullanalım. Bu nedenle $C(X'X)^{-}C'$ tekil değildir. Öte yandan $C(X'X)^{-}C'$ nün değişmezliği $X(X'X)^{-}X'$ nün değişmezliğinden görülür.

(ii) Bu durumda

$$E\left(C\hat{\beta}\right) = CE\left(\hat{\beta}\right) = C(X'X)^{-}X'X\beta$$

dır. Böylece (3.63) e göre, $E\left(C\hat{\beta}\right) = C\beta$ elde edilir. Ayrıca

$$kov\left(C\hat{\beta}\right) = Ckov\left(\hat{\beta}\right)C' = \sigma^2 C(X'X)^{-}X'X(X'X)^{-}C'$$

dir. Dolayısıyla (3.64) e göre, $kov\left(C\hat{\beta}\right) = \sigma^2 C(X'X)^{-}C'$ elde edilir. $C\hat{\beta}$ y nin bir lineer fonksiyonu olduğundan, (ii) ispatlanmış olur.

(iii) (ii) kısmına göre, $kov\left(C\hat{\beta}\right) = \sigma^2 C(X'X)^{-}C'$ dir. Buradan

$$\sigma^2 (C(X'X)^{-}C')^{-1} C(X'X)^{-}C' / \sigma^2 = I$$

olup istenen sonuç görülür. (iv) ve (v) benzer şekilde ispatlanabilir.

Teorem 3.28. X $k < p \leq n$ ranklı $n \times p$ matris olmak üzere $y \sim N_n\left(X\beta, \sigma^2 I\right)$ ve C, $C\beta$ ve $\hat{\beta}$ Teorem 3.27 de tanımlandığı gibi olsun. Bu takdirde eğer $H_0 : C\beta = 0$ hipotezi doğru ise,

$$F = \frac{SSH/m}{SSE/(n-k)} = \frac{(C\hat{\beta})'(C(X'X)C')^{-1}(C\hat{\beta})/m}{SSE/(n-k)} \sim F(m, n-k)$$

dır. Yeniden parametrelemede X, $k < p \leq n$ ranklı $n \times p$ matris olmak üzere $y = X\beta + \varepsilon$ eksik ranklı modelini Z, k ranklı $n \times k$ matris ve $\gamma = U\beta$, β nın k sayıda bağımsız tahmin edilebilir fonksiyonlarının bir kümesi olmak üzere, $y = Z\gamma + \varepsilon$ tam ranklı modeline dönüştürebiliriz. Bu nedenle $Z\gamma = X\beta$ olup ve $X = ZU$ olmak üzere,

$$Z\gamma = ZU\beta = X\beta$$

yazılabilir. Öte yandan UU' tekil olmadığından ($rank(UU') = k$),

$$ZUU' = XU'$$

ve

$$Z = XU'(UU')^{-1}$$

olacaktır. Şimdi Z tam –sütun ranklı bir matristir ($rank(Z) \geq rank(X) = k$) ve bu takdirde tam ranklı model için sonuçlar burada uygulanabilir. Bu nedenle

$$\hat{\gamma} = (Z'Z)^{-1}Z'y$$
$$s^2 = \frac{1}{n-k}(y - Z\hat{\gamma})'(y - Z\hat{\gamma}) = \frac{SSE}{n-k}$$

(en küçük kareler tahmin edicileri) elde edilir. $Z\gamma = X\beta$ olduğundan,

$$Z\hat{\gamma} = X\hat{\beta}$$

dır ve buradan

$$SSE = (y - X\hat{\beta})'(y - X\hat{\beta}) = (y - Z\hat{\gamma})'(y - Z\hat{\gamma})$$

olacaktır. Aynı zamanda herhangi bir $\lambda'\beta$ tahmin edilebilir fonksiyonu için

$$\lambda'\beta = a'X\beta = a'Z\gamma$$

ve buradan da

$$\lambda'\hat{\beta} = a'Z\hat{\gamma}$$

elde edilir, yani $\lambda'\beta$ nın tahmin edicisi yeniden parametrelemeyle değişmezdir.

Örnek 3.26. $y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad i = 1,2,\ j = 1,2$ modeli için yeniden parametreleme tekniğini açıklayalım. Matris formunda model

$$y = X\beta + \varepsilon = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \end{pmatrix}$$

olacaktır. *X*, 2 rankına sahip olduğundan iki lineer bağımsız tahmin edilebilir fonksiyon vardır. Bunları birçok şekilde seçebiliriz ki bunlardan birisi $\mu + \tau_1$ ve $\mu + \tau_2$ dir. Bu nedenle

$$\gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} \mu + \tau_1 \\ \mu + \tau_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix} = U\beta$$

ve

$$U = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

dir. $Z\gamma = X\beta$ ve $ZU = X$ olduğunu doğrulamak kolaydır.

Yan şartları koyma tekniği, parametreler bir tek ve ayrı ayrı tahmin edilebilir olacak şekilde bir tam ranklı olmayan model üzerindeki lineer kısıtlamaları temin eder. Yan şartlar için başka bir kullanım normal denklemleri basitleştirmek amacıyla tahminler üzerine kısıtlamalar koymaktır. Yan şartların, β nın tahmin edilemez fonksiyonları olmasına dikkat edelim.

X matrisi $k < p$ ranklı $n \times p$ tipinde bir matris olduğundan *X* in rankındaki eksiklik $p-k$ dır. Bu durumda tüm parametrelerin bir tek olması için, ranktaki bu eksikliği düzenleyen yan şartları tanımlamalıyız. Bu nedenle *T*, $T\beta$ tahmin edilebilir olmayan fonksiyonların bir kümesi olacak şekilde, $p-k$ ranklı bir $(p-k) \times p$ matris olmak üzere, $T\beta = 0$ yan şartlarını tanımlayabiliriz.

Teorem 3.29. Eğer X $k < p \le n$ ranklı $n \times p$ tipinde bir matris olmak üzere, $y = X\beta + \varepsilon$ ise ve T de $T\beta$ tahmin edilebilir olmayan fonksiyonların bir kümesi olacak şekilde, $p-k$ ranklı bir $(p-k)\times p$ matris ise, bu takdirde hem $X'X\hat{\beta} = X'y$ ve hem de $T\hat{\beta} = 0$ denklemlerini sağlayan bir tek $\hat{\beta}$ vektörü vardır.

İspat. İki denklemi birleştirerek,

$$\begin{pmatrix} y \\ 0 \end{pmatrix} = \begin{pmatrix} X \\ T \end{pmatrix}\beta + \begin{pmatrix} \varepsilon \\ 0 \end{pmatrix}$$

elde edilir. Bu nedenle $(X \quad T)(X \quad T)'$ p ranklı $p \times p$ matristir (yani, tekil değildir) ve bu durumda

$$\begin{aligned} \hat{\beta} &= \left[\begin{pmatrix} X \\ T \end{pmatrix}' \begin{pmatrix} X \\ T \end{pmatrix}\right]^{-1} \begin{pmatrix} X \\ T \end{pmatrix}' \begin{pmatrix} y \\ 0 \end{pmatrix} \\ &= \left(X'X + T'T\right)^{-1} X'y \end{aligned}$$

elde edilir ve böylece ispat tamamlanır.

Örnek 3.27. $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$; $i = 1, 2,\ j = 1, 2$ modelini göz önüne alalım. Bu durumda $\tau_1 + \tau_2$ nin tahmin edilebilir olmadığı gösterilebilir. $\tau_1 + \tau_2 = 0$ yan şartı $(0,1,1)\beta = 0$ olarak ifade edilebilir ve $X'X + T'T$ matrisi

$$X'X + T'T = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 3 & 1 \\ 2 & 1 & 3 \end{pmatrix}$$

olur. Bu takdirde,

$$\left(X'X + T'T\right)^{-1} = \frac{1}{4}\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & 0 \\ -1 & 0 & 2 \end{pmatrix}$$

dir. Ayrıca $X'y = \left(y_{..}, y_{1.}, y_{2.}\right)$ olmak üzere, $y_{1.}, y_{2.} = y_{..}$ ve $\bar{y}_{i.} = y_{i.}/2$ olduğundan,

$$\hat{\beta} = \left(X'X + T'T\right)^{-1} X'y = \begin{pmatrix} \bar{y}_{..} \\ \bar{y}_{1.} - \bar{y}_{..} \\ \bar{y}_{2.} - \bar{y}_{..} \end{pmatrix}$$

elde edilir.

Çizelge 3.2 Yeniden Parametrelenen Dengelenmiş Modellerde $H_0 : \gamma_1 = 0$ Hipotezini Test Etmek İçin ANOVA

Değişimin Kaynağı	s.d.	Kareler Toplamı	F - İstatistiği
γ_2 için düzeltilmiş γ_1 den dolayı	t	$SS(\gamma_1 \mid \gamma_2) = \hat{\gamma}'Z'y - \hat{\gamma}_2'Z_2'y$	$\dfrac{SS(\gamma_1 \mid \gamma_2)/t}{SSE/(n-k)}$
Hata	$n-k$	$SSE = y'y - \hat{\gamma}'Z'y$	
Toplam	$n-1$	$SST = y'y - n\bar{y}^2$	

Şimdi β, $p\times 1$ ve X $k < p \leq n$ ranklı $n \times p$ matris olmak üzere, $y = X\beta + \varepsilon$ tam ranklı olmayan modelinde $H_0 : \beta_1 = \cdots = \beta_q$ hipotezini test etmeyle ilgilendiğimizi farz edelim. Eğer H_0 test edilebilir ise,

$$H_0 : \gamma_1 = \begin{pmatrix} \lambda_1'\beta \\ \vdots \\ \lambda_t'\beta \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

a eşdeğer olacak şekilde $\lambda_1'\beta, \cdots, \lambda_t'\beta$ lineer bağımsız tahmin edilebilir fonksiyonlarının bir kümesini bulabiliriz. $k = rank(X)$ olmak üzere, $\lambda_1'\beta, \cdots, \lambda_k'\beta$ lineer bağımsız ve tahmin edilebilir olacak şekilde

$$\gamma_2 = \begin{pmatrix} \lambda_{t+1}'\beta \\ \vdots \\ \lambda_k'\beta \end{pmatrix}$$

yı da bulmak mümkündür.

$$\gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}$$

olsun. Şimdi $Z=(Z_1,Z_2)$, γ_1 ve γ_2 deki elemanların sayısı ile uyumlu olmak için parçalanmış olmak üzere, eksik ranklı modelden

$$y = Z\gamma + \varepsilon = Z_1\gamma_1 + Z_2\gamma_2 + \varepsilon$$

tam ranklı modeline yeniden parametreleme yapabiliriz. $y = Z\gamma + \varepsilon$ bir tam ranklı model olduğundan, $H_0 : \gamma_1 = 0$ hipotezi tam ranklı modeldeki gibi test edilebilir. Test, Çizgelge 3.2 de özetlenir. $SS(\gamma_1 \mid \gamma_2)$ için serbestlik derecesinin, yani, t nin, H_0 ı ifade etmek için gereken lineer bağımsız tahmin edilebilir fonksiyonların sayısı olduğuna dikkat edelim.

Yeniden parametreleme modelinde $X\hat{\beta} = Z\hat{\gamma}$ elde ettiğimizden $\hat{\beta}$, $X'X\hat{\beta} = X'y$ normal denklemine herhangi bir çözüm olmak üzere,

$$\hat{\beta}'X'y = \hat{\gamma}'Z'y$$

elde edilir. Benzer şekilde $y = Z_2\gamma_2^* + \varepsilon^*$ a karşılık getirerek $\beta_1 = \cdots = \beta_q$ alarak elde edilen bir $y = X_2\beta_2^* + \varepsilon^*$ indirgenmiş modelini elde edelim. Bu takdirde

$$\hat{\beta}_2'^{*} X_2' y = \hat{\gamma}_2' Z_2' y$$

dir.

Çizelge 3.3 Yeniden parametrelenen dengelenmiş modellerde $H_0 : \gamma_1 = 0$ ı test etmek için ANOVA

Değişimin Kaynağı	s.d.	Kareler Toplamı	F - İstatistiği
β_2 için düzeltilmiş β_1 den dolayı	t	$SS(\beta_1 \mid \beta_2) = \hat{\beta}'X'y - \hat{\beta}_2'X_2'y$	$\dfrac{SS(\beta_1 \mid \beta_2)/t}{SSE/(n-k)}$
Hata	$n-k$	$SSE = y'y - \hat{\beta}'X'y$	
Toplam	$n-1$	$SST = y'y - n\bar{y}^2$	

Bu takdirde test, Çizelge 3.3 teki gibi ifade edilebilir ki burada $\hat{\beta}'X'y$, $y = X\beta + \varepsilon$ tam modelinden ve $\hat{\beta}_2'X_2'y$, $H_0 : \beta_1 = \cdots = \beta_q$ hipoteziyle indirgenmiş olan, $y = X_2\beta_2 + \varepsilon$ modelinden elde edilir.

Şimdi dengelenmiş modellerde varyans analizi problemini göz önüne alalım. Bu durumda bir – yönlü dengelenmiş model aşağıdaki gibi ifade edilebilir.

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \ , \quad i = 1, 2, \ldots, k \ , \ j = 1, 2, \ldots, n \tag{3.65}$$

Eğer $\alpha_1, \cdots, \alpha_k$, her biri *n* deneysel birime uygulanan *k* sayıda işlemin etkilerini gösterirse, bu takdirde y_{ij} *i*–yinci işleme maruz kalan *n* birim arasında *j*–yinci işlemin tepkimesidir (cevap değişkenidir). Model için varsayımlar:

(1) Her i, j için, $E\left(\varepsilon_{ij}\right) = 0$.

(2) Her i, j için, $\operatorname{var}\left(\varepsilon_{ij}\right) = \sigma^2$.

(3) Her $(i, j) \neq (r, s)$ için, $kov\left(\varepsilon_{ij}, \varepsilon_{rs}\right) = 0$ dır.

(4) Bazen, ε_{ij} nin $N\left(0, \sigma^2\right)$ olarak dağıldığı dağılım varsayımına sahibiz.

Bu modelde, ekseriyle i –yinci işlem için ortalamayı, yani, $E\left(y_{ij}\right) = \mu_i$ yi göstermek için μ_i yi kullanırız. (1) varsayımını kullanarak, $\mu_i = \mu + \alpha_i$ elde ederiz. Bu nedenle, modeli

$$y_{ij} = \mu_i + \varepsilon_{ij}, \ i = 1, 2, \ldots, k, \ j = 1, 2, \ldots, n$$

biçiminde yazabiliriz. Modelin bu biçiminde $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$ hipotezi ilgi çekicidir. Bu durumda (3.65) bağlantısı genel bir *k* ve *n* ye genişletilerek, bir–yönlü model matris formunda

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix} = \begin{pmatrix} j & j & 0 & \cdots & 0 \\ j & 0 & j & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ j & 0 & 0 & \cdots & j \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{pmatrix}$$

veya

$$y = X\beta + \varepsilon$$

olarak yazılabilir. Burada j ve 0 ların her biri $n\times 1$ boyutludur ve y_i ve ε_i ler sırasıyla

$$y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in} \end{pmatrix}, \ \varepsilon_i = \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{in} \end{pmatrix}$$

olarak tanımlanır. Bu nedenle, $X'X\hat{\beta} = X'y$ normal denklemleri

$$\begin{pmatrix} kn & n & n & \cdots & n \\ n & n & 0 & \cdots & 0 \\ n & 0 & n & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ n & 0 & 0 & \cdots & n \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_k \end{pmatrix} = \begin{pmatrix} y_{..} \\ y_{1\cdot} \\ y_{2\cdot} \\ \vdots \\ y_{k\cdot} \end{pmatrix}$$

biçimini alır. Burada $y_{..} = \sum_{ij} y_{ij}$ ve $y_{i\cdot} = \sum_j y_{ij}$ dir. Bu durumda $X'X$ in bir genelleştirilmiş tersi

$$(X'X)^{-} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1/n & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1/n \end{pmatrix} \tag{3.66}$$

ile verilir. Bu takdirde normal denkleme bir çözüm

$$\hat{\beta} = (X'X)^{-} X'y = \begin{pmatrix} 0 & \bar{y}_{1\cdot} & \cdots & \bar{y}_{k\cdot} \end{pmatrix}' \tag{3.67}$$

olarak elde edilir. (3.67) deki tahmin ediciler farklı $\left(X'X\right)^{-}$ için farklıdır, fakat $\lambda'\hat{\beta}$ $\hat{\beta}$ nın seçimine göre değişmez kaldığından, onlar tahmin edilebilir fonksiyonların aynı tahminlerini verir. (3.67) deki $\hat{\beta}$ yı kullanarak SSE yi aşağıdaki biçimde ifade edebiliriz.

$$
\begin{aligned}
SSE &= y'(I - X(X'X)^{-} X')y \\
&= y'y - \hat{\beta}' X' y \\
&= \sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij}^{2} - \sum_{i=1}^{k} \overline{y}_{i\cdot} y_{i\cdot} \\
&= \sum_{ij} y_{ij}^{2} - \sum_{i} \frac{y_{i\cdot}^{2}}{n}
\end{aligned}
$$

Bu nedenle, s^2

$$
s^2 = \frac{1}{k(n-1)} \left[\sum_{ij} y_{ij}^2 - \sum_{i} \frac{y_{i\cdot}^2}{n} \right]
$$

ile verilir.

Şimdi de $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$ hipotezini göz önüne alalım. $\mu_i = \mu + \alpha_i$ ilişkisini kullanarak hipotez $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_k$ olarak ifade edilebilir. Bu durumda $H_0 : \alpha_1 - \alpha_2 = \alpha_1 - \alpha_3 = \cdots = \alpha_1 - \alpha_k = 0$ hipotezi test edilebilirdir. Basitlik için, $k = 4$ olmak üzere test etme yöntemini açıklayalım. Bu durumda, $\beta = (\mu, \alpha_1, \alpha_2, \alpha_3, \alpha_4)'$ olup hipotez $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$ dür. Böylece hipotez

$$
C = \begin{pmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \end{pmatrix} \tag{3.68}
$$

olmak üzere, $H_0 : C\beta = 0$ olarak ifade edilebilen

$$
H_0 : \begin{pmatrix} \alpha_1 - \alpha_2 \\ \alpha_1 - \alpha_3 \\ \alpha_1 - \alpha_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}
$$

biçiminde yazılabilir. (3.68) deki C matrisi tek değildir. Örneğin;

$$
C_2 = \begin{pmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}
$$

de C yerine kullanılabilir. (3.68) deki C yi ve (3.66) daki $(X'X)^{-}$ kullanarak,

$$C\left(X'X\right)^{-}C' = \frac{1}{n}\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \tag{3.69}$$

elde edilir. (3.69) un tersini bulmak için onu

$$C\left(X'X\right)^{-}C' = \frac{1}{n}\left\{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}\right\} = \frac{1}{n}\left(I_3 + j_3 j_3'\right)$$

biçiminde yazalım. J_3, 3×3 olmak üzere

$$(B+cc')^{-1} = B^{-1} - \frac{B^{-1}cc'B^{-1}}{1+c'B^{-1}c}$$

formülü vasıtasıyla

$$\left[C\left(X'X\right)^{-}C'\right]^{-1} = n\left(I_3 - \frac{I_3^{-1} j_3 j_3' I_3^{-1}}{1+ j_3' I_3^{-1} j_3}\right) = n\left(I_3 - \frac{1}{4}J_3\right) \tag{3.70}$$

olduğu görülür. Ayrıca j_n' ve $\underline{0}'$, $\left(A-C\right)$ olmak üzere,

$$C\left(X'X\right)^{-}X' = \frac{1}{n}\begin{pmatrix} j_n' & -j_n' & \underline{0}' & \underline{0}' \\ j_n' & \underline{0}' & -j_n' & \underline{0}' \\ j_n' & \underline{0}' & \underline{0}' & -j_n' \end{pmatrix} = \frac{1}{n}A \tag{3.71}$$

yazılabilir. Bu durumda (3.70) ve (3.71) i kullanarak,

$$
\begin{aligned}
SSH &= \left(C\hat{\beta}\right)'\left[C\left(X'X\right)^{-}C'\right]^{-1}C\hat{\beta} \\
&= y'X\left(X'X\right)^{-}C'\left[C\left(X'X\right)^{-}C'\right]^{-1}C\left(X'X\right)^{-}X'y \\
&= y'\left[\frac{1}{n}A'n\left(I_3-\frac{1}{4}J_3\right)\frac{1}{n}A\right]y \\
&= y'\left[\frac{1}{n}A'A-\frac{1}{4n}A'J_3A\right]y \\
&= y'\left[\frac{1}{4n}\begin{pmatrix} 3J_n & -J_n & -J_n & -J_n \\ -J_n & 3J_n & -J_n & -J_n \\ -J_n & -J_n & 3J_n & -J_n \\ -J_n & -J_n & -J_n & 3J_n \end{pmatrix}\right]y \\
&= y'\left[\frac{1}{4n}B\right]y
\end{aligned}
$$

olduğu görülür. Burada

$$
\frac{1}{4n}B=\frac{1}{n}\begin{pmatrix} 4J_n & 0 & 0 & 0 \\ 0 & 4J_n & 0 & 0 \\ 0 & 0 & 4J_n & 0 \\ 0 & 0 & 0 & 4J_n \end{pmatrix}-\frac{1}{4n}\begin{pmatrix} J_n & J_n & J_n & J_n \\ J_n & J_n & J_n & J_n \\ J_n & J_n & J_n & J_n \\ J_n & J_n & J_n & J_n \end{pmatrix}
$$

olduğuna dikkat edelim. Dolayısıyla

$$
\begin{aligned}
SSH &= \frac{1}{n}\sum_{i=1}^{4} y_i' J_n y_i - \frac{1}{4n}y'J_{4n}y \\
&= \frac{1}{n}\sum_{i=1}^{4} y_i' j_n j_n' y_i - \frac{1}{4n}y'j_{4n}j_{4n}'y \\
&= \frac{1}{n}\sum_{i=1}^{4} y_{i\cdot}^2 - \frac{1}{4n}y_{\cdot\cdot}^2
\end{aligned}
$$

elde edilir.

Çizelge 3.4 Üç Paketleme Yöntemi İçin Askorbik Asit (C Vitamini) (Mg/100g)

	A	B	C
	14.29	20.06	20.04
	19.10	20.64	26.23
	19.09	18.00	22.74
	16.25	19.56	24.04
	15.09	19.47	23.37
	16.61	19.07	25.02
	19.63	18.38	23.27
Toplamlar $(y_{i\cdot})$	120.06	135.18	164.71
Ortalamalar $(\bar{y}_{i\cdot})$	17.15	19.31	23.53

Örnek 3.28. (Askorbik Asit) Daniel (1974,s.196) tarafından dondurulmuş gıdaları paketlemenin üç yöntemi $(A-C)$ karşılaştırılmıştır. Tepkime değişkeni askorbik asittir. Veriler Çizelge 3.5 te verilmiştir. $H_0 : \mu_1 = \mu_2 = \mu_3$ hipotezini test etmek için,

$$\frac{y_{\cdot\cdot}^2}{kn} = \frac{419.95^2}{(3)(7)} = 8398.0001$$

$$\frac{1}{7}\sum_{i=1}^{3} y_{i\cdot}^2 = \frac{1}{7}\left[120.06^2 + 135.18^2 + 164.71^2\right] = 8545.3457$$

$$\sum_{i=1}^{3}\sum_{j=1}^{7} y_{ij}^2 = 8600.3127$$

yi hesaplayalım.

Çizelge 3.5 Askorbik Asit Verisi İçin Varyans Analizi

Kaynak	s.d.	Kareler Toplamı	Ortalama Kare	F
Yöntem	2	147.3456	73.6728	24.1256
Hata	18	54.9670	3.9537	
Toplam	20	202.312		

Bu takdirde

$$SSH = \frac{1}{7}\sum_{i=1}^{3} y_{i\cdot}^2 - \frac{y_{\cdot\cdot}^2}{21} = 8545.3457 - 8398.0001 = 147.3456$$

$$SSE = \sum_{ij} y_{ij}^2 - \frac{1}{7}\sum_{i} y_{i\cdot}^2 = 8600.3127 - 8545.3457 = 54.9670$$

$$SST = \sum_{ij} y_{ij}^2 - \frac{y_{..}^2}{21} = 8600.3127 - 8398.0001 = 202.3126$$

olup bu kareler toplamları Çizelge 3.5 te gösterildiği gibi, bir *F*–testi elde etmek için kullanılabilir. $F = 24.1256$ için *p*–değeri 8.07×10^{-6} dır. Bu nedenle, $H_0 : \mu_1 = \mu_2 = \mu_3$ hipotezi reddedilir.

Şimdi iki – yönlü dengelenmiş model durumunu göz önüne alalım. Bu durumda

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij},\ i = 1,\ldots,a,\ j = 1,\ldots,b$$

modeline sahip olduğumuz farz edelim. Bu model dengelenmiş veri ile iki–faktörlü modeldir. Burada A (α için) faktörü *a* sayıda, B (β için) faktörü *b* sayıda düzeye sahiptir. Her bir (i, j) kutucuğunda sadece bir y_{ij} gözlemi vardır. Matris formunda bu model, $y = X\beta + \varepsilon$ olarak yazılabilir. Burada

$$y = (y_{11}, \ldots y_{1b,} y_{21}, \ldots, y_{2b}, \ldots, y_{a1}, \ldots y_{ab})',$$

$$\beta = (\mu, \alpha_1, \ldots \alpha_a, \beta_1, \ldots, \beta_b),$$

$$\varepsilon = \left(\varepsilon_{11}, \cdots, \varepsilon_{1b}, \varepsilon_{21}, \cdots, \varepsilon_{2b}, \cdots, \varepsilon_{a1}, \cdots, \varepsilon_{ab}\right),$$

$$X = \begin{pmatrix} \mu & \alpha_1 & \alpha_2 & \cdots & \alpha_a & \beta_1 & \beta_2 & \cdots & \beta_b \\ 1 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

olup

$$X'X = \begin{pmatrix} ab & b & b & \cdots & b & a & a & \cdots & a \\ b & b & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \\ b & 0 & b & \cdots & 0 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ b & 0 & 0 & \cdots & b & 1 & 1 & \cdots & 1 \\ a & 1 & 1 & \cdots & 1 & a & 0 & \cdots & 0 \\ a & 1 & 1 & \cdots & 1 & 0 & a & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a & 1 & 1 & \cdots & 1 & 0 & 0 & \cdots & a \end{pmatrix}$$

dır. Burada $(X'X)^{-}$ genelleştirilmiş tersini bulmak kolay değildir. Bu nedenle $X'X\hat{\beta} = X'y$ normal denklemlerini çözmek için onun yerine, iki $\sum_{i=1}^{a} \alpha_i = 0$ ve $\sum_{j=1}^{b} \beta_j = 0$ yan şartını koyabiliriz. Dolayısıyla

$$X'X\hat{\beta} = \begin{pmatrix} ab\mu + b\sum_i \alpha_i + a\sum_j \beta_j \\ b(\mu+\alpha_1) + \sum_j \beta_j \\ \vdots \\ b(\mu+\alpha_a) + \sum_j \beta_j \\ a(\mu+\beta_1) + \sum_i \alpha_i \\ \vdots \\ a(\mu+\beta_b) + \sum_i \alpha_i \end{pmatrix} = \begin{pmatrix} ab\mu \\ b(\mu+\alpha_1) \\ \vdots \\ b(\mu+\alpha_a) \\ a(\mu+\beta_1) \\ \vdots \\ a(\mu+\beta_b) \end{pmatrix}$$

$$X'y = (y_{..}, y_{1.}, \ldots, y_{a.}, y_{.1}, \ldots, y_{.b})'$$

olup, çözüm

$$\begin{aligned} \hat{\mu} &= y_{..}/(ab) = \bar{y}_{..} \\ \hat{\alpha}_i &= y_{i.}/b - \hat{\mu} = \bar{y}_{i.} - \bar{y}_{..}, \ i = 1, \ldots, a \\ \hat{\beta}_j &= y_{.j}/a - \hat{\mu} = \bar{y}_{.j} - \bar{y}_{..}, \ j = 1, \ldots, b \end{aligned}$$

olmak üzere,

$$\hat{\beta} = \left(\hat{\mu}, \hat{\alpha}_1, \ldots, \hat{\alpha}_a, \hat{\beta}_1, \ldots, \hat{\beta}_b\right)'$$

dür. Şimdi Çizelge 3.3 teki çerçeveyi izleyerek, $H_0 : \alpha_1 = \cdots = \alpha_a$ hipotezini test edelim. Bu durumda $H_0 : \alpha_1 = \alpha_2 = \alpha_3$ hipotezi $H_0 : \alpha_1 - \alpha_2 = 0$ ve $\alpha_1 - \alpha_3 = 0$

olarak ifade edilebilir. Bu nedenle; eğer, $\alpha_1-\alpha_2$ ve $\alpha_1-\alpha_3$ tahmin edilebilir ise, H_0 hipotezi test edilebilirdir. Gözlemlerin her bir $E\left(y_{ij}\right)=\mu+\alpha_i+\beta_j$ beklenen değeri tahmin edilebilir ve $\left(\mu+\alpha_i+\beta_j\right)$ lerin herhangi bir lineer kombinasyonu tahmin edilebilir olduğundan, $\alpha_1-\alpha_2$ ve $\alpha_1-\alpha_3$ farklarının her ikisi de tahmin edilebilirdir.

$\sum_i y_{i\cdot}=y_{\cdot\cdot}$ ve $\sum_j y_{\cdot j}=y_{\cdot\cdot}$ olduğundan,

$$SS(\mu,\alpha,\beta)=\hat{\beta}X'y=\left(\hat{\mu},\hat{\alpha}_1,\ldots,\hat{\alpha}_a,\hat{\beta}_1,\ldots,\hat{\beta}_b\right)'\begin{pmatrix} y_{\cdot\cdot} \\ y_{1\cdot} \\ \cdots \\ y_{a\cdot} \\ y_{\cdot 1} \\ \cdots \\ y_{\cdot b} \end{pmatrix}$$

$$=\bar{y}_{\cdot\cdot}y_{\cdot\cdot}+\sum_{i=1}^{a}\left(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot}\right)y_{i\cdot}+\sum_{j=1}^{b}\left(\bar{y}_{\cdot j}-\bar{y}_{\cdot\cdot}\right)y_{\cdot j}$$

$$=\frac{y_{\cdot\cdot}^2}{ab}+\left(\sum_{i=1}^{a}\frac{y_{i\cdot}^2}{b}-\frac{y_{\cdot\cdot}^2}{ab}\right)+\left(\sum_{j=1}^{b}\frac{y_{\cdot j}^2}{a}-\frac{y_{\cdot\cdot}^2}{ab}\right)$$

ve SSE hata kareler toplamı

$$y'y-\hat{\beta}'X'y=\sum_{ij}y_{ij}^2-\frac{y_{\cdot\cdot}^2}{ab}-\left(\sum_{i=1}^{a}\frac{y_{i\cdot}^2}{b}-\frac{y_{\cdot\cdot}^2}{ab}\right)-\left(\sum_{j=1}^{b}\frac{y_{\cdot j}^2}{a}-\frac{y_{\cdot\cdot}^2}{ab}\right)$$

ile verilir. Çizelge 3.3 teki $\hat{\beta}_2'X_2'y$ yi elde etmek için, $y_{ij}=\mu+\alpha+\beta_j+\varepsilon_{ij}=\mu+\beta_j+\varepsilon_{ij}$ indirgenmiş modelini kullanalım. Burada $\alpha_1=\alpha_2=\alpha_3=\alpha$ ve μ yerine $\mu+\alpha$ alınırsa indirgenmiş model için $X_2'X_2\hat{\beta}_2=X_2'y$ normal denklemleri

$$\begin{aligned} ab\hat{\mu}+a\hat{\beta}_1+a\hat{\beta}_2&=y_{\cdot\cdot} \\ a\hat{\mu}+a\hat{\beta}_1&=y_{\cdot 1} \\ a\hat{\mu}+a\hat{\beta}_2&=y_{\cdot 2} \end{aligned}$$

olacaktır. Dolayısıyla $\hat{\beta}_1+\hat{\beta}_2=0$ yan şartı kullanılarak, indirgenmiş normal denklemlerin çözümünün

$$\hat{\mu}=\bar{y}_{..},\quad \hat{\beta}_1=\bar{y}_{.1}-\bar{y}_{..},\quad \cdots,\quad \hat{\beta}_b=\bar{y}_{.b}-\bar{y}_{..}$$

olduğu kolayca görülür. Böylece

$$SS(\mu,\beta)=\hat{\beta}_2' X_2' y=\hat{\mu} y_{..}+\hat{\beta}_1 y_{.1}+\cdots+\hat{\beta}_b y_{.b}=\frac{y_{..}^2}{ab}+\left(\sum_{j=1}^{b}\frac{y_{.j}^2}{a}-\frac{y_{..}^2}{ab}\right)$$

$$SS(\alpha\mid\mu,\beta)=\hat{\beta}' X' y-\hat{\beta}_2' X_2' y=\sum_{i=1}^{a}\frac{y_{i.}^2}{b}-\frac{y_{..}^2}{ab}$$

yazılabilir. Test Çizelge 3.6 da özetlenmiştir. $H_0:\beta_1=\cdots=\beta_b$ için test istatistiği benzer şekilde elde edilebilir.

Çizelge 3.6 İki Yönlü Modeller İçin ANOVA

Varyansın Kaynağı	s.d.	Kareler Toplamı	F istatistiği
μ ve β için düzeltilen α dan dolayı	$a-1$	$SS(\alpha\mid\mu,\beta)=\sum_{i=1}^{a}\frac{y_{i.}^2}{b}-\frac{y_{..}^2}{ab}$	$F_1=\frac{SS(\alpha\mid\mu,\beta)/(a-1)}{SSE/(a-1)(b-1)}$
μ ve α için düzeltilen β dan dolayı	$b-1$	$SS(\beta\mid\mu,\alpha)=\sum_{j=1}^{b}\frac{y_{.j}^2}{a}-\frac{y_{..}^2}{ab}$	$F_2=\frac{SS(\beta\mid\mu,\alpha)/(b-1)}{SSE/(a-1)(b-1)}$
Hata	$(a-1)(b-1)$	$SSE=y'y-\hat{\beta}' X' y$	
Toplam	$ab-1$	$SST=\sum_{ij} y_{ij}^2-\frac{y_{..}^2}{ab}$	

4. SONUÇ ve ÖNERİLER

Bu çalışmada matris teorisi ve istatistikle ilgili olan bazı temel bilgiler tanıtılarak bunların lineer modellerde kullanımı ele alınmıştır. Lineer modellerde lineer kısıtlamalar altında ve herhangi bir kısıtlama olmaksızın parametre tahmini elde edilmiştir. Öte yandan determinasyon katsayısı ve bunun varyans analizi üzerindeki etkisi üzerinde durulmuştur. Ayrıca determinasyon katsayısının farklı geometrik yorumları ifade edilmiştir.

Lineer modellere geometrik bakış açısı ile soyut kavramların somutlaştırılması hedeflenmiştir. Bununla ilgili olarak öncelikle geniş bir literatür taraması yapılmış, bu konuda yazılan son makaleler ele alınarak mevcut gelişmelere yeni katkılar yapılmaya çalışılmıştır. Elde edilen sonuçlar örneklerle ve şekillerle daha anlaşılır hale getirilmeye çalışılmıştır. Tam rank durumu ele alındıktan sonra bu durum eksik rank durumuna genişletilerek, eksik ranklı modellerdeki tahminler üzerinde durulmuştur.

Bu tezde ele alınan konuların tam ranklı modellerde yapılan çalışmaların eksik rank durumunda da yapılabilmesi konusunda araştırmacılara bir miktar da olsa katkı sağlayacağı düşünülmektedir.

5. KAYNAKLAR

Ackley, FR. (1967). The use of the generalized Inverse in the General Linear Statistical Model. Msc. Thesis, Naval Postgraduate School, Monterey, California.

Bean D. (2013). Linear Models. Lecture03 - Fall 2013 Statistics 151(Linear Models Lecture Three Derek Bean 05 September 2013 1 Linear Algebra Review contd Result for matrix A rank(A | Course Hero-(Erişim tarihi:22.03.2021).

Blalock HM Jr (1967). Status inconsistency, social mobility, status integration and structural effects. American Sociological Review 32: 790–801.

Cayley, A. (1858). A Memoir on The Theory of Matrices. *Philosophical Transactions of the Royal Society of London*, 148, 17-37.

Cline, RE. (1964). Representations fort he Genaralized Inverse of a Partitioned Matrix. *Journal SIAM*, 12, 588-600.

Duncan OD (1966). Methodological issues in the analysis of social mobility. In: Smelser NJ and Lipset SM, editors. Social structure and mobility in economic development. Chicago: Aldine, 51–97.

Graybill, FA (1961). An Introduction to Linear Statistical Models. Volume 1. *New York: McGrawHill Book Company, Inc*.

Greville, TNE. (1966). Note on the Generalized Inverse of a Matrix Product. *SIAM Review*, 8, 518-21.

Holford TR (2006) Approaches to fitting age-period-cohort models with unequal intervals. Statistics in Medicine, 25, 977–993.

Kuang D., Nielsen, B Nielsen, J.P. (2008) Identification of the age-period-cohort model and the extended chain ladder model. Biometrika, 95, 979–986.

Larsen RJ. & Marx. ML. (2012). The Standard Linear Model: Hypothesis Testing. [PDF] OF TECHNOLOGY Ma 3 / 103 KC Border Introduction to Probability and Statistics Winter 2015 Lecture 24 : The Standard Linear Model : Hypothesis Testing | Semantic Scholar-(Erişim Tarihi:22.03.2021).

Mason KO, Mason WH, Winsborough HH Poole K (1973) Some methodological issues in cohort analysis of archival data. American Sociological Review, 38, 242–258.

Mazumdar S, Li CC, Bryce GR (1980) Correspondence between a linear restriction and a generalized inverse in linear model analysis. The American Statistician, 34, 103–105.

Moore & Eliakim M. (1935). General Analysis. Philadelphia: *The American Philosophical Society*.

O'Brien RM. (2011) Constrained estimators and age-period-cohort models. Sociological Methods & Research 40: 419–452.

O'Brien RM. (2011) Intrinsic estimators as constrained estimators in ageperiod-cohort accounting models. *Sociological Methods & Research*, 40, 467– 470.

O'Brien RM. (2012) Visualizing Rank Deficient Models: A Row Equation Geometry of Rank Deficient Matrices and Constrained-Regression. *Plos One* 7(6)*:e38923. doi:10.1371.*

Owen A. (2011). Linear Least Squares. ch2.pdf (stanford.edu)-(Erişim tarihi:22.03.2021).

Penrose, R. (1955). A Generalized Inverse for Matrices. *Proceedings of the Cambridge Philosophical Society*, 51, 406-13.

Rao, CR. (1965). On Discrete Distributions Arising out of Methods of Ascertainment. *Sankhya*; *The Ind.J. Statist*, *Series A*, 27, 311-324

Shalabh S. (2020). Multiple Linear Regression Model. Microsoft Word - Chapter3-Regression-MultipleLinearRegressionModel.doc (iitk.ac.in)-(Erişim tarihi:22.03.2021).

Yang Y, Schulehoffer-Wohl S, Fu WJ, Land KC (2008) The intrinsic estimator for age-period-cohort analysis: what it is and how to use it. *American Journal of Sociology*, 113, 1697–1736.

ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı	Hakan GEZGİN
Doğum Yeri	Ünye
Doğum Tarihi	18.10.1990
Uyruğu	❒ T.C. ❒ Diğer:
Telefon	0544 827 27 20
E-Posta Adresi	hakanngezgin@gmail.com

Eğitim Bilgileri Lisans	
Üniversite	Mustafa Kemal Üniversitesi
Fakülte	Fen Edebiyat Fakültesi
Bölümü	Matematik
Mezuniyet Yılı	16.09.2020
Yüksek Lisans	
Üniversite	Ordu Üniversitesi
Enstitü Adı	Fen Bilimleri Enstitüsü
Anabilim Dalı	Matematik Anabilim Dalı
Programı	Program Adı
Mezuniyet Tarihi	23.05.2016
Doktora	
Üniversite	
Enstitü Adı	
Anabilim Dalı	
Programı	
Mezuniyet Tarihi	

Yayınlar

[1]Gezgin, H. , Yalçın F. B. "Easy Correlation Tricks, Geometric View on Partial Correlation and Correlation Between Two Random Events" , İcanas 2017 Antalya.

[2]Yalçın F. B., Gezgin, H. " Equality Constrained Least Squares and Maximum Likelihood Estimators in Classical Linear Regression Model" , İcanas 2017 Antalya.

[3] Gezgin H., Korkmaz M., "A Brief Overview on the Geometry of the Uncentered Coefficient of Determination" Icmme 2018 Ordu.